Abstract

The present doctoral research investigates the role of local knowledge in supporting disaster response, by applying cognitive, predictive and ontological models to the study of text message exchange in relevant Social Networks.
The lack of studies on local knowledge in risk domains, along with the growing attention of scholars and decision makers to open governance processes, underpin the three main research questions that have been addressed.
As for the relevance of local knowledge to disaster response, human communication – in real or simulated emergency situations – seems to be imbued with culturally mediated understandings of spatiality, relationality and actions. Innovative spatial data science tools are therefore needed for tacit and vernacular knowledge to be adequately modelled and operationalized. With respect to the use of text messages in disaster response, it appears that distributed systems (which combine crowdsourcing methods with collaborative hypertext editors) may effectively complement volunteered or public participation GIS – to harness the potential of all-purpose social networks to reach out to the wider internet community.
Finally, to foster the interpretation of text messages, three separate taxonomies (regarding Spatial Location, Needs and Actors) – each being linked to a terminal entity in DOLCE foundational ontology –, helped develop a shared conceptualization of risk.
Future developments of the present work could concern the further integration between machine learning and ontological models, and advancing text classification methods.

2017

Pasquale Balena

Local Knowledge and Social Sensor: Integrated Models of Text Analysis for Disaster Response

03

POLITECNICO DI BARI

D.R.R.S.

03

2017

PhD in Risk and Environmental, Territorial and Building Development

Coordinator: Prof. Michele Mossa

XXIX CYCLE
Curriculum: Infrastructure, Transport and Territory

DICATECh
Department of Civil, Environmental, Land, Building Engineering and Chemistry

Pasquale Balena

Local Knowledge and Social Sensor: Integrated Models of Text Analysis for Disaster Response

Prof. Dino Borri
Department of Civil, Environmental, Land, Building Engineering and Chemistry
Technical University of Bary, Italy
Prof. Caterina De Lucia
Department of Economics
University of Foggia, Italy

D.R.R.S.

POLITECNICO DI BARI

03

PhD in Risk and Environmental, Territorial and Building Development

2017

Coordinator: Prof. Michele Mossa

XXIX CYCLE
Curriculum: Infrastructure, Transport and Territory

DICATECh

Department of Civil, Environmental, Land, Building Engineering and Chemistry

Local Knowledge and Social Sensor: Integrated Models of Text Analysis for Disaster Response

Prof. Dino Borri
Department of Civil, Environmental, Land, Building Engineering and Chemistry
Technical University of Bary, Italy

Prof. Caterina De Lucia
Department of Economics
University of Foggia, Italy

Pasquale Balena

POLITECNICO DI BARI

D.R.R.S.

03

2017

Dottorato di Ricerca in Rischio e Sviluppo ambientale, territoriale ed edilizio

Coordinatore: Prof. Michele Mossa

XXIX CYCLE
Curriculum: Infrastrutture, Trasporto e Territorio

DICATECh

Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica

Conoscenza Locale e Social Sensor: Modelli Integrati di Text Analysis in Supporto alla Gestione dei Disastri

Prof. Dino Borri
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica
Politecnico di Bari, Italia

Prof. Caterina De Lucia
Dipartimento di Economia
Università di Foggia, Italia

Pasquale Balena

Table of Contents

Extended Abstract

English

The present doctoral work aims to contribute to the academic debate on the role of local knowledge in collective decision-making processes. The work analyses if and how local knowledge can contribute, through crowdsourcing and social sensing, to the creation of knowledge frameworks in risk domains.

The current state of the art highlights a long-term intellectual deadlock about local knowledge, which might explain why the international scientific community tends to refrain from carrying out (or to ignore the few existing) studies on local knowledge in risk domains. On the one hand, this marginalization is due to long-standing cultural traditions in several scientific disciplines, which emphasize expert knowledge rather than local knowledge; on the other hand, the importance of technical-methodological complexities is such that local knowledge seems to be less attractive than expert knowledge from a computational point of view.

Over the last few years, the institutional attention towards public participation in urban and regional governance has gained momentum. Thus, the international scientific debate is becoming more interested in local knowledge as a key factor in collaborative decision-making processes. The recent evolution of Information and Communication Technology (ICT) and mobile devices has strongly encouraged public e-participation as a tool for decision-support systems. These e-participation tools are labelled in the literature as Volunteered Geographic Information (VGI) or Participatory Geographic Information System (PGIS). The use of these e-participation tools has also extended to several domains – such as natural disasters, humanitarian crises, political conflicts – with the main aim to help affected populations and provide useful information for survival (Haiti, Kenia, etc.).

Nonetheless, e-participation tools present some drawbacks for managing non-structured information retrieved from large databases and Social Networks. The limitations concern either the need to understand knowledge in (almost) real time and to share it with experts and decision makers, or to facilitate mutual understanding in the context of those events – such as humanitarian crises and natural disasters – that mobilize heterogeneous cultures, languages and organizations.

The present research dwells on the above limitations, to tie information, analysis, interpretation and sharing of local knowledge in disaster response and, hence, to contribute to close the gap in the international literature. To reach this main purpose, the work focuses on developing an ontology for local knowledge in disaster response and on investigating, through predictive models, the effectiveness of information retrieval from VGI/PGIS technologies.

The dissertation is structured in four parts. After a brief introduction (**Chapter 1**) which provides a description of the aims, objectives and research questions, **Chapter 2** describes the conceptual background, with regard to the four main knowledge domains that are relevant to the investigation. The research design and methodology is covered in **Chapter 3**. **Chapter 4** deals with the empirical work and illustrates the main results obtained by the ontological model and the machine learning classification. Finally, main discussion, conclusions and recommendations are drawn in **Chapter 5**.

The conceptual background (in **Chapter 2**) includes four themes:

- Firstly, the issues of risk are analysed under their various forms and phases, to single out the potential applications of interest for the present work.
- Secondly, a two-stage framework knowledge is described, in which a critical discussion in support of specific aspects of disaster response follows a description of general concepts. This two-stage framework does not provide a classification of terms given that various shades across definitions and concepts exist and are widely acknowledged in the international debate
- Thirdly, the work delves into the state-of-the-art of, and trends in, ontological analysis

- Finally, the main research issues ensuing from data collection through crowdsourcing (VGI and PGIS) are presented.

Chapter 3 illustrates the design of the research process, by focussing on 3 methodological packages. The first trend provides a simulation analysis with the use of PGIS technologies. These platforms (e.g. Ushahidi) retrieve geo-referenced – data either directly (through specific functions of the platform) or indirectly (through streaming from Social Networks). Simultaneously, an open-ended web survey was organised. The second theme concerns applying text-mining and lexical statistics to create specific matrices for stochastic classification models (Aaboost, Random Forest, Support Vector Machine, Naïve Bayes, Neural Network) to support machine learning analysis. The third domain is the semantic and contextual analysis of the retrieved messages. It aims at identifying the lexicon (also based on two or three words) from natural language and local context to define the following concepts: needs for survival and for other requirements, actors and their roles in post-disaster domain, spatial aspects of actors and needs. In a second step, a spatial-relational model is applied to the above approaches to identify further topological closeness, directional and distance relations.

The empirical work is illustrated in **Chapter 4**, following the outline of the previous methodological part. Thus, an application of PGIS platforms to gather generic territorial information on the metropolitan area of Bari was run in parallel with an open-ended web survey: after illustrating a hypothetical earthquake in the metropolitan area of Bari, participants were asked to write a request for, or offer of, help. Data retrieved from Ushahidi and the on-line survey were stored in a database containing 1.300 messages in total. A subset of 314 observations contained messages from the post-earthquake domain only. The making of the full database proved to be useful to apply and validate text classification techniques in a subsequent stage of the research work. As for the stochastic classification models selected for the machine learning analysis, the results point to the robustness of such models, which are widely also used in the economic and medical literature. More in detail:

a) ***needs*** are articulated into 5 categories such as objects, services, communications, environmental and functional needs – each of the above categories is split into two sub-fields: 'needs for survival' and 'comfort';
b) ***actors*** encompass both single individual and groups. For each of these actors, specific roles in post-disaster domain are identified (i.e. exposed people, affected people, caregivers, or information providers);
c) ***spatial locations*** are defined through a double approach: the first one identifies the elements of a location in the framework of a reference system; the second one, based on natural language, provides a location (e.g. address, landmark, meeting places).

In respect of the ontological analysis, the previous results helped set up a taxonomy associated to the *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) foundational ontology, with the aim to create a shared ontology on disaster response.

A discussion of the results against the conceptual background and some concluding remarks are offered in **Chapter 5**, together with recommendations for better-informed practices and policies in disaster response, and prospects for future research developments. It is suggested that the present doctoral research contributes to closing the gap between understanding local knowledge through non-structured texts from mobile devices and designing more effective disaster management systems. The new tools and models for text analysis, though affected by the limitations of computational analysis, are capable of developing structured knowledge from un-structured data, by means of identify through *tokenization*. Future research may build on these functions and work towards integrated platforms to retrieve, analyse and manipulate unstructured data to support decision making. Important issues, such as information reliability from crowdsourcing, remain open. Also, the existence of false negatives from text classification systems may undermine the robustness of classification models. As a result, these types of messages are neither included in disaster response domains nor considered as local knowledge. Future developments of the present work should therefore be geared towards advancing text classification models and improving the analysis of latent knowledge in the disaster response domain.

Keywords

Local Knowledge, Ontologies, Social Sensing, Crowdsourcing, Risk, Disaster, Knowledge Discovery in Text, Machine learning

Italiano

Questo dottorato di ricerca contribuisce al dibattito accademico relativo al ruolo della conoscenza locale nei processi di *decision-making*. Il lavoro analizza se e come la conoscenza locale diffusa può contribuire, attraverso i nuovi paradigmi del *crowdsourcing* e del *social sensing*, alla creazione dei quadri di conoscenza in situazioni di rischio.

Dalla letteratura internazionale di riferimento emerge una evidente situazione di stallo intellettuale di lunga durata sulla conoscenza locale che potrebbe spiegare il motivo per cui gran parte della comunità scientifica tende a emarginare (se non ignorare) gli studi sulla conoscenza locale nel dominio del rischio.

Questo è dovuto, in parte a una emarginazione legata a vecchi retaggi culturali diffusi in diverse discipline scientifiche, che enfatizzano le conoscenze esperte, relegando e assegnando un ruolo secondario alle conoscenze locali; in parte alle criticità tecniche-metodologiche legate alla complessità intrinseca della conoscenza locale (in quanto conoscenza non strutturata) e quindi poco adatta ad essere trattata da un punto di vista computazionale.

Da quando negli ultimi anni l'attenzione dei governi ai processi di partecipazione collettiva alla *governance* del territorio è diventata più alta, l'interesse del mondo scientifico nei confronti della conoscenza locale e delle sue dinamiche, è aumentata. Un'attenzione agevolata anche dall'evoluzione dell'ICT e dei sistemi mobili, che ha stimolato la realizzazione di applicazioni di partecipazione diretta dei cittadini, che intervenendo con proprie istanze in arene virtuali, hanno contribuito alla costruzione di quadri di conoscenza locale affidandogli un ruolo di strumento di supporto alle decisioni. Strumenti conosciuti in letteratura come *Volunteered Geographic Information* (VGI) e *Public Geographic Information System* (PGIS).

L'utilizzo di questi strumenti si è esteso ben presto anche a modelli di partecipazione in altri ambiti, con la nascita di numerose piattaforme finalizzate alla segnalazione di situazioni di emergenze a seguito di disastri naturali, crisi umanitarie e conflitti politici con l'obiettivo di aiutare le popolazioni a fornire informazioni utili alla sopravvivenza (haiti, kenia, ecc.).

Tuttavia, questi sistemi hanno evidenziato alcuni limiti e criticità legati soprattutto alla gestione di informazioni non strutturate, archiviate in data base di grandi dimensioni. A queste si sono aggiunte tutte quelle informazioni, che trattano gli stessi argomenti provenienti dai *social network* pur essendo dislocate in conversazioni 'non coordinate'.

La criticità sono legate sia alla necessità di estrarre conoscenza da database in tempi rapidi, da distribuire, in varie modalità, a competenze esperte; sia alla facilitazione della comprensione e della condivisione di conoscenze più approfondite, soprattutto per quelli eventi che impattano vasti territori (come le crisi umanitarie), in cui si trovano culture, lingue e comportamenti molto eterogenei.

Il lavoro di tesi affronta queste criticità cercando di riannodare tra loro i temi della raccolta dell'informazione, dell'analisi, dell'interpretazione e della condivisione, attraverso metodologie differenti, con l'idea che strumenti attuali, che fanno uso di tecnologia VGI/PGIS, possano integrare al loro interno le metodologie qui sperimentate, in modo da ricoprire un ruolo strategico nel supportare le situazioni di emergenza nelle fasi successive ad un evento disastroso.

La dissertazione è strutturata in quattro parti. Dopo una breve introduzione (**Capitolo 1**) che fornisce una descrizione di *aims, objectives* e *research questions*, il **Capitolo 2** descrive il *Conceptual Background* in merito ai quattro principali domini rilevanti che sono stati analizzati.

La Research Design e la Methodology sono trattate nel **Capitolo 3 e Capitolo 4** e sono state affrontate con un lavoro empirico, i cui principali risultati sono: una classificazione attraverso modelli di machine learning, l'estrazione di concetti attraverso analisi testuali e la costruzione di modelli ontologici per la condivisione e l'interoperabilità dell'informazione estratta.

Infine nel **Capitolo 5** sono riportati: discussioni conclusioni e principali limitazioni del lavoro e future raccomandazioni di ricerca.

Il *Conceptual Background* (in **Capitolo 2**) include quattro temi:

- Il primo tema analizza le tematiche attinenti al rischio osservato nelle sue diverse forme e nelle sue diverse fasi, per comprenderne i campi di applicazione delle metodologie usate.
- Il secondo tema descrive la conoscenza, esaminata nei suoi concetti generali, cercando di fornire, al contempo, alcune definizioni per meglio argomentare aspetti specifici di questo lavoro, pur consapevoli che le sfumature tra i vari concetti non consentono una reale classificazione.
- Il terzo tema affronta l'analisi ontologica nelle sue diverse declinazioni.
- Il quarto tema è dedicato alla letteratura relativa ai principali paradigmi che si sono affermati negli ultimi anni relativamente alla raccolta di informazioni generate in modalità *crowdsourcing* (VGI e PGIS).

Il **Capitolo 3** illustra il lavoro empirico (i risultati sono riportati poi nel **Capitolo 4**), dipanato su tre temi metodologici differenti. i) Il primo tema affronta la raccolta delle informazioni attraverso una piattaforma basata su tecnologia PGIS. Tale piattaforma ha consentito di raccogliere informazioni geolocalizzate su argomenti generici sia in modalità diretta (attraverso le funzionalità della piattaforma), sia in modalità indiretta (attraverso operazioni di streaming da social network) su argomenti territoriali generici della città di Bari. Contestualmente, si è simulato un questionario on-line a riposta aperta che, dopo aver ipotizzato e descritto uno scenario catastrofico dovuto ad un terremoto nella città di Bari, chiedeva ai rispondenti di scrivere un messaggio breve (140 caratteri) di richiesta di aiuto o di offerta di aiuto. I dati acquisiti nelle due modalità (piattaforma Ushahidi e simulazione tramite questionario) sono stati successivamente uniti tra loro per creare un archivio unico eterogeneo (1300 messaggi). Un subset (400 messaggi) si riferisce a messaggi in situazioni di emergenza post terremoto. La creazione di un unico database si rende necessario per applicare e validare le procedure di classificazione dei messaggi in una fase successiva dell'elaborato. ii) Il secondo tema è quello delle applicazione di tecniche di *text mining* e di statistica lessicale per la creazione di matrici da analizzare e di tecniche di *machine learning* per verificare se alcuni modelli di classificazione (*Aaboost, Random Forest, Support Vector Machine SVM, Naï ve Bayes, Neural Network*) usati dalla letteratura corrente ed applicati ai testi fossero in grado di delimitare solo i messaggi appartenenti al dominio del terremoto. I risultati ottenuti hanno dimostrato l'affidabilità di questi sistemi, già affermati in letteratura soprattutto in discipline mediche ed economiche, evidenziando alte percentuali di affidabilità. iii) Il terzo tema, infine, fa riferimento all'analisi semantica e contestuale dei messaggi acquisiti con lo scopo di individuare, nel linguaggio naturale e in un contesto locale, il lessico (organizzato anche su due o tre parole) necessario a descrivere i seguenti concetti: bisogni necessari per la sopravvivenza e bisogni secondari, attori e ruoli ricoperti nell'ambito del dominio, aspetti geo-localizzativi di attori e bisogni. In particolare: a) I bisogni, sono stati declinati in cinque categorie (oggetti, servizi, comunicazioni, esigenze ambientali e funzioni), ognuna delle quali è stata suddivisa in due sottocategorie: 'necessarie per la sopravvivenza' e 'comfort'; b) Gli attori, divisi tra singole persone e gruppi di persone, per i quali sono stati individuati i possibili ruoli durante le fasi post evento (persone esposte, colpite, assistenti o semplici fornitori di informazioni). c) Le locazioni spaziali, sono state esaminate con un doppio approccio: uno, più scientifico, individuando tutti gli elementi che compongo una localizzazione rispetto ad un sistema di riferimento; l'altro, basandosi sulle modalità utilizzate nel linguaggio naturale per indicare un luogo (indirizzo, landmark, luoghi famosi, ecc.). Per entrambe è stato associato un modello di relazione spaziale per consentire di allocare ulteriori relazioni topologiche, di prossimità, direzionali e di distanza. I risultati di quest'analisi si sono concretizzati di fatto in tassonomie che successivamente attraverso un'analisi ontologica è stata allineata al modello ontologico fondazionale DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) con lo scopo di creare una ontologia condivisa per mitigare i rischi legati a successive traduzioni in sistemi di condivisioni in una logica di interoperabilità.

Infine, il **Capitolo 5** del lavoro di tesi ha delineato formulazioni conclusive e potenziali sviluppi futuri di ricerca. In particolare, il lavoro di tesi ha fornito un proprio contributo fondamentale ed ha colmato alcuni gap della letteratura internazionale circa il trattamento della conoscenza locale espressa in modo non strutturato attraverso forme di comunicazione testuali tramite *device* mobili. I nuovi strumenti e i nuovi modelli di analisi dei testi, seppur legati ai limiti del calcolo computazionale, riescono attraverso operazioni di destrutturazioni dei testi, a trasformare la conoscenza non strutturata in strutturata. Studi di questo tipo potrebbero rappresentare una base per la creazione di piattaforme integrate di raccolta, analisi e supporto alle decisioni. Rimangono aperte alcune questioni importanti circa l'affidabilità delle informazioni catturate dalla rete dove potrebbero annidarsi segnalazioni errate. A queste si aggiungono i falsi negativi derivanti dall'applicazione dei sistemi di classificazione, che per quanto abbiano un grado di affidabilità molto alto, rischiano di classificare come 'rumore' richieste di aiuto reali ed escluderle dal dominio. Il futuro di questa ricerca potrebbe dedicare a quel rumore ulteriori approfondimenti basati su tecniche più raffinate in grado di indagare su conoscenze latenti escluse dal dominio.

Acronyms and Abbreviations

CDD	Canadian Disaster Database
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
UGC	User-generated content
CGI	Contributed geographic information
VGI	Volunteered Geographic Information
SVM	Support Vector Machine
PGIS	Participatory Geographic Information System
PPGIS	Public Participation Geographic Information Systems
ICT	Information and Communication Technology
IT	Information Technology
DIKW	Data, Information, Knowledge, Wisdom
GPS	Global position system
VTC	Volunteer and Technical Communities
DH	Digital Humanitarians
SDI	Spatial Data Infrastructure
WHO	World Health Organisation
OSI	Open Source Initiative
UNDRO	United Nations Disaster Relief Office
UNISDR	United Nations Office for Disaster Risk Reduction
SMS	Short-Message Services
IAT	Inference Anchoring Theory
SWEET	Semantic Web for Earth and Environmental Terminology
SUMO	Suggested Upper Merged Ontology
FOAF	Friend of a Friend

NCGIA	National Center for Geographic Information and Analysis
GIS	Geographic Information System
PLA	Participatory Learning and Action
HOT	Humanitarian OpenStreetMap Team
TtT	Tweak the Tweet
HCDE	Department of Human Centered Design and Engineering
API	Application Programming Interface
KDT	Knowledge Discovery in Text
TDM	Text Data Mining
TM	Text Mining
NLP	Natural Language Processing
KDD	Knowledge Discovery in Database
IR	Information Retrieval
IE	Information Extraction
IM	Information Mining
NED	Named Entity Detection
NERD	Named Entity Recognition and Disambiguation
UIMA	Unstructured Information Management Architecture
GATE	General Architecture for Text Engineering
HXL	Humanitarian eXchange Language
MOAC	Management of A Crisis Vocabulary
SoKNOS	Service-oriented architectures supporting networks of public security
UTM	Universal Transverse Mercator

List of Figures

List of Tables

1 Introduction

Scientific research has always paid great attention to the issues and problems linked, in all its forms, to risk. Despite the relevant development of techno-scientific methodologies, the growing complexity of urban systems and socio-ecological interactions, also caused by the high degree of uncertainty due to territorial transformations and the awareness of global climate change, calls for new conceptual challenges.

Over time, the theories, models and methods to study and assess the above issues have used scientific knowledge and structured data in an attempt to build decision-support systems to prevent and mitigate the negative impacts caused by disaster events. These methods have favoured rational approaches based on deterministic models which have often highlighted the limitations caused by uncertainties and instability of the natural system (Gardner, 2002; Heisenberg, 1932; Hewitt, 1983; Prigogine, 1989). Uncertainties and instabilities have also caused the occurrence of environmental and socio-economic damages. Moreover, the persistence of damages has further supported the evidence against the *problem-solving* approach (as understood in terms of empirical approach based on scientific and technical knowledge) pioneered by Schön (1983). On the other hand, the same scholarship (*ibid.*) argued that a lack of problem setting, which takes into account complexity and uncertainties of the natural system (i.e. *reflection-in-action*), fully justifies the adoption of the problem-solving approach.

The above limitations are also due to computational problems for the availability of territorial data, high costs of databases, and lack of collaboration of public (and private) entities to supply and share data information. During the 1970s, Simon (1972), in his theory of bounded rationality, argued how a lack of information heavily affects the computation of problem-solving models in complex systems.

Fiori (2009), few decades later, built on Simon's legacy to highlight two critical points raised by the bounded rationality theory: On the one hand, the lack of useful information to implement mathematical modelling; on the other hand, an excess of (useless) information difficult to define and typify.

Recently, the diffusion of Social Networks is opening new research scenarios in the field of risk. Huge flows of Data are being made available from Social Networks: although they introduce further elements of complexity and uncertainty, they contribute to the diffusion of local knowledge and to unravel several complex system dynamics (Surowiecki, 2004).

The topics developed in this Thesis follow this direction to focus on local knowledge and knowledge diffusion in risk dynamics. The latter are importantly affected by opinions, attitudes and perceptions which, however, appear to be often neglected by formal analyses and thus require further investigations to better identify their contribution to the abovementioned risk dynamics.

1.1 Context

Over the last decades, natural disaster events (such as flooding, earthquakes, tsunamis and hurricanes) have caused extensive damage (to housing and infrastructure) and severe loss of lives in vast regions worldwide.

When these events occur, the role played by international organisations and cooperation to ease the management of emergencies and available resources are key aspects, which are widely debated in the literature (Kopena et al., 2008; Quarantelli, 2006; Reddy et al., 2009). Similarly, since the 1950s, scholars have highlighted the valuable contribution of citizens as active participants to handle emergency events. Lately, this contribution is closely linked to the diffusion of new *Information and Communication Technology* (ICT) (Simon et al., 2015; Whittaker et al., 2015) which has enabled a wider public participation in the decision-making process.

To understand the effects of ICT on consumer behaviour at global level, the following information on ICT use is provided (**Tab 1.1**):

Tab. 1.1 Current state of global digital statistics

PERIOD		TOTAL POPULATION	ACTIVE INTERNET USER	ACTIVE SOCIAL MEDIA USER	UNIQUE MOBILE USER	ACTIVE MOBILE SOCIAL USERS
JAN 2014	BILLION	7.095	2.485	1.857	n.d.	1.297
JAN 2015		7.210	3.010	2.078	3.649	1.685
JAN 2016		7.395	3.419	2.307	3.790	1.968
ANNUAL GROWTH		TOTAL POPULATION	ACTIVE INTERNET USER	ACTIVE SOCIAL MEDIA USER	UNIQUE MOBILE USER	ACTIVE MOBILE SOCIAL USERS
2014/2015	%	+ 1,6	+ 17,4	+ 10,6	-	+ 23,0
	m	+ 115	+ 525	+ 221	-	+ 388
2015/2016	%	+ 2,5	+ 12,0	+ 9,9	+ 3,7	+ 14,4
	m	+ 185	+ 409	+ 229	+ 141	+ 283

Source: (Kamp, 2016)

Tab. 1.1 reports key analytics on the use of digital technologies by the world population over the period 2014-2016. It is important to note that these statistics should also take into account the distribution of digital technologies across countries to understand the relative impact of the global Digital Divide[1]. **Fig. 1.1** shows a proxy representation of the distribution of digital technologies by relying on the percentage of individuals using the internet in the year 2015 in large regions or continents.

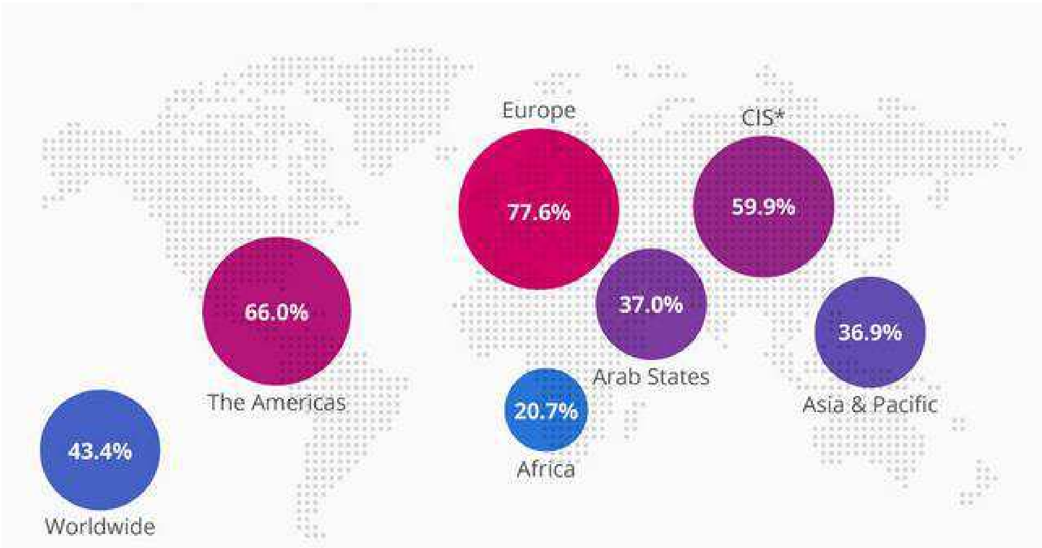
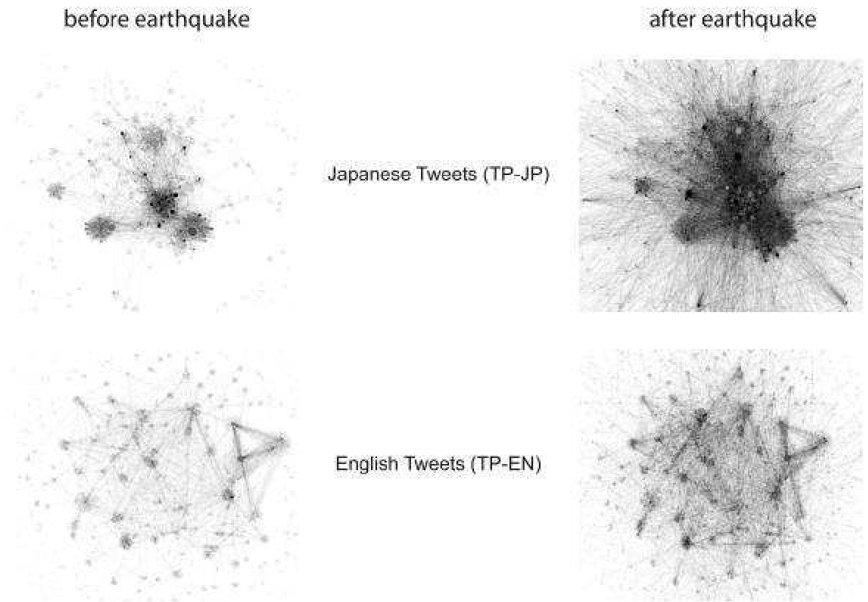


Fig. 1.1 Percentage of individuals using internet in 2015 - ITU United Nations specialized agency for information and communication technologies – ICTs

The large increase in the use of Social Networks in risk dynamics is a relatively recent aspect. The international literature considers several studies on the Haiti Earthquake of 2010, Tōhoku (Japan) earthquake and tsunami of 2011, Christchurch (New Zeland) earthquake of 2011, Queensland (Australia) flooding of 2012 and/or Haiyan (The Filippines) hurricane of 2013 to cite a few (Hughes and Palen, 2009; Vieweg, et al., 2010; Brun and Stieglitz, 2012). These studies have in common the analysis of *Tweets* posted on some Social Networks during and after the occurring of a disaster event. Also, text messages are analysed to shed light on their dynamics during rescue operations (Sakazi et al. ,2010; Neubig et al., 2011; Qu, et al., 2011).

[1] Although the digital divide is decreasing in recent years, it does pose, however, issues of unequal chances of survival in extreme situations (Laituri, 2008).

Fig. 1.2 shows the findings of a study by Lu and Brelsford (2014): the authors emphasise a communication stream across thousands of people on Twitter[2] soon after the 2010 earthquake in Japan. The work also draws attention to data exchange and availability on the web during the event, and the potential of exploring local knowledge associated to this data.



LANGUAGE	PERIOD	USER	TWEETS
Twitter Post-JP	Before	36893	42678
	after	36893	114247
Twitter Post-EN	before	36234	49095
	after	36234	64507

Fig. 1.2 Network structure and community evolution on Twitter: Before and after the earthquake in Japan 2010. Adapted from Lu and Brelsford (2014)

[2] Twitter is a social networking service where users post and read short 140-character messages called 'Tweets'. Registered users can post and read tweets, but those who are unregistered can only read them. Users access Twitter through the website interface, SMS or mobile device app (www.twitter.com).

The present Thesis analyses the communication streams occurred on Twitter before, during and after some of the above-mentioned disaster events. The contents of the text messages are analysed according to the event (e.g. earthquake, flooding) and phase types (ex-ante, in-itinere, ex-post).

1.2 Aims and Objectives

The new paradigms of communication and the openness to new forms of knowledge (giving due consideration to their limitations and drawbacks) that have been previously illustrated, seem unrelated topics. Based on this consideration, the present work aims to unravel latent relationships occurring across these main domains.

1.2.1 Aims

The consequences of a natural disaster can be large or small, direct or indirect, linked to main causes or side effects. These can be distant over space and time from the initial event, and depend on the contexts and their magnitude (Crozier et al., 2005).

Nonetheless, the content of communication streams in emergency situations can be sometimes useless and inappropriate. This is the case of an information overload as suggested by Simon, or rather because there are ambiguous lexical expressions (Cutler, 1983).

There exist some common elements that should accompany the study of risk in natural disaster events. These are all rooted in the concepts of ‘primary needs’ and ‘survival needs’ and are often neglected in the relevant literature (Kovács et al., 2007).

The main aim of this Thesis is to contribute to advancing the understanding of the above elements. These elements require a precise location in time and space to save human lives. Local knowledge arises from places, needs, contexts; the attitudes and feelings expressed with the use of digital technology when properly collected and analysed represent a fundamental contribution to the complex architecture of assessment of the intervention criteria during disaster events.

Social Networks are the main sources to achieve information in unstructured crowdsourcing systems and favour the construction of local knowledge. This form of communication is used by those individuals who directly or indirectly adopt digital technology to ask for, or provide assistance.

The stream of communication is often difficult to interpret. This is due to some forms of intrinsic complexity because data is not structured or rather semantic or linguistic ambiguities are present.

The re-organization into structured knowledge and sharing in open-source can contribute to solve complex risk dynamics and enrich knowledge frameworks used in risk forecasting and decision-support systems during management, coordination and rescue operations.

1.2.2 Objectives

The present Thesis deals with the following research questions. The first one is strictly linked to Simon's thought in terms of complex dynamics for the understanding of data: *How useful is the interpretation of text messages to grasp local knowledge in disaster response?*

To answer to this research question, there is the need: (i) to define and describe the study domain; (ii) to identify its main parts; and (iii) to assess the usefulness of these parts according to a scale of values.

From a scientific point of view, the use of ontological models, built through *bottom-up* (starting from exploratory data analysis) and *top-down* (based on the existing literature) approaches, appears the most appropriate strategy for the present work. The ontological models used in this Thesis can be summarised as follows: risks, needs, agents, and spatial location.

To discriminate between 'event' and 'non-event' data (text messages) for risk dynamics in disaster contexts, the present Thesis adopts the following criteria: 'event data' and 'non-event data' are considered as dichotomous variable. Event data is coded with a value of 1, otherwise 0. The former is selected; the latter is dismissed.

The second research question is the following: *Under what conditions, and with what limitations, may cognitive, predictive and onto-logical models facilitate the interpretation of text messages in the context of disaster response?*

To answer to this question, the present Thesis adopts two methodological approaches. The first one is a cognitive approach to understand the elements that contributes to define risks, needs, agents, and spatial location in a text message; the second one is an empirical approach, that provides data recognition and extraction to classify information in the selected ontological models.

The third research question asks: *How relevant is local knowledge to the governance of disaster response?* To answer to this research question the present thesis analyses through a cognitive and empirical approach the role of local knowledge in disaster response and its relevance for governance.

Similarly to other studies, the present work deals with and is applied to emergency or risk events in the context of natural disasters and assumes the existence of all necessary infrastructural and objective conditions for the exchange of text messages in Social Networks.

2 Conceptual Background

As illustrated in the previous chapter, the present Thesis's main aim is to analyse the problem of 'risk' in all its forms. The Thesis implements an interdisciplinary approach based on emerging paradigms in risk research, such as *local knowledge* and *social sensing*. As a result, the conceptual background is mainly concerned with the following three domains: Risk, Local Knowledge, and Social Sensing. The review also covers methodological aspects (including semantic, ontological, statistical, and computational issues) which are key to a holistic understanding of risk-related intellectual and operational challenges. **Fig. 2.1** shows the relevant domains and sub-domains of the scientific literature described in the present *Chapter*.

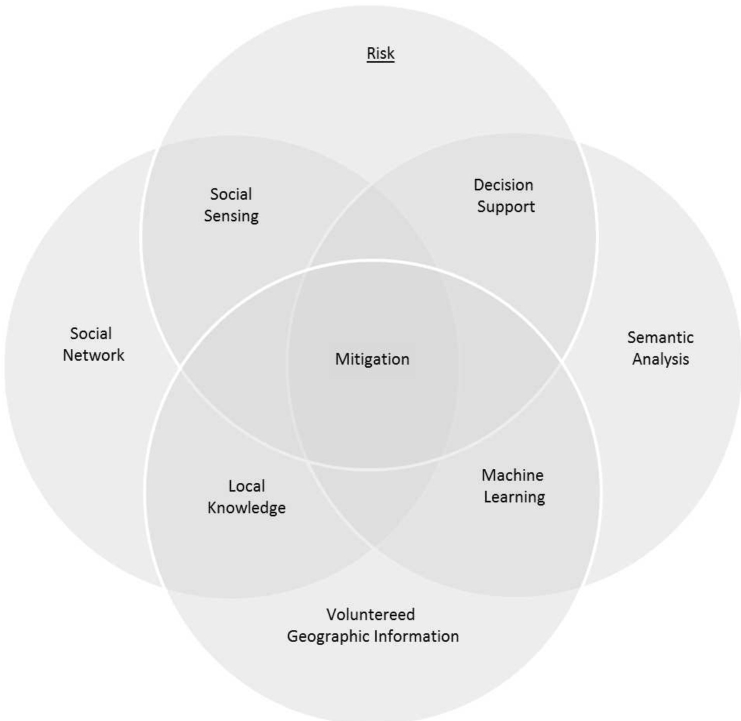


Fig. 2.1 Conceptual diagram of the research domains

To get a better understanding on the multifaceted conceptual background, the relevant literature has been codified in dedicated databases and specific text analysis techniques have been applied to better examine some interesting aspects of the evolution of the scientific literature. Thus, important strengths and drawbacks of the relevant bodies of literature have been highlighted.

Fig. 2.2 shows an application of the Fruchterman-Reingold algorithm to the relationships occurring across keywords assigned to the relevant scientific works. The algorithm simulates the physical law of particles in which the nodes represents the force (i.e. the magnitude of the keywords' frequency) and the edges represent the frequency between the nodes (i.e. the magnitude of the relationship between keywords/literature themes) (Kobourov, 2012).

The analysis highlighted a massive selection of papers from scientific journals in the field of computer science than other subject areas such as environmental and cognitive sciences and others.

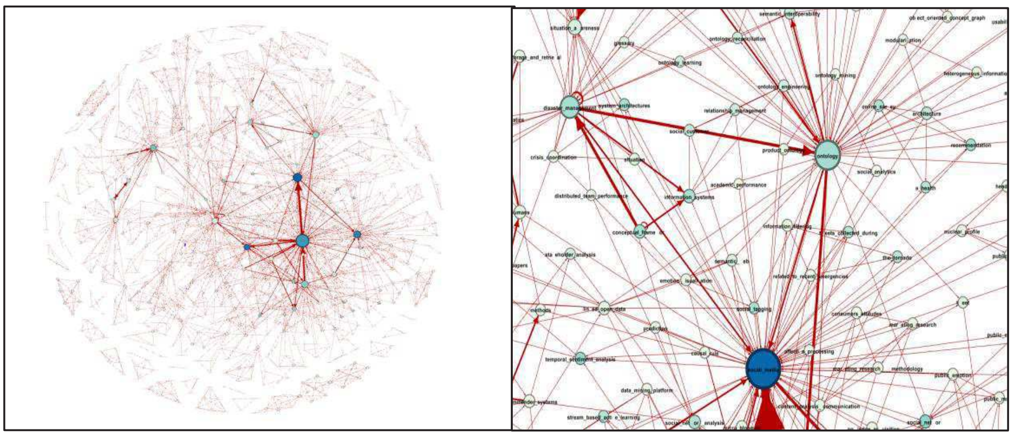


Fig. 2.2 Applications of the Fruchterman-Reingold algorithm - Keyword relationships in the relevant research domains selected for this study

2.1 Risk, Disaster, Catastrophe, Crisis

At first, the title of this section may appear ambiguous. The terms used may be considered as synonyms; or refer to different spatio-temporal dimensions; and/or may depend on the context in which they are placed. Also, the addition of other terms such as vulnerability, hazard, threat and resilience may further undermine the understanding and definition of the research domain. There has indeed been a long-lasting debate within the international community on lexicological and semantic aspects of these terms and concepts. Since 1917, Samuel Henry Prince in his psychological investigation of the explosion of a vessel at the Halifax port in Nova Scotia (Canada), started the debate on the use of concepts and terms related to risk in emergency situations (Smith, 2009). Between the 1960s and the 1970s, the debate involved other scientific disciplines aimed at analysing the physical aspects of disasters (*ibid.*). Different points of view, based on various theoretical backgrounds, have enriched over time the scientific debate with targeted studies on hazard, vulnerability, risk, and disasters (Cutter, 2001). Perry and Quarantelli (2005) argue that one of the main effects of the advancements over the last 20 years is a misunderstanding about the proper use of terms; by means of an example, the terms 'disaster' and 'catastrophe' are not often accurately picked up.

Following Cutter (2005), the present section focuses on the development of functional knowledge, based on the conceptual and theoretical understanding of the domains linked to 'risk' and their relationships – as applied to problem solving of real world cases. Consistently with an emerging trend, less attention is paid to the definition of 'what is a disaster event', when compared to advancing the understanding of why people and settlements are (or become) vulnerable to threats and unpredicted events through the study of the aspects affecting vulnerability and resilience.

The first part of the present section reflects on the literature on terms and definitions such as: i) risk and its components; ii) disaster; iii) catastrophe and its related classifications; and iv) forms of knowledge. In addition, the section sheds some light on recent events known as 'Humanitarian Crises' and 'Disease Outbreaks', which involve population dynamics such as migrations and epidemics.

2.1.1 Risk and Threat

In 1979, the United Nations Disaster Relief Office (UNDRO) publishes the Natural Disaster and Vulnerability Analysis report, which underlined major risk-related concepts and definitions in the following terms (UNDRO, 1979: 5):

- "***Natural Hazard***, *meaning the probability of occurrence, within a specific period of time in a given area, of a potentially damaging natural phenomenon.*
- ***Vulnerability***, *meaning the degree of loss to a given element at risk or set of such elements resulting from the occurrence of a natural phenomenon of a given magnitude and expressed on a scale from 0 (no damage) to 1 (total loss).*
- ***Elements at Risk***[3] *meaning the population, buildings and civil engineering works, economic activities, public services, utilities and infrastructure, etc., at risk in a given area.*
- ***Specific Risk*** *meaning the expected degree of loss due to a particular natural phenomenon and as a function of both natural hazard and vulnerability*".

Few years later, Varnes and IAEG (1984) defined (total) risk (R) for a given element at risk as the product of 3 factors – *hazard* (H), the *vulnerability* of the element at risk (V) and the presence of the *elements* at risk (E), that is:

$$R = H * V * E$$

Where H, V and E are still defined consistently with UNDRO (1979). The novelty of the work by Varnes and IAEG (1984) is to set, in their definition of (total) risk, an integrated quantitative assessment of risk which requires a broad-based knowledge from a wide range of disciplines.

In the following decades, the definition of risk remained almost unchanged. Godschalk (1991; 132) maintained that "*risk is the probability that a hazard will occur during a particular time period*" Crozier et al. (2005) highlighted that the term 'hazard' refers to those processes and situations (actions or non-actions) of potential damage which can cause losses or other

[3] Element at risk is also known in literature with the term exposure (Crozier and Thomas Glade, 2005).

damaging effects to human beings. The United Nations Office for Disaster Risk Reduction (UNISDR)[4] pointed out that "*risk is the combination of the probability of an event and its negative consequences*" (UNISDR, 2009: 25). The Office also defined 'hazard' as "*a dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage*" (*ibid.*: 17). In most cases, it is the specific source of risk that comes to define the hazard, such as natural hazards or hazards that are induced by human processes. The concept of hazard is applied to heterogeneous situations, where both natural aspects (such as earthquakes, landslides and floods) and economic aspects (e.g. capital investments in productive activities) may interact.

Generally, the term 'risk' can be seen as the combination of the probability that a threat is occurring and its anticipated consequences (Crozier et al., 2005). Following this argument, Alexander (2002) considers risk as a relationship between the hazard and the value of the elements at risk, given a certain degree of vulnerability (Fig. 2.3).

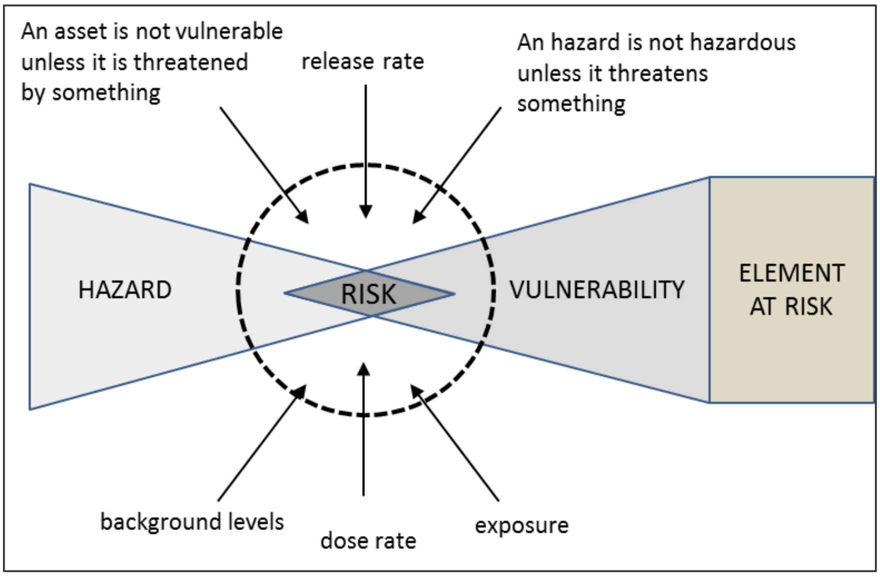


Fig. 2.3 Source: Conceptual relationship between hazard, elements at risk, vulnerability and risk (Alexander, 2002)

[4] Which evolved out of UNDRO, following a paradigmatic shift of focus from relief to disaster risk reduction – see https://www.unisdr.org/who-we-are/history.

The effects of a hazard event can be large or small, direct or indirect; linked to the main impact or to the occurrence of any further effects over time and space. They also depend on the context of the occurrence including particular elements and features which can provide a better understanding (i.e. quali-quantitative) of the risk assessment (Crozier et al., 2005).

More recently, the concept of resilience has been gaining momentum to stand for "*The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to, and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions*." (UNISDR, 2009: 24).

2.1.2 Disaster, Catastrophe and Crisis

A universally accepted definition of the term ‘disaster’ is yet to be agreed upon (PDM, 2002; Turner et al., 1997), given that the context which it refers to may vary, and the heterogeneity of scientific disciplines that it covers is vast (Shaluf, 2003). Similar conclusions may be drawn for the term ‘catastrophe’.

To explain the existing differences between the two concepts, Perry and Quarantelli (2005) shed light on the misunderstanding about the proper use of terms ‘disaster’ and ‘catastrophe (i.e. quantitative assessment) rather than semantic aspects. The latter generally require a more rigorous analysis before the terms disaster or catastrophe are cleared for use. Other scholars have it that the main difference between the terms disaster and catastrophe lies in the degree to which a community covers the (economic) costs of the damages, without external aid (Drabek, 1996): a **disaster** may accordingly be described as “*an event in which a community undergoes severe danger and incurs, or is threatened to incur, such losses to persons and/or property that the resources available within the community are exceeded. In disasters, resources from beyond the local jurisdiction, that is State or Federal level, are required to meet the disaster demands*”. Similarly, a **catastrophe** is defined as “*an event in which a society incurs, or is threatened to incur, such losses to persons and/or property that the entire society is affected and extraordinary resources and skills are required, some of which must come from other nations*” (*ibid.*).

Quarantelli (2005), on the contrary, argues that the difference between the two terms is more complex. The differences can be particularly evident at the organizational, community and social level. The two terms can also differ with regard to the scale of the phenomenon (i.e. the magnitude), dimension, population vulnerability, preparedness of public institutions and government, and efficacy of the decision-making processes before and during the event. In other words, the degree of the damage and preparedness significantly differ from a catastrophe to a ‘normal disaster’ (GAO, 2006) with serious societal impacts (Tobin and Montz, 1997).

A further concept to consider which is largely used in the literature is the term ‘accident’. **Accident** is a concept characterised by specific spatial boundaries. The effects of an accident involve a small group of people and the damages to buildings appear modest (Drabek, 1996; Dynes, 1998).

Over the last years, the term crisis has been widely analysed in the literature, but still there is a lack of a universally accepted definition of this term (Shaluf, 2003). Apparently, the meaning of **crisis** is similar to that of disaster or catastrophe. It represents a serious threat; it creates uncertainties, and it requires urgent interventions. The main feature of a crisis event is its molteplicity (Kim and Lee 2001, 502) which involves large groups, communities and nations (Quarantelli, 1998) such that it can also be traced over time. Farazmand (2001) maintains that it can be classified in:

i. natural disasters (e.g. hurricanes, floods);
ii. terroristic attacks (e.g. New York World Trade Center and Istanbul airport);
iii. nuclear accidents (Three-Mile Island, Chernobyl) or other technological risk-related events; and
iv. civil unrests, riots.

Fig. 2.4 summarises the concepts of Catastrophe, Crisis and Disaster. Each of these concepts is associated to that of Event, Hazard and Threat. In particular, the associations ‘Catastrophe-Event’, ‘Crisis-Event’ and ‘Disaster-Event’ refer to the occurrence of the event (part a); ‘Catastrophe-Threat’, ‘Crisis-Threat’ and ‘Disaster-Threat’ highlight potential damages (part b); and ‘Catastrophe-Hazard’, ‘Crisis- Hazard’ and ‘Disaster-Hazard’ refer to the probability with which the event might occur (part c).

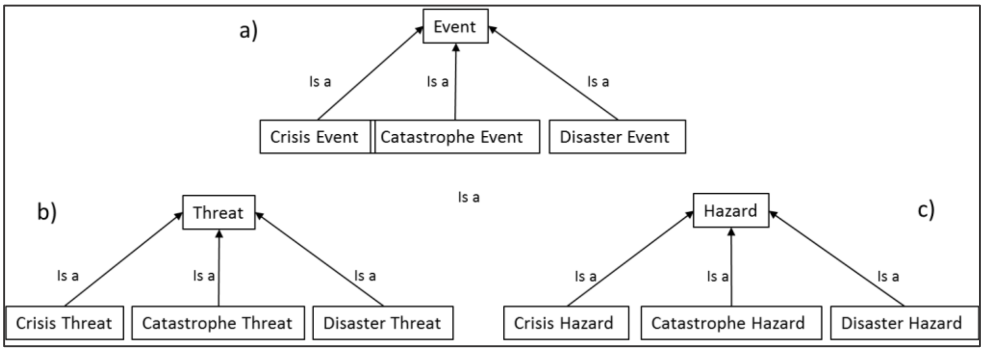


Fig. 2.4 Event, crisis, threat, disaster, hazard and catastrophe relationship

2.1.3 Event (Accident, Catastrophe, Crisis, Disaster) Phases

Various models – of the like of *disaster cycle*, *disaster model*, *disaster management* – are acknowledged in the literature (Platt, 1999; Poser et al., 2010; Kirillov et al., 2012). The novelty of these models is that they go beyond the mere description and analysis of events. A common element across these models is, in fact, the analysis of agents and organizations' behaviours and actions relatively to the event under investigation. These behaviours aim at positively affecting the resilience of societies and ecosystems through processes based on past experiences and/or immediate responses (Kreps, 1991).

The illustration of these models, as described below, is relevant to the present Thesis. In particular, the features of these models serve to support and justify the adoption of knowledge processes to affect resilience and vulnerable systems (Baharin et al., 2009) - as it will be dealt with in further details in **Chapter 3**.

Fig. 2.5 illustrates the main phases of a disaster-cycle model, which include Response, Recovery, Mitigation/Prevention, and Preparedness.

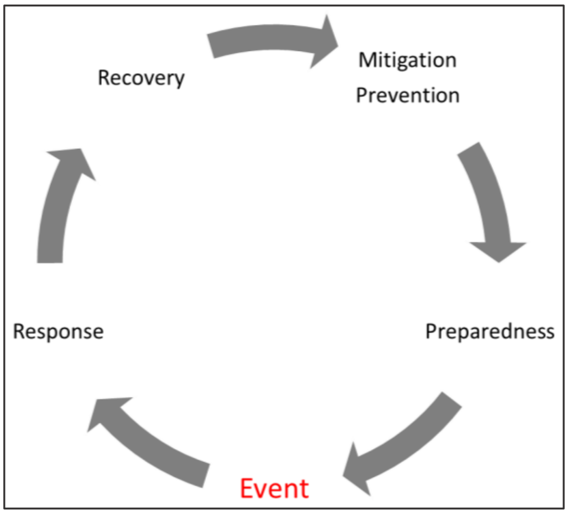


Fig. 2.5 Representation of event management life-cycle. Adapted from Platt (1999), Poser et al. (2010) and Kirillov (2012).

Lindell et al. (2006) extended the above representation to include specific aspects of risk management, risk assessment, and disaster emergency, and add further details to the analysis.

The above models and extensions appear on the whole adequate enough to help frame crisis/disaster management for the aims of the present work.

The first stage that follows a disaster event is a *response* phase. This consists of emergency operations to assist the affected population, including first aid (in terms of medical care and personal assistance), protection from looting actions and property (private and public) damage estimates (Perry, 1991).

Second, the *recovery* phase, which begins with the accident and ends when the community re-establishes socio-economic and political relations, aims at returning the affected community to the state before the event (UNISDR, 2009; Lindell, 2013).

Mitigation/Prevention actions precede the event and aim to reduce the probability of damage and its related effects. Generally, these actions are considered as normal prevention actions and include community safeguard, good practices of land use planning and environmental management, and building and infrastructures construction and maintenance (Lindell and Perry, 2000).

Finally, like mitigation, p*reparedness* precedes the event and aims to set up and test emergency plans; run adequate professional courses for emergency; and training to improve population response to the event (Tierney, 1993; Lindell, 2013).

2.1.4 Threat and the Interaction

The previous sections have defined some important notions linked to the concept of risk without specifying the types of threat and their nature. This section closes the gap and describes into details the threats of a typical crisis/disaster life cycle. The scientific debate considers several repositories and databases about information (i.e. time series) on disaster events. These are typically classified according to the nature and type of an event such as the EM-DAT[5], and the Canadian Disaster Database[6] (Liu et al., 2013).

The main drawback of these databases lies in the way information is classified. This is often not clear enough to detail a given event univocally, and available data tend to assume some forms of aggregation. The use of taxonomies and ontologies in the present Thesis has the advantage to go beyond the above issues. Ontologies, in particular, allow an understanding of their applicability in real world cases. The use of a recent taxonomy is illustrated below. The Centre for Risk Studies at the University of Cambridge has developed an interesting taxonomy of threats. The study, under review from October 2011 to March 2012, was completed in collaboration with a wide community of researchers (about 350). To date, the taxonomy on threats is released in its 2.1 version: this was partially adapted, for the purposes of this Thesis, and the outcome is shown in **Tab. 2.1** and **Tab. 2.2** (Coburn et al., 2014).

A 'threat' is defined as "*a potential cause of a socio-economic catastrophe that would threaten human and financial capital, damage assets, and disrupt the systems that support our society, with ability to have international or global impact*" (ibid.).

---

[5] EM-DAT - EM-DAT was created with the initial support of the World Health Organisation (WHO) and the Belgian Government. The main objective of the database is to serve the purposes of humanitarian action at national and international levels. EM-DAT contains essential core data on the occurrence and effects of over 22,000 mass disasters in the world from 1900 to the present day. The database is compiled from various sources, including UN agencies, non-governmental organisations, insurance companies, research institutes and press agencies. The database may be accessed at http://www.emdat.be.

[6] The Canadian Disaster Database (CDD) contains detailed disaster information on more than 1,000 natural, technological and conflict events (excluding war) that have happened since 1900 at home or abroad and that have directly affected Canadians. The CDD tracks 'significant disaster events' which conform to the Emergency Management Framework for Canada definition of a 'disaster'.

Tab. 2.1 Taxonomy of macro-threats (i)

Natural Catastrophe				
Naturally occurring phenomena causing widespread damage and disruption				
Earthquake	Windstorm	Tsunami	Flood	Volcanic Eruption
Seismic fault rupture causes high levels of damage to infrastructure of a major populated area	Hurricane/typhoon/cyclone wind system makes landfall onto a major populated area; European-type windstorm system, large scale, fast-moving, gale force wind speeds	Coastal impact of a tidal wave, caused by offshore earthquake, marine landslide, or meteorite in the sea,	River Flood from high rainfall/sudden water release across one or more river systems; Coastal Flood from sea surge caused by low pressure weather systems, exceptional tides and extreme winds	Ash, pyroclastic hot gasses, lava, and lahar-triggered mudflows cause localized destruction and regional disruption
Climatic Catastrophe				
Climatic anomalies or extremes causing severe and unusual weather conditions				
Drought	Freeze Event	Heatwave		
Extended period of below-average precipitation	Extended period of below-average temperatures	Extended period of above-average temperatures		
Environmental Catastrophe				
Crises leading to significant and widespread change to environmental or ecological equilibriums				
Sea Level Rise	Ocean System Change	Atmospheric System Change	Pollution Event	Wildfire
Thermal expansion of the oceans or sudden ice shield melt changes coastline geography	Sudden switch in the circulatory systems of the ocean, such as the Gulf Stream, caused by salination or thermal changes, causes regional climatic change	Rapid or sustained periods of change in patterns of meteorological circulation, such as jet stream, causes regional climatic change	Spillage or major release of toxic chemicals into land or sea systems that causes environmental destruction	Uncontrolled inferno, enhanced by natural landscape and environmental factors
Technological Catastrophe				
Accidental or deliberate industrial events affecting local and global stakeholders				
Nuclear Meltdown	Industrial Accident	Infrastructure Failure	Technological Accident	Cyber-Catastrophe
Major core meltdown of a nuclear power station, causing radioactive fallout over a large area of population and economic and agricultural productivity	Fire, explosion or release of toxic chemicals from an industrial complex, storage facility or during transportation	Blackouts in the electricity supply network and other systems failures due to accidents and technical breakdowns	New technological advance proves to have unexpected societal effects and causes disruption or harm to human populations	Computer networks, communications and information technology systems destabilized by computer virus, hacking, denial of service attacks or other cyber-security issues
Geopolitical Conflict				
Military engagements and diplomatic crises between nations with global implications				
Conventional War	Asymmetric War	Nuclear War	Civil War	External Force
The engagement of two or more nations in military conflict, using conventional weapons to target military infrastructure and invade/defend sovereignty	Military action, insurgency and violent resistance carried out between combatants of significantly different power, resources, and interests	Military Conflict pursued using nuclear weapons	Internal conflict within a country, including wars of succession and coups d'etat	Blockades, No-Fly zones, missile attack or other military action by external forces to prevent national authorities pursuing internal policies deemed harmful or repugnant
Political Violence				
Acts or threats of violence by individuals or groups for political ends				
Terrorism	Separatism	Civil Disorder	Assassination	Organized Crime
Politically-motivated single or coordinated attack(s) to inflict societal and/or economic fear and disruption	Sustained campaign of violence for regional independence	Riots and civil disobedience, through to uprisings and revolutions	Assassination of a major political leader	Crime waves, Campaigns of criminal extortion, piracy, or mass illegal activities that debilitates commercial activity

Source: Cambridge Centre for Risk Studies (Coburn et al., 2014)

Tab. 2.2 Taxonomy of macro-threats (ii)

Humanitarian Crisis				
Impact of conditions on mass populations of people				
Famine	Water Supply Failure	Refugee Crisis	Welfare System Failure	
A large population suffers failure of their food supply, food distribution, or agricultural production system	A large population suffers failure of their water supply due to water resource conflicts, river diversion, aquifer depletion, or other cause	Mass population movements cause instability and collapse of social infrastructure in the areas newly populated and depopulated	Collapse of pension schemes, health programs and social security systems leading to deprivation and hardship for dependents. Breakdowns triggered by underfunding, and imbalances e.g. ageing populations	

Financial Shock				
Events in the financial system causing short-run fluctuations and/or significant changes in long-run economic growth				
Asset Bubble	Financial Irregularity	Bank Run	Sovereign Default	Market Crash
Pricing inflation and sudden collapse for a major sector or asset class	Corporate or accounting fraud; Rogue trading; Ponzi schemes; or other major irregularities	Bank failure; Credit default for major banks, banking system or market participant	Debt default, currency devaluation or government failure and/or change	Extreme correlated mass movement of share prices , possibly driven by information or perception about economic fundamentals

Trade Dispute				
Events causing widespread change or disruption to international trading conditions				
Labour Dispute	Trade Sanctions	Tariff Wars	Nationalization	Cartel Pressure
Strikes, mass refusal of employees to work, or picketing by aggrieved workforce to prevent commercial activity	Country-to-country trade embargos denying entry or passage of commercial goods and services	Protectionism through the imposition of taxation a particular set of goods or services	Sovereign appropriation of foreign-owned assets in that country	Trading bloc of suppliers applies pricing or supply pressures

Disease Outbreak				
Disease outbreaks affecting humans, animals and/or plants				
Human Epidemic	Animal Epidemic	Plant Epidemic		
Influenza pandemics, emerging infectious diseases and re-emergent disease epidemics that cause death and illness in human populations	Diseases in animals that cripple agricultural production of meat and poultry or destroy wildlife	Diseases in plants that impact food production in many agricultural areas or cause destruction of the ecological environment.		

Externality				
Threats originating from outside the earth's atmosphere including astronomical objects and space weather				
Meteorite	Solar Storm			
Ground impact of meteors that cause localized destruction, and dust clouds capable of causing periods of ash winter	Solar flare activity that can impact satellites, communication technology, power distribution systems and other infrastructure			

Source: Cambridge Centre for Risk Studies (Coburn et al., 2014)

In their seminal work, Coburn and co-workers (2014) sort threats into 11 classes and 44 sub-classes, provide related descriptions, and include recently acknowledged threats such as Humanitarian Crises, Geopolitical Conflicts, and Financial Shocks. The study over-takes the typical classification of threats in *natural* (e.g. floods, tsunamis, earthquakes) and *anthropic* (e.g. technological, conflicts, terroristic) events. An interesting aspect of this work is also the analysis of relationships among threats, which is formalised in a (weighted) correlation matrix (i.e. cause/effect of threats) as shown in **Tab.** *2.3* (Coburn et al., 2013).

Tab. 2.3 Correlation and causation dependencies of threat categories

			Consequence											
			1	2	3	4	5	6	7	8	9	10	11	12
			Financial Shock	**Trade Dispute**	**Geopolitical Conflict**	**Political Violence**	**Natural Catastrophe**	**Climatic Catastrophe**	**Environmental Catastrophe**	**Technological Catastrophe**	**Disease Outbreak**	**Humanitarian Crisis**	**Externality**	**Other**
Primary Trigger	1	**Financial Shock**	4	3	2	2	1	1	1	1	1	2	1	1
	2	**Trade Dispute**	3	4	2	3	1	1	1	1	1	1	1	1
	3	**Geopolitical Conflict**	3	2	4	3	1	1	1	1	1	2	1	1
	4	**Political Violence**	2	2	3	4	0	0	0	3	3	2	1	1
	5	**Natural Catastrophe**	2	2	2	1	4	2	3	3	2	2	1	1
	6	**Climatic Catastrophe**	3	2	3	2	3	4	3	2	2	3	1	1
	7	**Environmental Catastrophe**	3	2	2	2	3	3	4	2	2	2	1	1
	8	**Technological Catastrophe**	2	2	2	2	2	2	0	4	1	1	1	1
	9	**Disease Outbreak**	3	2	1	1	1	1	1	2	4	2	1	1
	10	**Humanitarian Crisis**	2	2	3	3	1	1	1	1	2	4	1	1
	11	**Externality**	3	2	2	1	3	3	3	3	2	2	4	1
	12	**Other**	1	1	1	1	1	1	1	1	1	1	1	4

Legend

0. The two threat types are uncorrelated, and if they occurred coincidentally, their consequences would be broadly the same as if they occurred independently
1. No mechanism for this threat to directly cause an event of the second threat type, but the consequences of a coincidental second event shortly afterwards would be made significantly worse, for example because resources would be already committed and abilities to respond and contain would be weakened
2. There is some potential for an event to contribute to the causal mechanisms that would trigger the occurrence of an event of the second type
3. An event of this type potentially can directly trigger an event of the second type
4. An event of this type potentially can directly trigger another sub-category of threat within the same threat category.

Source: Cambridge Centre for Risk Studies (Coburn et al., 2014)

The main implication of the analysis shown in Tab. *2.3* is that the event is (almost) never classified as an isolated event. There exists a *domino* effect which may cause much more damage than the one brought about by the primary threat.

Further computations of the threats matrix are illustrated as radar charts reported in Fig. 2.6-Fig. 2.16. These highlight the main effects of each *primary trigger*. It is important to notice that climatic and environmental catastrophes are the main threats in terms of capacity to trigger ecological, economic and financial domino effects on the affected population and ecosystems.

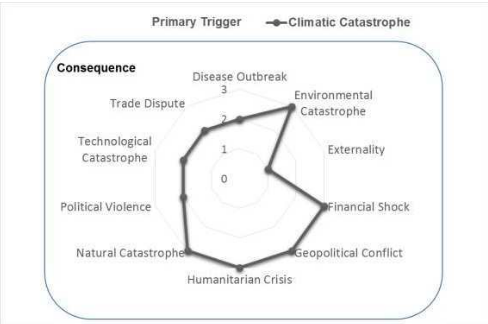


Fig. 2.6 Primary trigger: Climatic catastrophe

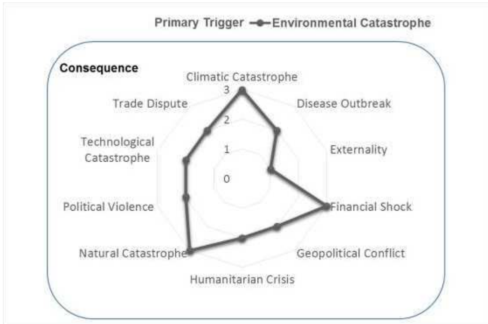


Fig. 2.7 Primary trigger: Environmental catastrophe

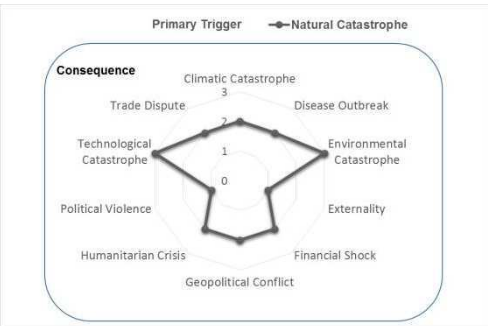


Fig. 2.8 Primary trigger: Natural catastrophe

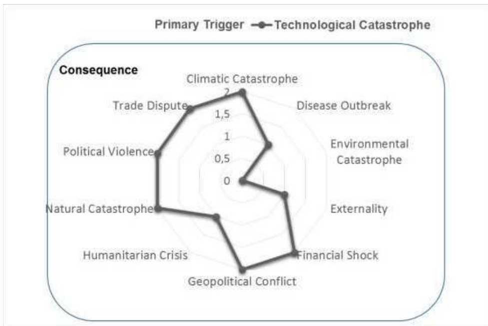


Fig. 2.9 Primary trigger: Technological catastrophe

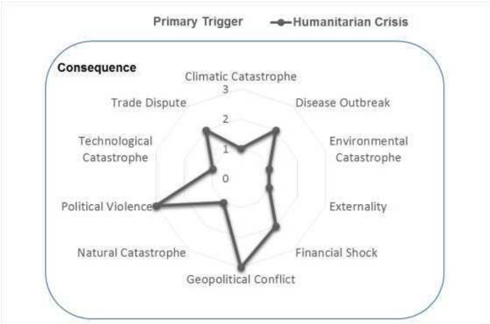


Fig. 2.10 Primary trigger: Humanitarian crisis

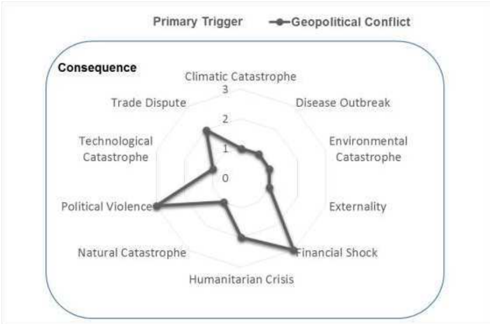


Fig. 2.11 Primary trigger: Geopolitical conflict

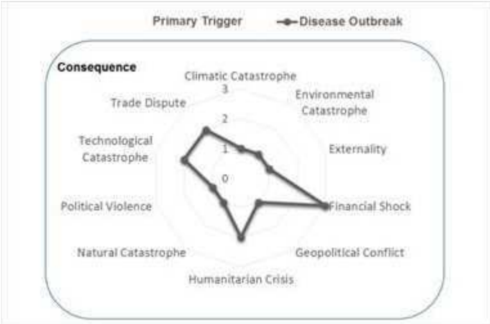


Fig. 2.12 Primary trigger: Disease outbreak

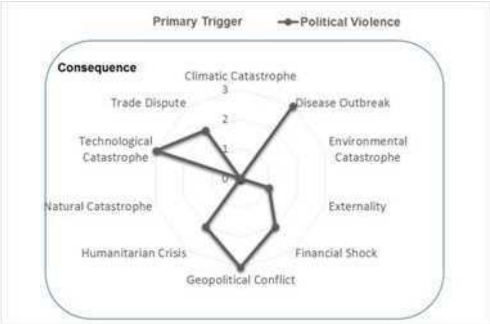


Fig. 2.13 Primary trigger: Political violence

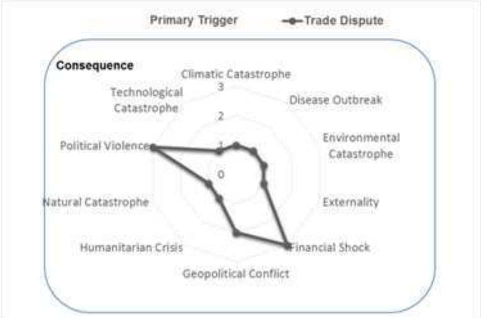


Fig. 2.14 Primary trigger: Trade dispute

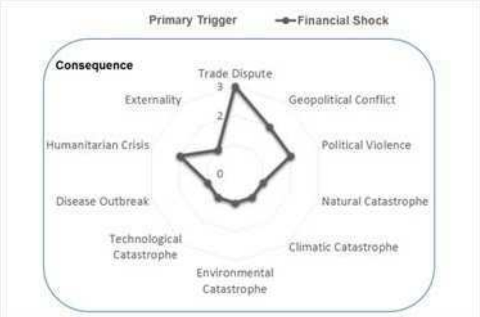


Fig. 2.15 Primary trigger: Financial shock

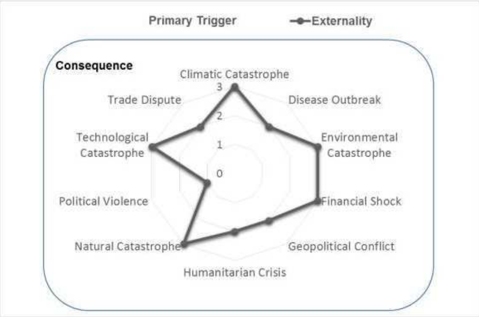


Fig. 2.16 Primary trigger: Externality

2.2 Knowledge and Risk

To define knowledge and cognitive processes is a challenge that, from Plato's theory of innate ideas to Karl Popper's stress on falsifiability, has fascinated all generations of philosophers and triggered many epistemological debates.

For the purposes of the present Thesis, and in line with Alavi and Leidner (2010), it is interesting to shed light on the views of the main authors of organizational theories of knowledge, mostly explored within the boundaries of *computer science* literature. This is because the methodological and applied parts of the present Thesis describe and use ontological models and machine learning algorithms.

2.2.1 Knowledge Hierarchy

For several decades, *Information Technology* (IT) has been adopting the 'knowledge hierarchy' (**Fig.** *2.17*) paradigm, which is best known as the Data, Information, Knowledge, Wisdom (DIKW) model, developed by Ackoff (1989). DIKW is adopted for defining data, information, knowledge and wisdom of computer science literature (Rowley, 2007).

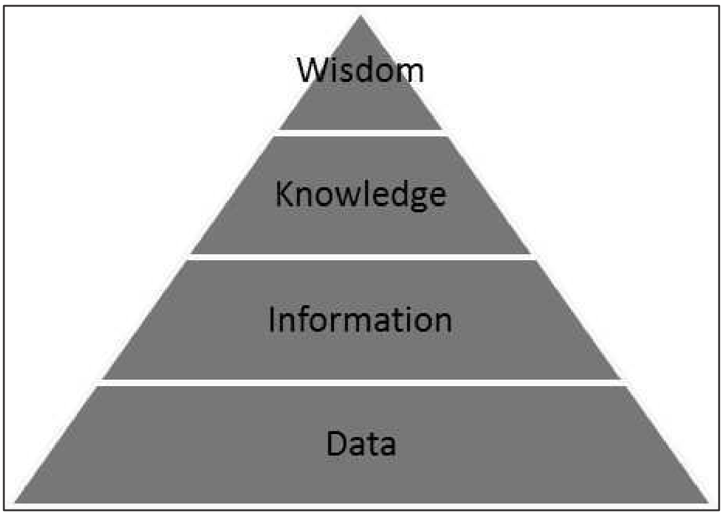


Fig. 2.17 source : The DIKW pyramid Ackoff (1989)

Sharma (2004) and Bernstein (1989) recognise, based on the seminal work of Ackoff (1989), the importance of the *knowledge hierarchy* – at least in the context of knowledge management. Nonetheless, over the last fifty years, computer science literature seems to have favoured *Data* and *Information*, although *Wisdom* and *Knowledge* are considered as higher-level concepts (Carlisle and Pinn, 2015). This might be traced to the limited reach of programming across the layers of the hierarchy (**Fig. 2.18**) (Awad and Ghaziri, 2004; Rowley, 2007).

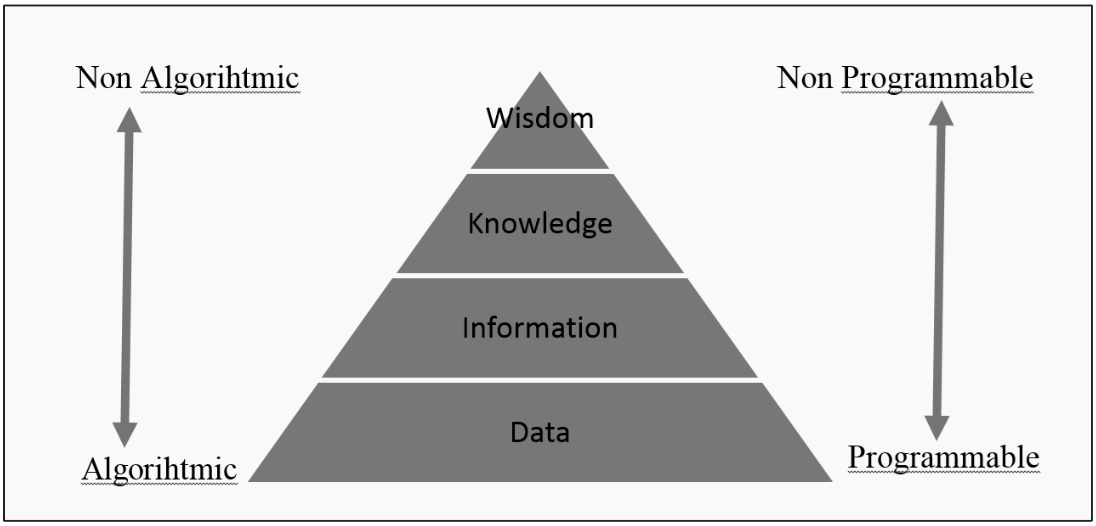


Fig. 2.18 The DIKW pyramid adapted from Awad and Ghaziri (2004).

In the present work, Knowledge and Wisdom play an important role. In line with Tronti (2014), these are seen as part of a collective process where individual knowledge and scientific research contribute to reach socially desirable goals.

The aim of the present section and sub-sections is to provide a scientific background on the role that the concept of knowledge plays in the overall theoretical approach, and to single out the aspects of Knowledge and Wisdom that might prove more key to support advancements in disaster response. Under these premises, the concepts of Data and Information will play a secondary role (Aven, 2013).

2.2.2 Data, Information, Knowledge, Wisdom

There is a consensus within the international scientific community about the diversity of, and the complex relationships between, the concepts of DIKW. This consensus is oriented to clarify meanings and assign hierarchical relationships (Tuomi, 2000; Zins, 2007).

A common goal seems the search for defining DIKW and highlighting divergences across meanings, which, at first, appear similar to each other, but at the same time, originate ambiguities and/or mis-interpretations (Hoppe et al., 2011).

The works by Rowley (2007), Baškarada and Koronios (2013), and Liew (2013) are considered relevant reviews of DIKW definitions. To understand the heterogeneity of the DIKW definitions in the literature and the ambiguities and conflicts developed over time, **Tab. 2.4** reports a synthesis drawn from Baškarada and Koronios (2013), who adapted it from Rowley (2007).

Defining Data

According to Blair (2002) and Zins (2007) *Data* are defined as *primitive symbolic entities*; or rather as a sequence of symbols (Jankowski and Skowron, 2007; Hoppe et al., 2011); alternatively, Jessup and Valacich (2003) and Groff and Jones (2003) refer to *Data* as symbols with no meaning because of a lack of a contextual space. In other words, "*Data are physical signs. They have no meaning because they reside outside of a human mind*" (Baškarada and Koronios, 2013:11).

Defining Information

There is a broad consensus on the interpretation of the concept of *Information*. Recent literature (Nonaka and Takeuchi, 1995; Turban et al., 2005; Laudon and Laudon, 2006;

Boddy et al., 2008) considers *Information* as an aggregate of organized (or modelled) data represented in a meaningful way for individuals. Zins (2007) and Groff and Jones (2003) also argue that the contextual space is an important aspect which provides meaning to data and retrieve/processes information.

"*Information (or meaning) emerges through cognitive processing of data*" (Baškarada and Koronios, 2013:13).

Tab. 2.4 Defining ambiguous and/or conflicting definitions developed for the Data, Information, Knowledge, Wisdom concepts.

Wisdom	Wisdom is accumulated knowledge, which allows you to understand how to apply concepts from one domain to new situations or problems (Jessup and Valacich, 2003).
	Wisdom is the highest level of abstraction, with vision foresight and the ability to see beyond the horizon (Awad and Ghaziri, 2004, p. 40).
	Wisdom is the ability to act critically or practically in any given situation. It is based on ethical judgement related to an individual's belief system (Jashapara, 2005, pp. 17-18).
Knowledge	Knowledge is the combination of data and information, to which is added expert opinion, skills, and experience, to result in a valuable asset which can be used to aid decision making (Chaffey and Wood, 2005, p. 223).
	Knowledge is data and/or information that have been organised and processed to convey understanding, experience, accumulated learning, and expertise as they apply to a current problem or activity (Turban et al., 2005, p. 38).
	Knowledge builds on information that is extracted from data ... While data is a property of things, knowledge is a property of people that predisposes them to act in a particular way (Boddy et al., 2005, p. 9).
Information	Information is data which adds value to the understanding of a subject (Chaffey and Wood, 2005, p. 233).
	Information is data that have been shaped into a form that is meaningful and useful to human beings (Laudon and Laudon, 2006, p. 13).
	Information is an aggregation of data that makes decision making easier (Awad and Ghaziri, 2004, p. 36).
Data	Data has no meaning or value because it is without context and interpretation (Jessup and Valacich, 2003, Bocij et al., 2003, Groff and Jones, 2003).
	Data are discrete, objective facts or observations, which are unorganised and unprocessed, and do not convey any specific meaning (Awad and Ghaziri, 2004, Chaffey and Wood, 2005, Pearlson and Saunders, 2004, Bocij et al., 2003).
	Data items are an elementary and recorded description of things, events, activities and transactions (Laudon and Laudon, 2006, Turban et al., 2005, Boddy et al., 2005).

Source: adapted from (Rowley, 2007: 170-174), as quoted in Baškarada and Koronios (2013).

Defining Knowledge

The definition of *Knowledge* appears more ambiguous and contentious, when compared to those of *Data* and *Information* (Barnes, 2002). Some authors also refer to specific concepts such as 'mix of experiences' (Davenport and Prusak, 1998) values, insights, and contextual information, often recorded in various forms (e.g. documents, processes, practices and norms). Other works, in the DIKW context, define *Knowledge* as a combination of data and information coupled with expert opinions, particular abilities and experience (Chaffey, 2005, Turban, 2005), or collective scrutiny "*Knowledge constitutes a person's beliefs which have been socially judged to be true"* (Baškarada and Koronios, 2013:12).

Defining Wisdom

Compared to Knowledge, the concept of *Wisdom* appears more complex to define. First, there exist several definitions of the term; second, the existing definitions seem at odds with each other (Hoppe et al., 2011). Ackoff (1989) argues that *Wisdom* is a process that makes use of *Knowledge* to answer to difficult questions. Zeleny (2006), on the other hand, supports the idea that to answer to difficult questions it is sufficient to get available information from experts. Wise people are able to explain the reason why something should be done and from which the following metaphors emerge: *Information = 'know–what' - 'knowledge' - 'know–how' - wisdom = 'know–why'* (*ibid.*). Along the same line of reasoning, Rowley (2007) stresses the role of socio-ethical principles to legitimate a wise behavior. Hoppe et al. (2011) analyse the concept of Wisdom in various domains such as Philosophy, Psychology, Neurology and Computer Science and provide some contrasting views. Finally, Baškarada and Koronios (2013) build on their definition of knowledge to assert that "*Wisdom constitutes a person's normative judgements which have been socially judged to be desirable*" (*ibid.*)

Relationship between Data, Information, Knowledge and Wisdom

The DIKW model, under the knowledge hierarchy representation by Ackoff (1989), appears more complex to explain when one moves to the top layers. According to Stanmark

(2002), the representation of the DIKW model is probably misleading (see **Fig. 2.18**). At first, the relationships occurring across the concepts appear linear and constant to each other. In fact, the transition from *Data* to *Information* seems to be fairly easy. On the other hand, the transitions 'Information to Knowledge' and 'Knowledge to Wisdom' are unclear. Stanmark (*ibid.*) criticises the asymmetry of the transitions: these would not be bi-directional but they rather proceed one way only.

This detail, which might have been apparently trivial a few decades ago, is becoming ever more important – as we advance further into the era of Big Data[7]. Under certain aspects, it is linked to one of the initial research questions of the present Thesis: *How useful is the interpretation of text messages to grasp local knowledge in disaster response?* In other words, how is it possible, through automated processes, to transform non-structured data into knowledge (i.e., real-time text retrieval and analysis during disaster events). Is this transformation useful to understand knowledge in disaster response?

Understanding a verbal or written expression may represent knowledge or wisdom when the agent is a person. When the agent is a computer, the task of understanding that verbal or written expression appears more complex to perform. The computer is not able to understand the text as a whole: hence, it should perform a dual step:

- the first step is a form of reverse engineering to deconstruct the text in search of the information and data contained in the text;
- the second one, through specific inferential methods on the extracted data, is a reconstruction of Knowledge.

Should the above argument be true, then data and information cannot be considered hierarchically inferior to information and knowledge. Data and Information represent the backbone to information and knowledge and are the variables that a computer processes (Tuomi, 2000).

Then, a further issue arises, as Frike (2008) questions the DIKW model: the Data and Information obtained through an inductive process may be unreliable due to an error component which is present during the processing phase. This error would then alter, as a loop,

[7] 'Big data' is taken to mean the booming increase in data quantity, which become difficult to store, elaborate and analyse by means of traditional database technologies (Hashem et al., 2015).

the knowledge extraction process. A final issue is that of ignorance, misinformation, stupidity and belief: should these belong to the DIKW hierarchy or require a different model? (Bernstein, 2011). The answers to the previous issues might somehow emerge following interdisciplinary dialogue with those branches of pedagogy and psychology that deal with *critical incidents*, i.e., problematic experiential events (Nuzzaci, 2011). According to these fields of studies, errors, ignorance, misinformation stupidity and belief may be considered as critical learning tools, and should therefore be treated as knowledge.

2.2.3 Forms of Knowledge

Over time, researchers have treated and compared different types of knowledge, although they have been always aware that there is no borderline that can be used to sort knowledge into clear-cut classes. Nonetheless, besides the definitions provided in the previous section, knowledge may be articulated under several forms.

To this end, the study by Raymond at al. (2010), although originally aimed at reviewing the different understanding of knowledge in the field of environmental management, covers several aspects that are relevant risk studies. **Fig. 2.19** summarises the main steps of this review. There exist four types of knowledge, which are as follows:

a) Local and Generalised knowledge which focuses on spatial aspects;
b) Formal and Informal knowledge which focuses on legal/administrative aspects;
c) Novice and Expert knowledge which focuses on the dimension/background aspects of knowledge; and
d) Tacit, Implicit and Explicit knowledge which focuses on the degree of knowledge expression.

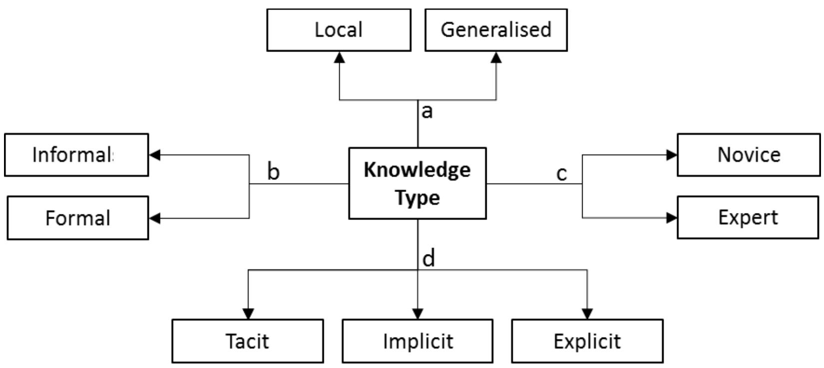


Fig. 2.19 Dimensions of knowledge types derived from the environmental management literature. Adapted from (Raymond, 2010)

For the purposes of the present Thesis, Local, Informal, Novice and Tacit knowledge play an important role since they are strongly rooted in the personal (and emotional) sphere of an individual, and hence difficult to formalise, share and transfer. The remaining forms of knowledge such as Generalised, Formal, Expert and Explicit are generally easier to understand and analyse because they incorporate words and numbers, scientific formulas, specifications, manuals (Nonaka, 2001; Fazey, 2006; Olaide and Omolere, 2013).

It should be stressed that it is possible to have combinations between two or more forms of knowledge. For example, knowledge can simultaneously be treated as Local, Expert and Informal. Raymond (2010) argues instead on the forms of knowledge such as local, scientific and hybrid. The author specifies the existence of two sub-categories of local knowledge: 'Personal Experience' and 'Traditional Cultural Rules and Norm'. These are seen as the main drivers of specific local knowledge (**Fig. 2.20**). Over the last years, it is this level of disaggregation that has received great attention in the international literature. A detailed description of each relevant type of knowledge classified by Raymond (*ibid.*) is provided in the following paragraphs. For the purposes of the present Thesis, the term 'Local knowledge' is used to include all such forms of knowledge.

Personal

This is a form of knowledge based on (individual) experiential processes that can be distinguished into different forms (Polany, 1962; Fazey et al., 2006).

Lay

It refers to informal knowledge which reflects the most common interpretation given by individuals about a specific situation (Jones, 1985; Halfacree, 1995; Hansen at al. 2003).

Local or situated

This is a form of knowledge based on the understanding of local phenomena. It is often used to distinguish it from external expert knowledge. The latter, although rich in technical experience and expertise, lacks of local views and nuances (Smith, 2001; Robertson e McGee, 2003; Kettle, 2014).

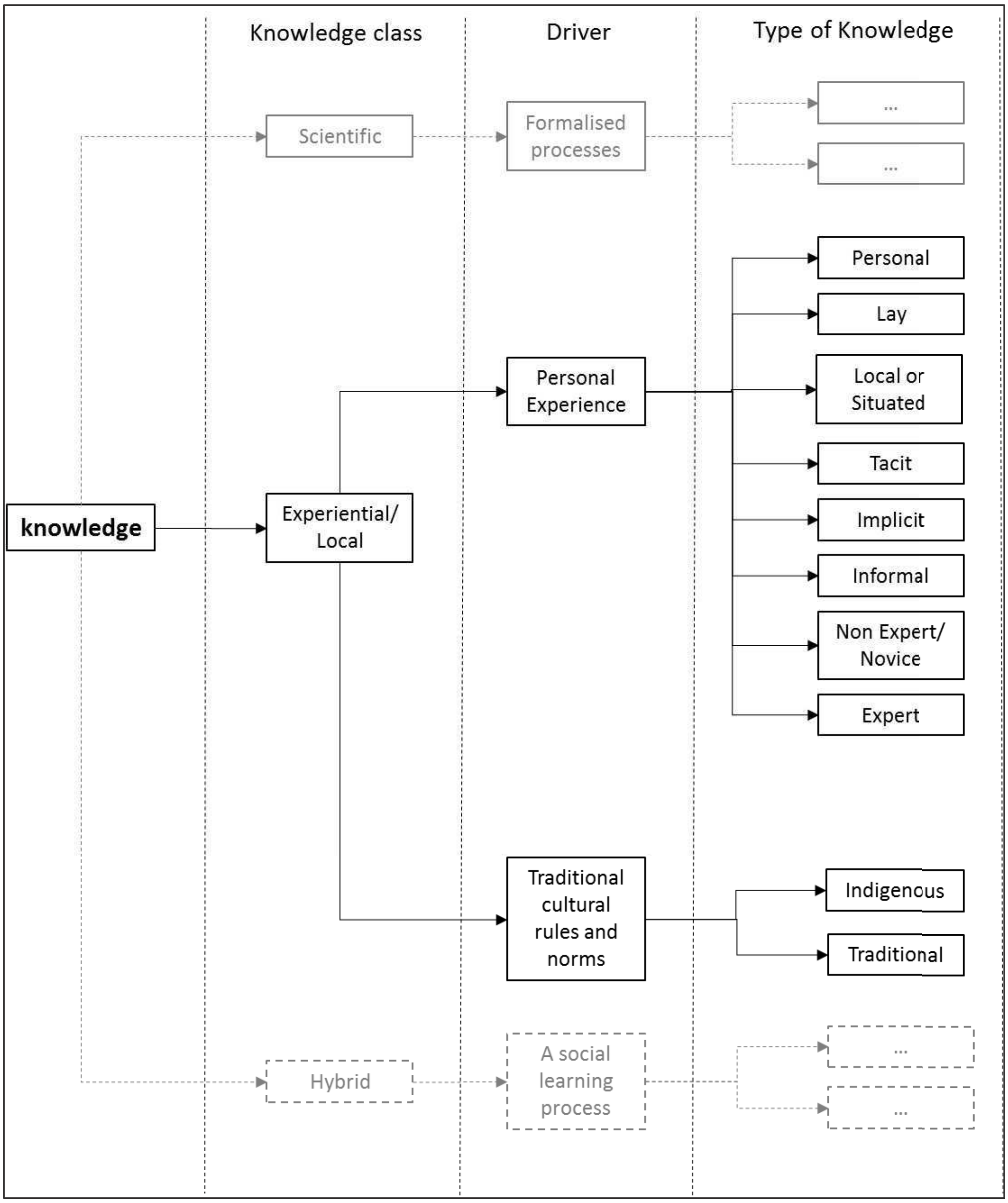


Fig. 2.20 Different types of knowledge within management literature (dashed and dotted lines relate to knowledge fields not considered in this Thesis). Adapted from Raymond et al. 2010.

Tacit

It is an unconscious knowledge that is often latent, difficult to express. Therefore, it has a significant influence on the individual thinking and behaviour, because it is deep and closely linked to each person's worldviews, values, personal experiences and expertise (Smith, 2001; Fazey at al., 2006; Kumar, 2012;).

Implicit

This form of knowledge is rooted in an individual but it is not accessible to others. It differs from tacit knowledge because it can be expressed (Fazey at al., 2006; Frappaolo, 2008). Some authors, including Ambrosini and Bowman (2001), argue that no difference exists between the two forms of knowledge (implicit and tacit), and that these are part of the same concept.

Informal

It is similar to personal and tacit knowledge. It refers to the knowledge acquired with different experiences, without a structured planning processes (rules and procedures) and which is aimed at improving understanding and learning (Van Herzele, 2004; Pasquini and Alexander, 2005; Bond at al. 2010).

Non Expert, Novice

This form of knowledge generally does not incorporate the depth of experience and expertise which, in contrast, are clear features of expert knowledge (Nonaka et al., 2001; Fazey et al., 2006).

Expert

It is a form of knowledge which is represented by the degree of experience of an individual gained over the years through practice, not necessarily formalized in scientific frameworks or structured processes (Kuhnert et al., 2005; Pollock et al., 2007).

Indigenous

Is it exclusively related to a specific culture and social group, or indigenous peoples (Mackay, 2009; Mercer, 2010; Howden, 2011).

Traditional

It can be classified as a local knowledge handed down across generations (Mercer, 2010; Howden, 2011).

Specific knowledge is rooted in local populations. This is fundamental to activate environmental protection actions aimed at: i) balancing natural, bio- and climatic systems; ii) conserving biodiversity to promote risk mitigation and to reduce vulnerability of territories; iii) strengthening resilience to natural, climatic and environmental disasters (Olson and Folke, 2001; Jawahar, 2003; Newport, 2003; Mercer et al., 2010; Kettle at al., 2014).

Also, specific knowledge is the result of experiences, awareness and sensitivity. Hence, it interweaves biophysical and social contexts, it simplifies the understanding of several local phenomena and it allows the assessment of priority interventions – as implemented through regional government policies and proper ecosystem management practices (Berkes and Folke, 2002; Picketts et al., 2012).

2.2.4 Role of Local Knowledge in Disaster event

Despite scientific knowledge supported by technological advances in Earth monitoring through *remote and proximity sensing* systems has considerably improved, risks and vulnerability arising from natural hazards has increased in both developed and developing countries, almost at same frequency and intensity (Burton and Van Aalst 2002; Gardner 2002; Dekens, 2007).

Consequently, the effects of these events on socio-economic and natural systems have also increased. The anthropogenic processes – such as building infrastructure – have opened up new immigration routes and created new risks, which especially affect the poorest people. The latter have moved at the edge of this infrastructure and cannot migrate in other urban areas. (Dekens, 2007).

In the face of these changes, a lot of information is embodied almost exclusively in local knowledge because it is related to behaviours and lifestyles of the people living in these contexts.

Generally, the areas in which individuals live generate a high degree of local knowledge. People usually know interesting details of their surroundings such as the busiest streets, on- and off peak times, the places and times for free parking spaces, hangout and commercial places. This information would be unknown to a newcomer, unless someone local would transfer him/her their knowledge. This is actually what is meant by local knowledge: it is the knowledge and awareness that people who live in a particular place have about their territory (Smith, 2011).

After having long been met with distrust or neglect on the side of international scientific research, local knowledge features in most recent studies on resilience, vulnerability, complexity and uncertainty in the context of environmental planning and management. This trend reflects a shift in the attitude of scholars, in search of fresh perspectives to investigate disaster events – such as closing the gap between local and scientific knowledge (Gaillard and Mercer, 2013).

Adversity towards local knowledge studies is probably linked to the difficulty of dealing with these topics from a computational point of view. **Fig. 2.21** shows the relationships between the DIKW hierarchical model and different forms of knowledge adapted from a work by Weichselgartner and Pigeon (2015).

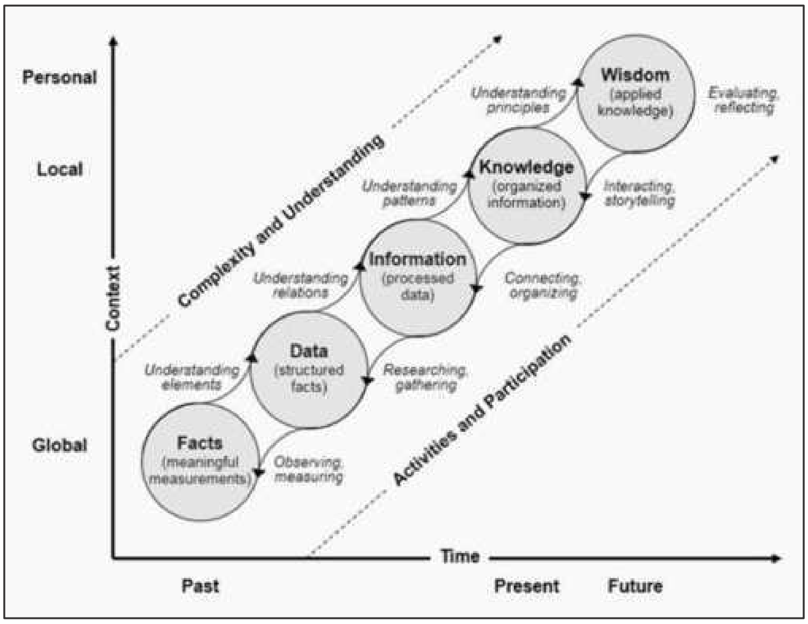


Fig. 2.21 Relationship between DIKW hierarchy and local knowledge (Weichselgartner and Pigeon, 2015)

In Fig. 2.21, knowledge is mapped according to time (X axis) and context (Y axis), with the former ranging from past, through present to future, and the latter being represented as a linear gradient (global, local, personal). An increase in the degree of both complexity and understanding requires the contextualisation at local level of knowledge and the direct involvement of affected actors, whilst data and Information are best handled when they concern global/general aspects of past events (i.e., it is more difficult to quantify local knowledge with data and Information).

Following the abovementioned paradigm shift, scholars have recognised the importance of using knowledge and local practices to manage complexity and risk-related problems (Olsson and Folke, 2001; Fabricius et al., 2006; Raymond, et al., 2010). Recent studies focus on local knowledge as a potential tool of vulnerability, resilience and mitigation in, among other fields, flooding risk (Slinger et al. 2007; Peters-Guarin, 2012), hydro-meteorological risk (Hiwasaki, 2014), Tsunami risk (McAdoo et al., 2009).

All experiences related to typical local attitudes, such as: community participation is thought to support the identification of local resources, strengthen the ability to adapt, improve risk perception with respect to familiar places, empower collective action, facilitate the willingness of individuals and organisations to share cultures, values and existing structures. When involving community participation, vulnerability assessment appears more effective in implementing practical systems which are more likely to satisfy the needs of the local community (Newport, 2003; UNISDR, 2009; Harwood et al. 2015).

Forms of local knowledge, when observed in a dynamic context, can take on specific features. There emerge complex cognitive systems including language, the sense of belonging to a place, spirituality and worldviews (Van Camp, 2007). These forms of indigenous knowledge, which can be considered ‘at margin’, could be decisive for understanding social, cultural and religious dynamics, and hence solve local problems. (Tickner, 2015).

Among the different conceptual models explaining how to deal with disasters and what strategies are to put into practice (Kates, 1971; Clarke-Guarnizo 1992; Hall and Davis, 1999), only a few venture into defining local knowledge and explaining the complexities associated with it. One such case presents an interesting framework (**Fig. 2.22**), developed by Dekens (2007). This is based on a review which identified relevant elements that come into play in a local knowledge system, which are organized and grouped into subsystems that analyse spatio-temporal contexts.

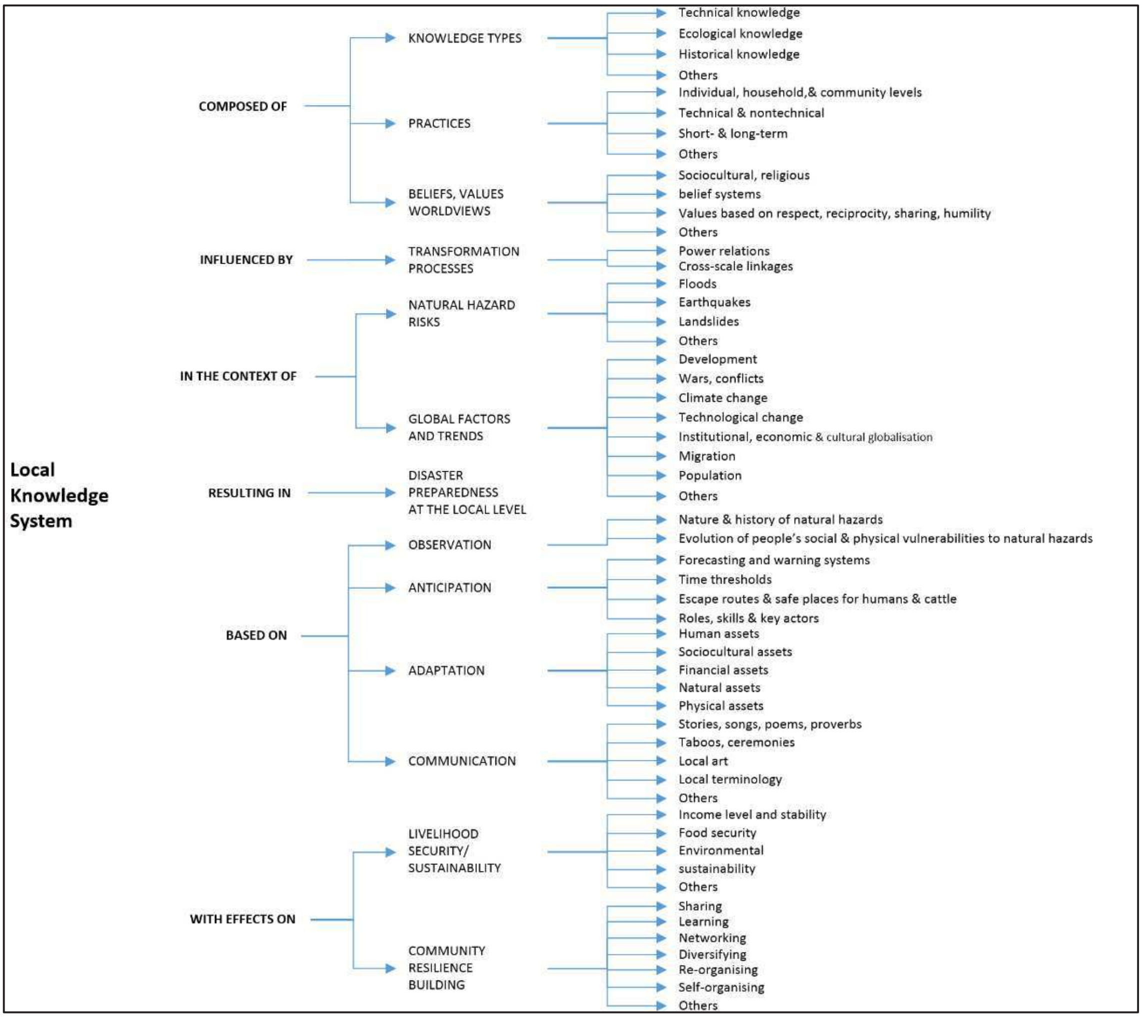


Fig. 2.22 Framework for local knowledge in disaster preparedness. Adapted from Dekens (2007)

2.2.5 Knowledge and Action

After examining definitions and forms of knowledge, this section analyses knowledge from an operational point of view, starting from two pioneers of the literature of the 20th century: Donald A. Schön e John Friedmann.

Although they both were scholars of Urban Planning, they had different cultural backgrounds. Their main contribution concerns, from different viewpoints, the relationship between knowledge and action. This is a relevant aspect for the purposes of this Thesis, and it is directly linked to several topics of interest (such as Response, Recovery, Mitigation, Preparedness).

Schön (1983), in his seminal work *The reflective practitioner*, explains the concepts of knowledge, reflection and action and discusses the relationships occurring across these topics in an attempt to overcome the separation between thinking and acting, knowing and doing, deciding and implementing. These can be defined as dichotomies belonging to an old-fashioned epistemological stance. The author draws a distinction between 'reflection in action' and 'reflection on action'. These concepts are both centred on knowledge (Zeichner, 1990) and have influenced current literature in many areas of research, highlighting chronology as an important element during the reflective activity (which might take place before, during, or after the action) (Nuzzaci, 2011).

More in detail, '*knowledge in action'*, is practised for those events involving intrinsic knowledge. These are processes or patterns of non-logical action (such as riding a bicycle) that silently manage rational moves. Under risk situations these 'moves' would save lives (Schön, 1983). Therefore, reflective skills imply '*knowledge in action*' based on tacit and latent knowledge to face events and deal with specific tasks (Nuzzaci, 2011).

Argyris and Schon (1979) in *Organizational Learning: A Theory of Action Perspective* explain how knowledge is acquired and stored over time. The authors also explain how this learning mechanism could improve either the specific process of which it is part as 'single-loop learning' or, if reiterated, similar processes related to them as 'double-loop learning'. Through the latter, the members of an organization would be able to uncover problems and adapt and improve any learning strategy (Senge, 1990; Drupsteen and Guldenmund, 2014).

In *Planning in the public domain: from knowledge to action*, Friedmann (1987), delves into the relationship between knowledge and action, which he addresses as a possible key to identifying opportunities and limits of planning the public domain. Friedmann critically discusses the definitions of knowledge and action and concludes on a non-existence of definitive answers. The author highlights the limits of scientific knowledge as this is expressed through theories and simplistic models. The latter assumes that the world remains unchanged and questions whether knowledge is at the exclusive power of experts or it arises from the experience of a 'non-scientific' personal or experiential action (Polanyi, 1962).

Also, Friedman (1987) points out that actions are more important than decisions and hence more crucial than the intrinsic knowledge: in planning a decision which is not followed by an action would be meaningless. However, this does not mean that acting necessarily means performing well. Friedman argues that it is not possible to predict the long-term outcomes of actions, and he also cast doubts on the reliability of predictive models in planning.

This uncertainty could affect people if the planning process and the predictive models are not efficient. Actions should be, for this reason, dynamic and implemented through adequate strategies (Friedman, 1987).

2.2.6 From Knowledge to Words and Acts: The Speech Act Theory in Risk Response

So far, knowledge has been analysed from a cognitive point of view, leaving aside the limitations due to the transitions from thought to (written and oral) language. To address these limitations, the Speech Act Theory is an important reference point (Austin, 1962)[8]. This is also due to the particular context of the exchange of communication under risk situations. The methodological and applied aspects of the Speech Act Theory are appropriate to support the type of information that are being exchanged on social media.

During a disaster event, both demand for and offer to help are written in such a way that they assume an assertive force which resonates with the theories of Frege (1879) and Austin (1962) and Searle (1969).

The Speech Act Theory has played a key role in shaping contemporary pragmatic linguistics, a discipline that deals with the contextual and interactional aspects of language (Sbisà, 2011).

Philosophy, linguistics, psychology, law, artificial intelligence, literature, and other scientific disciplines have been influenced by the Speech Act Theory (Gormon, 1999). The Speech Act Theory changed the understanding of communication. It focuses on the 'intentions to communicate' of the speaker, and overcomes a communication model based on the encoding, where the sender and receiver are simple terminals of a mechanical process that transmits and decodes messages (Sbisà, 2011).

Studies on syntax and semantics are not exhaustive to understand a language. For the Speech Act Theory, revealing statements means to perform acts. It is necessary to detect the linguistic acts of conversation, since all verbal or written actions have their own counter effects and corresponding reactions (Schegloff, 2007).

The actual understanding of utterances in terms of speech acts is based on a complex process that involves the use of a general knowledge organized into conceptual systems

[8] The dissemination of the Speech Act Theory is attributed to John Searle, with the book *Speech acts* of 1969, in which he redefined Austin's thought, albeit partially modifying it.

(Van Dijk, 1977) such as speaking and writing. The type of verb and/or the presence of auxiliary or modal verbs, adverbs and conjunctions, intonation and accompanying gestures affect the meaning of a sentence (Austin, 1962: 56-59). In his theory, Austin distinguishes between 'performative utterances' and 'constative utterances'. The latter declares something that is true or false (i.e. it is not action). Performative utterances, instead, are considered actions (i.e. they do something) (Austin, 1962)[9].

By analysing different messages exchanged during a disaster event, it is possible to observe that, in many cases, requests for, and offers of, help have the same characteristics of performative utterances as theorized by Austin as well as recent studies.

Also, Austin (1962) argues on the importance of the senses during speech. When a sense is relevant, then an act is named *locutionary act*. A simple *locutionary act* can be performed as follows:

- Making the *phonetic* act;
- Making the *phatic* act (i.e. use expressions that follow rules, belong to a vocabulary and have proper intonation);
- Making a *rhetic* act (i.e. use expressions containing synonyms).

For the purposes of the present Thesis, the most relevant aspect of Austin's thought is to consider an utterance such that it is possible to understand knowledge from the communication between the sender and receiver of a message.

As for this aspect, Austin uses: i) the concept of *locution* to refer to the simple structure of an utterance; the concept of *illocution* to refer to the object of communication; iii) the concept of *perlocution* to refer to the validation of the message by the receiver (i.e., the type of effect that the message produces on the receiver).

In a risk situation the utterance 'hurry up' (one locution), for example, could correspond to the following illocutions:

- The simple intention to ascertain something by way of information;
- The intention to communicate that time is limited;
- The intention to invite someone to move speedily;

[9] A constative utterance may be reworked in a performative manner by adding predicates that imply an action – e.g.: constative ('The Earth is round'); performative ('I maintain that the Earth is round').

- The intention to communicate that someone's life is threatened.

A complex conception of the relationship between language and action was also established by the philosopher J. Searle in the 1960s and 1970s, which is based on the dependence of any meaning and any intentionality by a shared context and made of practices, rules and conventions social (Di Lorenzo, 2001). Searle disagrees on the paradigms that reduce semantics to computational aspects: cognitive, perceptual and mental structures are important to understand communication (*ibid.*).

He extends Austin's Speech Act Theory and, by defining the set of *illocutory forces*, creates a taxonomy of five basic types of illocution (see **Tab. 2.5**)[10].

Tab. 2.5 Detailed description of the speech act

Speech act	Description	Verbs associated with Speech Act
Assertive	Statements that can be verified as true or false	Assert, claim, affirm, assure, inform, predict, report, suggest, insist, hypothesize, swear, admit, confess, blame, praise
Directive	Statements that call upon the listener to do something	Direct, request, ask, urge, demand, command, forbid, suggest, insist, recommend, implore, beg
Commissive	Statements that commit to a course of action	Promise, vow, pledge, swear, consent, refuse, assure, guarantee, contract, bet
Expressive	Statements that express a psychological position about state of affairs	Apologize, thank, condole, congratulate, complain, protest, compliment, praise, welcome
Declarative	Statements that, through their utterance, perform act	Fire, pronounce, declare, appoint, confirm, endorse, renounce, denounce, name, call, repudiate

Source: Adapted from Searle (1976) and Searle and Vanderveken (1985)

[10] Wierzbicka (1987), building on the works illustrated in Tab. 2.5, classified 229 English verbs (representing as many speech acts) into 37 groups – a very useful endeavour for application testing.

2.2.7 Using The Speech Act Theory in Risk Response

The use of the Speech Act Theory to analyse messages in emergency/risk situations is a complex task. Often communication systems place limits on the number of characters used in a text message. Given the wide diffusion of short-message services (SMS) in risk situations, recent studies highlights the need to teach and accustom individuals to communicate properly and effectively (Gomez and Elliot, 2013). In the aftermath of the Haiti earthquake of 2010, to cite an example, about 100,000 SMS were sent to ask for help through the free service 'Mission 4636'[11] – made available to all providers (Heppler, 2010). However, out of 100,000 SMS, only 16,000 were readable; among these, only 3,000 SMS contained useful information to implement a rescue action (Gomez and Elliot, 2013).

From a linguistic point of view, the study of SMS in emergency situations through the speech act model shows significant potentials which are relevant to support the classification of thousand instances (Vosoughi and Roy, 2005). These require the adoption of theoretical analysis and interpretation of new models (Budzynska and Redd, 2011).

Ogasawara and Ginsburg (2015) analysed the Japanese tsunami in April 2014. In their study, the authors assess the effect of the calls issued by the rescue bodies on the population in terms of strength. The calls were classified and ranked second as illocutionary acts. These illocutionary acts implicitly required the caller to perform actions such as an 'order', a 'declaration', a 'request', a 'promise', or a 'permission'. The results show that SMSs declaring imperative requests such as 'Please immediately move away from the shore' and employing a major communicative strength/power seem to corroborate the theory of Searle (1969). This highlights that the success of the illocutory act lies in its preparation, and that it is a relevant step to reach the reader.

In an attempt to fully understand the structure of the debate, it is appropriate to consider the *Inference Anchoring Theory* (IAT) (Budzynska and Redd, 2011), which provided significant advances in automatic and semi-automatic search (Budzynska et al., 2016), similary

[11] Mission 4636 was a Haitian initiative launched in the immediate aftermath of the 2010 earthquake to facilitate relief actions, mainly based on free text messages services and complementary support (e.g. translation and mapping). See http://www.mission4636.org.

the *Argumentation mining* – a young discipline stemming from law studies (Moens et al., 2007) deals with zoning argumentative segments of a speech in different types of information (Peldszus and Stede, 2013).

2.3 Ontology and Domain Description

Several studies on modelling knowledge in risk/emergency domains show that the use of ontologies is nowadays an accepted and recognised approach from the international scientific community (Wang, 2009).

In the following sections, a definition of ontology is provided, in a following step the difference between foundational and lightweight ontologies is introduced, as following as an illustration of the use of ontologies in the domain of event response knowledge.

2.3.1 What is an Ontology?

Since of Aristotle, the concept of Ontology is linked to a philosophical endeavour oriented to the observation of entities and the development of categories to represent them.

However, only a quarter of a century ago, the computational/engineering aspects of an ontology has gained momentum within the international scientific debate. Though rooted in the above mentioned philosophical conceptions, the engineering ontology represents a new approach to modelling the real world.

The main purpose of an ontology is to create knowledge that is the engine (inference engines) to specific systems which can simulate and represent the human reasoning in the use of semantics (Poli and Obrst, 2010). In their paper *What is an ontology?*, Guarino et al. (2009), describe the difference between the term 'Ontology' (under the philosophical conception) and the term 'ontology' (i.e. the computational conception). The authors also refer to a definition of the term ontology by Studer et al. (1998: 184) stating that: '*An ontology is a formal, explicit specification of a shared conceptualization*', and point out that this definition extends and mixes two previous definitions by Gruber (1993) and Borst (1997): the former considered an ontology as an '*explicit specification of a conceptualization*' and the latter as a '*formal specification of a shared conceptualization*'. A conceptualization, in turn, refers to a simplified view of the world (Gruber, 1995), by identifying the concepts (i.e. the entities) involved in it.

To build a conceptualization of a particular domain it is required to identify the elements that are part of it.

These elements are immaterial entities that are part of an individual mind. These entities should be translated into a concrete artefact (i.e., a model) and represented by a modelling language (Guizzardi et al., 2005) in order to be documented, tracked, communicated, shared and analysed.

Fig. 2.23 shows the relationships occurring between the four aspects of conceptualization, domain abstraction, model and modelling language.

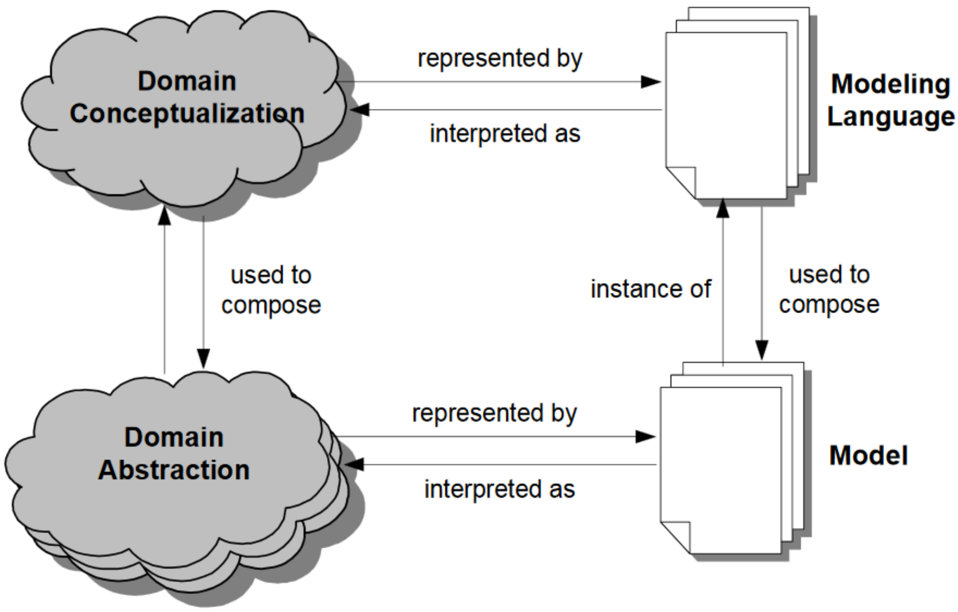


Fig. 2.23 Relationship between conceptualization, abstraction, modelling language and model. Adapted from Guizzardi et al. (2005)

Fig. 2.24, adapted from a work by Guarino et al. (2009), illustrates the setup of an ontology, from the perception of a real world case to the realization of the ontology, along with the following main steps: 1) the perception of real world situations at different stages; 2) the conceptualization (abstract); 3) the definition and use of a language to describe it; 4) the identified models and ontologies[12].

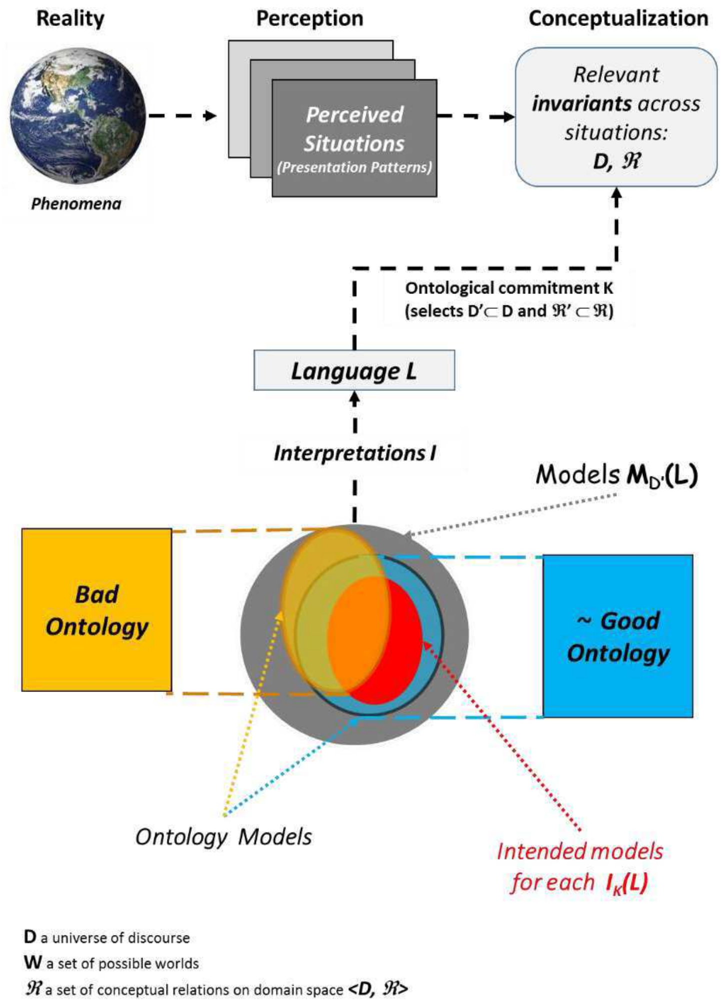


Fig. 2.24 From phenomena to ontology models. Adapted from Guarino et al. (2010)

[12] For further details, see Guarino et al. (2009).

A conceptualization defines all possible eligible domain abstractions (representing the state of the world), and a modelling language defines all possible specifications allowed by the language (Guarino, 1998 and Guizzardi, 2005).

From a methodological point of view, the ontological analysis sets the boundaries of a domain (including all its entities), and through the use of a specific language it realises the ontology.

The occurrence of errors during the implementation of this process would create a serious impact on the final design of the ontology. **Fig. 2.25** shows the outcome of the biased model as follows: a) an efficient model in which the ontological model approximates the initial conceptual domain; b) an oversized model; c) an undersized model; d) a model that partly meets the requirements identified with the ontological analysis.

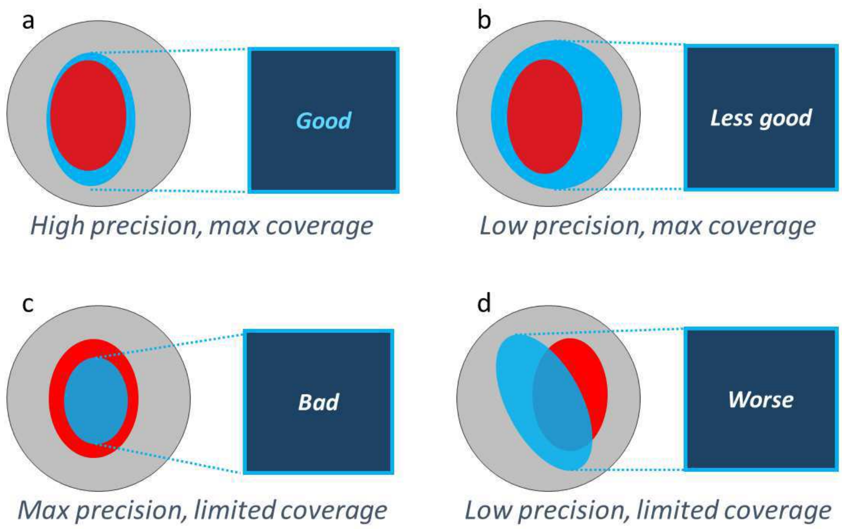


Fig. 2.25 Ontology domains' precision. Adapted from (Guarino, 2008)

Fig. *2.26* illustrates how the perceptions (translated into ontologies) of the same concept (for example, 'apple'), when realised by two different communities (farmer and food processing company) that interact with each other for common interests, can lead to different results and cause misunderstandings in trade relationships.

The mismatch may be due, for example, to professional semantic 'biases'. As a result, the information system (realised on the basis of the two ontologies) does not share the same assumptions about the meaning of the terms that are used.

Thus, the domains are restricted in such a way that the agreement on the meanings occurs in a small area (the *area of false agreement*). As a result, the ontology partly meets its purpose, according to two of the cardinal principles, namely *understanding* and *agreement*.

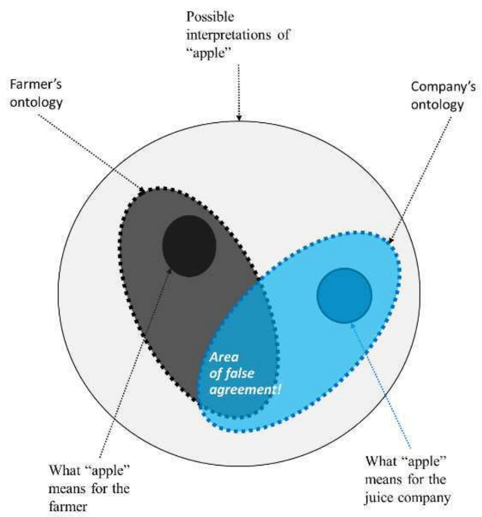


Fig. 2.26 Ontology domains' precision. Adapted from Guarino (2008)

The *agreement* principle requires a general consensus: it is not based on individual subjective vision/perception (Gaio et al., 2010).

Nonetheless, the risk of creating areas of false agreement in decision support systems is high. This is particularly relevant in a globalised world where migrations (caused, among other factors, by wars, poverty, epidemics) involve large populations with mixed ethnic backgrounds, languages, and cultures – thus potentially triggering knowledge biases.

Shared knowledge spaces are important to improve exchange of information and favour semantic integration. In particular, the domain of communication in emergency situations, as already described in **Section 2.2.3**, has developed a mismatch between expert and scientific knowledge (generally in a dominant position in rescuing operations), and local knowledge. Recent literature (Raymond, 2010) recognises a key role played by local knowledge in communication as well as other domains (e.g. marketing, economics) because it is considered as depository of important information to solve complex problems.

2.3.2 Formal ontology: Foundational and lightweight ontologies

A formal ontology identifies, within a specific domain, entities and their respective properties and relations on a basis of a logical system. The ontology generally pursues one or more among the following three goals (Gaio et al., 2010: 108):

- the 'representation of information';
- the 'description of a certain domain'; and
- 'the development of a systematic theory for a certain type of entity'.

Over the last years, several types of ontologies have been established. These differ in terms of the level of abstraction of the real world and the type of formalisation and representation. One of the key differences within the field of formal (or computational) ontologies is the one drawn between *foundational* (upper) and *lightweight* ontologies.

Generally, the latter are structured as simple taxonomies of concepts, often hierarchically structured (Oltramari et al., 2003; Gaio et al., 2010). The former, on the other hand, faces more general and cross-cutting issues between different domains and communities. Therefore, ambiguities across meanings arise. Formal ontologies can be considered as bridges between different communities (Oltramari et al. 2003; Gaio et al., 2010).

The remainder of the section will focus on the elements of formal ontologies, given their complexity. The literature (Obrst, 2010) distinguishes formal ontologies through multiple categories:

Descriptive vs. Revisionary

These types of formal ontologies are based on two different approaches regarding the ontological modelling and conceptualization mode:

- Descriptive ontology establishes that what is predicated must be true.
- Revisionary ontology (or prescriptive) establishes that some things do exist while others are only figures of speech.

Both approaches can be rigorous and formal.

Multiplicative vs. Reductionist

The main difference between multiplicative and reductionist ontologies is at conceptual level. The former has a vast, heterogeneous and inclusive vision of any entity linked to the real world, provided that it is detectable to the purposes of an ontology. The latter incorporates a limited number of concepts, with a consequent reduction of primitives, to derive the complexity of the real world (Obrst, 2010).

Universals and Particulars

The discussion on the difference between Universal and Particular is complex and philosophically-oriented, and goes beyond the scope of the present Thesis. A brief reminder will therefore suffice.

As for Universal, it refers to an entity of a general nature which exists independently of us and of our ways of thinking and talking about the world (Hamlyn, 1984). A Universal ontology identifies general entities (natural and abstracted) such as person, position, *etc*. Based on these entities, it is possible to classify instances named 'Particular' (element with specific properties).

Endurants and Perdurants

The literature distinguishes between enduring entities (*endurants*) and perduring entities (*perdurants*). These are compared using the time dimension.

As for endurants, they present a dynamic behaviour over time, they exist in every moment of their life cycle, and when they exist they are fully equipped.

On the contrary, perdurants are almost static and sometimes are partially equipped (Masolo et al. 2003).

The above types of ontologies are both still present in different foundational ontologies.

The following description refers to some of the most common foundational ontologies, as illustrated by Apisakmontri et al. (2013):

- *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE)[13]
  DOLCE is part of a larger international project (WONDERWEB)[14], and it aims to be a foundational ontology that covers a set of higher/superior concepts, which represent a domain including basic categories drawn from natural language and common sense (see also Section 3.5.2).
- *Semantic Web for Earth and Environmental Terminology (SWEET)*
  According to its structure, SWEET is a reductionist ontology, primarily designed to deal with issues associated with Earth sciences in all its forms. It has a very extensive vocabulary of thematic terms (Raskin and Pan, 2005).
- *Suggested Upper Merged Ontology (SUMO)*[15]
  SUMO is a top-level ontology which is characterized by its architecture (oriented to the Semantic Web. It is able to link categories and relationships belonging to other foundational ontologies, with the aim of creating a standard ontology and promoting interoperability between data, analysis of natural language and search for information.
- *Friend of a friend (FOAF)*[16]
  FOAF is an ontology that has the propensity to representing people and t heir behaviours and actions. These can also be considered in relationship with objects or other people.

Among the above ontologies, the DOLCE ontology seemed more appealing assist with to the research activities planned in the framework of the present work, and is therefore discussed in further details in Section 3.5.2.

---

[13] http://www.loa.istc.cnr.it/old/DOLCE.html. Accessed: 23/09/2016.
[14] http://wonderweb.semanticweb.org. Accessed: 12/06/2016.
[15] http://www.adampease.org/OP. Accessed: 18/06/2016.
[16] http://www.foaf-project.org. Accessed: 18/06/2016.

2.4 Crowdsourcing and Social Sensing

Over the last twenty years under the lens of the sustainable development concept and numerous international accords (Agenda 21, Millennium Development Goals, Aarhus Convention) there has been a growth of various initiatives to involve public participation in the decision making process (Banini, 2011).

Against the systematic adoption of public practices (e.g. workshops, round tables, focus groups), a lack of understanding local knowledge within decision making processes arises.

Omitting both ideological aspects, related to the usefulness and political will to use this knowledge, and the debate of participatory democracy *vs* deliberative democracy largely addressed in other contexts (Bobbio 2007 et al.; Bifulco, 2009), it is clear the difficulty to recognise local knowledge as well as that of expert and scientific knowledge. The difficulty arises because local knowledge derives from unstructured processes and is not easy to use it in computational processes.

The latest social changes due to the rapid development of Information and communication Technology (ICT), have primarily affected social interactions through countless applications oriented to share textual and multimedia content.

These multimedia applications represent bottom-up local knowledge and draw a line between old and new generations, between new on-line systems of public participation and traditional ones.

In addition, Web 2.0 technology (Roche et al., 2012) platforms like Twitter, Ushahidi, and OpenStreetMap have witnessed the rise of terminologies including geospatial and participatory aspects such as Volunteered Geographic Information (VGI), Public Participation Geographic Information Systems (PPGIS), the participatory geoweb, neogeography, citizen sensing, crisis mapping, crowdmapping, and citizen science (Elwood, 2010; Fast and Rinner, 2014; Leszczynski, 2014, Kar et al., 2016).

In particular, PPGIS and VGI find application for several governmental activities (Verplanke, 2016). However, while these new tools open new collaborative scenarios, they still leave some complex issues unresolved.

During the 18th and 19th centuries, crowdsourcing and citizen science were born in Great Britain thanks to research projects such as the Whewell tide monitoring (Cooper, 2016). Citizen science can be seen as an approach in which non-experts and professionals collaborate in scientific research to define research questions, collect and analyse data, and interpret and disseminate results (Kar et al., 2016).

The citizen science used in the present Thesis refers to scientific activities in which non-professional scientists voluntarily participate in data collection, analysis and dissemination of scientific projects (Cohn 2008 ; Silvertown 2009).

The digital divide (Helbig et al., 2009) between developed and developing countries and between richer and lagging behind regions within developed countries still causes unsolved questions in terms of both the degree of representativeness of public participation and the understanding local knowledge.

2.4.1 Crowdsourced Data: Volunteering or Contributing in Disaster and Crisis

For several years the growth of geo-referenced data was due to private firms providing their services to governmental organizations (Goodchild and Glennon, 2010).

With the introduction of web 2.0 technology and diffusion of mobile devices such as tablet and smartphones endowed, through user friendly interfaces, with *Global position system* (GPS) technology, the decision making process evolves to the digital paradigms such as e-Democracy, e-Participation, and Gov 2.0 (Floreddu, 2012; Latre, 2013). The exchange of data through these new technologies provides detailed information (Goodchild, 2007; Gill e Bunker, 2012, Horita, 2013) and supports the diffusion of local geo-referenced knowledge (Frommberger and Schmid, 2013) .

Over the last 15 years, PPGIS/PGIS and VGI have been the main technologies used for e- participation.

PPGIS/PGIS

PPGIS/PGIS (Participatory Geographic Information System) and VGI although based on similar information architectures pertain different aims such that the literature has confined them in two different domains.

PPGIS/PGIS aim to reinforce local active participation and achieve detailed regional/territorial analysis. VGI aims to the implementation of methodologies oriented to large scale data acquisition (Tulloch, 2008; Sieber and Haklay, 2015).

The term PPGIS is used for the first time in 1996, at the meeting of the National Center for Geographic Information and Analysis (NCGIA). It describes a new model that combines geographic information system (GIS) with the public. The main aim is to provide a direct and responsible participating of citizens, non-governmental organizations, and local stakeholders to the activities of the decision maker (Sieber 2006 Gulnerman, and Karaman, 2015).

The use of PPGIS has largely developed in developed countries (Obermeyer, 1998) while PGIS was found more practical in developing countries (Rambaldi et al, 2006).

From a methodological point of view, PGIS and PPGIS are based on geographical information systems known as Participatory Learning and Action (PLA) (Rambaldi et al. 2006, Verplanke et al., 2016).

Brown et al. (2014) argue that there exist ideological differences between PGIS and PPGIS. PGIS has often been used by local and non-governmental organization groups of developing countries to promote protests and dissent actions to political parties for income distribution and the rights of indigenous peoples. PPGIS represents a tool for public participation initiatives in developed countries.

Both practices emerge on the emphasis of active local participation (local spatial knowledge) which aims at strengthening local powers and rights (McCall et al., 2015) which, through geo-referenced tools, facilitate the contribution of local population in the decision-making process.

Nonetheless, McCall et al., (2015) note that the difference between PPGIS and PGIS in recent literature is non-existent.

Beyond the debate of defining the two systems (Tulloch, 2007), it is reasonable to argue that these tools represent a real step enabling ordinary/non-expert individuals to public participation. Non-expert and expert knowledge (already existing on GIS) combine together to widen local knowledge system (Brown et al. 2014).

PGIS will be used in the next sections to refer to both PPGIS and PGIS paradigms.

*VGI/*CGI

The term VGI was introduced for the first time by Goodchild (2007). It refers to a set of approaches, systems and methods for the collection of local knowledge and organization, based on architectures that use techniques such as UGC[17], Web 2.0, Geoweb[18] in which the geo-spatial component plays a key role (Goodchild, 2007; McCall et al., 2015; Xu and

---

[17] User-generated content (UGC) are Websites platforms, where citizens can publish their own comments, photos, videos, and more online such as YouTube, Facebook, Twitter, and Wikipedia (Hermida and Thurman, N, 2008)

[18] Geoweb is defined as a distributed set of geographic services that are user-controlled and available on the internet (Scharl and Tochtermann, 2007).

Nyerges, 2016). It is important to distinguish between Contributed Geographic Information (CGI) and Volunteered Geographic Information (VGI).

The former refers to geographic information collected through a system in which the individual is not aware of being tracked; vice-versa for the latter (Harvey, 2013)[19].

VGI conceptualization is not yet rigorous and somewhat inaccurate. The term 'volunteered' refers to voluntary acts that the user performs when he/she provides information about a particular domain.

To reach the aims of the present Thesis, both VGI and CGI are useful tools. The former contributes to extend knowledge about affected places in post-disaster events and collecting help requests; the latter contributes to assess damages and detect the presence of people exposed to some form of risks.

Verplanke (2016) points out that the success of VGI is mainly due to two aspects. The first aspect is the social phenomenon, born outside academic and/or scientific environments. It is linked to the process of public participation in local government and particularly relevant in the planning sphere; the second aspect is related to the technological nature of the tool. It develops tools that can handle geo-referenced applications (i.e. software applications and platforms that exploit the potential of GIS often as Open Source)[20].

Open Source technologies provide cost reductions and easy access to geographic information. This has also led to increase the diffusion of the world wide web and mobile devices in developing countries, to overcome the issue of the digital divide, and opening up access to spatial data previously available to professionals. For the above reasons, VGI applications

[19] The issue of privacy goes beyond the purpose of the present Thesis.

[20] 'The Open Source Initiative (OSI) is a California public benefit corporation founded in 1998. The corporation is also actively involved in Open Source community-building, education, and public advocacy to promote awareness and the importance of non-proprietary software. OSI Board members frequently travel the world to attend Open Source conferences and events, meet with open source developers and users, and to discuss with executives from the public and private sectors about how Open Source technologies, licenses, and models of development can provide economic and strategic advantages.' Source: https://opensource.org

are an important tool for local knowledge often defined as 'momentary engagement' by non-profit agencies (international development, emergency services and humanitarian organizations) and particularly relevant during post-disaster and risk events (Goodchild, 2007; Coleman et al. 2009; Haklay, 2013; McCall et al, 2015; Xu and Nyerges, 2016; Verplanke, 2016).

To date, there are several successful stories and experiences in different contexts such as Quakemap in 2015; Chennai Flood Help in 2016; Apartheid In Palestine in 2016, to cite a few[21] which are born through specific tools such as the grassroots maps of the Humanitarian OpenStreetMap Team (HOT)[22] and Wikimapia[23], or as the crisis maps of Ushahidi[24] (Okolloh, 2009; Ziemke, 2012). These initiatives involve thousands of people in disaster management and are often referred to in the literature as Volunteer and Technical Communities (VTC) or Digital Humanitarians (DH) (Horita, 2013).

The recent development of the Citizen Science has helped the legitimacy on the use of this information also for scientific purposes. An 'authoritative knowledge' in different disciplines has been recognised (Verplanke, 2016).

Brown and Kyttä (2014) argue on the interest and distributions of PGIS and VGI. On October 31, 2013, they searched for the keywords 'public participation GIS', 'participatory GIS', and 'volunteered geographic information' in scientific peer-reviewed journals indexed by bibliographic databases dealing with emergency issues (see **Tab. 2.6**).

---

[21] https://chennaifloodhelp.ushahidi.io/views/map - https://siddadel.ushahidi.io/ - https://quakemap.ushahidi.io/
[22] https://hotosm.org/
[23] http://wikimapia.org
[24] https://www.ushahidi.com/

Tab. 2.6 Search results from selected bibliographic databases (as of October 31, 2013)

Database	PPGIS	PGIS	VGI
Search term	"Public participation GIS"	"Participatory GIS"	"Volunteered geographic information"
ISI Web of Science[a]	82	87	84
Scopus[b]	129	105	177
Google Scholar[c]	1770	2410	2060
Most published journal	*Applied Geography*		*Transactions in GIS*

[a] Search for articles, books, book chapters, and published proceedings in "topic" which includes title, abstract, and keywords.
[b] Search for articles or conference papers in title, abstract, keywords.
[c] Search results are estimates.

Source: Brown and Kyttä (2014)

However, several critical issues remain open for discussion to respond to the scope of the present Thesis.

These critical issues concern: i) the problem of interpreting information structured as texts and expressed in natural language; and ii) the storage and sharing of knowledge in compliance with international standards for the Spatial Data Infrastructure (SDI) (Bishr and Janowicz, 2010).

Far more important are the questions raised about reliability and quality of data collected through crowdsourcing (Flanagin and Metzger, 2008; Elwood, 2008).

The use of platforms dedicated to public participation increase the risks of manipulating real world information. The solution to this issue is a challenge for future scientific research.

2.4.2 Crowd Sourcing and Tweak the Tweet

The methodology Tweak the Tweet (TtT) was proposed for the first time by Starbird and Stamberger (2010) who introduced a series of hashtags (e.g. #need, #iamok, #offer) to use in Tweets during disaster events. Nonetheless, in the case of the Haiti earthquake, Starbird and Palen (2011) found that few users adopted the specific proposals of the TtT paradigm.

Soon after its launch, the TtT paradigm was used in pilot projects carried out by various academic institutions. The EPIC project (Empowering the public with information in Crisis) aimed at increasing public awareness about a disaster (see Fig. 2.27). Tweets were gathered using specific hashtags about the disaster events (Wendling et al., 2013).

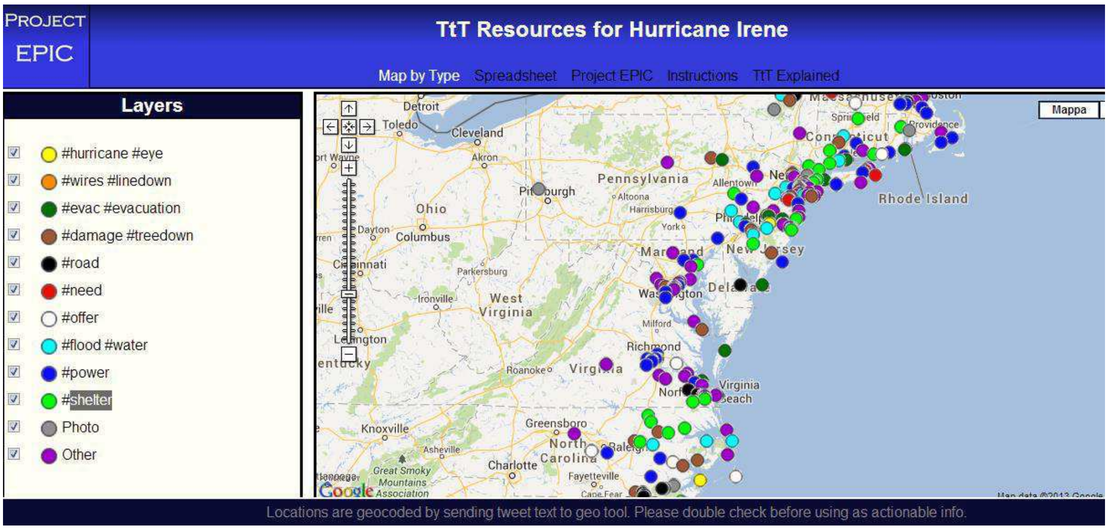


Fig. 2.27 TtT Resources for Hurricane Irene – Project Epic

Another noteworthy initiative is that of the Department of Human Centered Design and Engineering (HCDE), Washington University, that has created a platform[25] that collects and locates reports regarding Hurricane Sandy organized on specific hashtags (see Fig. 2.28).

[25] http://faculty.washington.edu/kstarbi/TtT_Hurricane_Map_byEvent.html

Fig. 2.28 TtT Sandy Map - Department of Human Centered Design and Engineering (HCDE)

2.4.3 Social Media role in Event Communication

The term social media refers to a type of communication that uses web platforms where users discuss and share general topics. The platforms include blogging and micro-blogging, social networking sites, platforms, and wikis. Social media has nowadays reached a global level such to occupy all areas of public life (van Dijck and Poell, 2013).

Over the past decade, social media applications have radically changed the lives of millions of people, influencing the interaction between people and playing an important role in sharing crisis information (Bruns, 2014).

The late nineties witnessed the first experience of using social networks to collect and share information from citizens being in an emergency, crisis or protest. It dates back to 1998 the protests in Indonesia coordinated through internet points, newsgroups and email messages (Poole et al. 2005); another case appeared in January 2001 where more than one million citizens in Manila organised a protest against the President Joseph Estrada of the Philippines using SMS and e-mail messages. The protest led to the president's dismissal (Rafael, 2003).

One of the first websites with message board was created soon after the Indian Ocean tsunami of December 26 in 2004 to support the relief efforts of the affected population (Imran et al., 2015). A similar experience was repeated a year later (2005) by the citizens of New Orleans on Myspace platform immediately after the devastation of the city caused by Hurricane Katrina (Sklovskij et al., 2010).

Twitter was used for the first time in disaster event in 2007 during the fires in the areas surrounding San Diego, California (United States). Since then, Twitter is one of the most used platforms by affected people in disaster response aiming at collecting and disseminating information and supporting relief operations (Vieweg et al., 2010; Cobb et al. 2014; Imran et al., 2015).

The case of the Haiti earthquake of 2010 is perhaps the most cited in the literature (Blum et al., 2014; Haworth and Bruce, 2015; Xu and Nyerges, 2016). The magnitude and the severe impact on populated areas, led initially to a local response that was crucial to support rescuing operations and to guide the search of missing people. At a second stage, Twitter was also central to organise international humanitarian aid.

The Tsunami in Japan (Acar and Muraki, 2011), as well as the flooding in Queensland (Bruns, et al., 2012), and Hurricane Sandy in Usa (2012) are just some of the several examples in which Twitter has played a key role.

Social networks have also been strategic to support the response phase for other types of risks.

In September of 2013, a group of terrorists laid siege for four days a shopping center in Nairobi, Kenya, killing 67 people and injuring 175. During the event, Twitter was the media used by governmental forces to communicate the evolving situation and invite people to stay away from the mall (Simon et al. 2014).

In the terroristic attacks in Paris (13th November 2015), Twitter, has been a point of reference for many people during and after the event. The citizens of Paris offered hospitality to the desperate people seeking shelter, through the messages posted on the platform and labeled with the hashtag '#PorteOuvertes' (see Fig. **2.29**).

Fig. 2.29 Examples of Tweets

2.4.4 Characteristics of Message Content During an in Event

From the previous section it is evident how social media constitutes a valid and reliable source of information at all stages of an event (Response, Recovery, Mitigation, Preparedness).

This is acknowledged by the actors involved in emergency activities (local and international organizations), which are aware that social media is a central tool to collect and disseminate timely information about the event (Vieweg, 2010).

According to a recent study by Imran et al. (2015) it is possible to identify the main factors influencing the content of messages during an event. In particular, the content of a message may be characterized by several factors given in the **Tab. 2.7** and described below:

- By application type: This considers the social network or messaging service used. Many applications set limits to the number of characters to use in a text message (140 characters in the case of Twitter). In this case, the user tends to delete words, namely 'empty word' such as articles or prepositions, or shortening long words.
- By event type and information provided: The type of the event and its magnitude affect the content of the message. A natural disaster such as an earthquake, is different from a terrorist attack. Both have different impacts on people and objects, generating different needs. As for the same type of events, instead, it is the magnitude of the event that affects people and, consequently, the content of the messages.
- By factual, subjective or emotional content: The message may report objective facts or subjective and emotional contents. Objective contents are experienced by the person who writes the message; subjective contents include personal and emotional comments about the affected places. It should be emphasised that the emotional aspect under- or overestimates the magnitude of the event and its effects.
- By information source: Messages can be posted by individuals playing different roles in disaster response. These people can be directly affected by the event, or can be volunteers to the rescuing operations or participate to fundraising. They could be mere observers (followers) of tweets or write on behalf of organizations to provide useful

information for coordination and organizational issues (such as telephone numbers of or locations for fund raising).

- By credibility: This feature refers to actors (individuals and organizations) that can be associated with a degree of reliability to the information contained in the message, which in turn, can be more or less credible or plausible. False or incorrect information can trigger widespread mis-information due to replication of the news through the network.
- By location: The message, although relating to the event, can be sent by people who are nearby or far away to the area of the event. The perception and knowledge of the event affects the degree of knowledge of the event itself.
- By time: The phases of an event affect the content of the message. Messages sent ex-ante to an event have a different content than messages sent ex-post. The latter focuses on the request for help, while the former on reporting the probability of the occurrence of the event.

Tab. 2.7 Classification of contents posted on social media during high impact events with description and related work references

Classification dimension	Description/examples
By factual, subjective, or emotional content	
Factual information	(Examples under 'By information provided')
Opinions	opinions, criticism (e.g. of government response)
Sympathy	[Kumar et al. 2013]; **condolences** [Acar and Muraki 2011]; **support** [Hughes et al. 2014b]; **thanks, encouragement** [Bruns 2014]; **prayers** [Olteanu et al. 2014]
Antipathy	**schadenfreude, animosity against victims** (e.g. because of long-standing conflict)
Jokes	**jokes, trolling** [Metaxas and Mustafaraj 2013]
By information provided	
Caution and advice	**caution and advice** [Imran et al. 2013b]; **warnings** [Acar and Muraki 2011]; **hazard, preparation** [Olteanu et al. 2014]; **tips** [Leavitt and Clark 2014]; **advice** [Bruns 2014]; **status, protocol** [Hughes et al. 2014b]
Affected people	**people trapped, news** [Caragea et al. 2011]; **casualties, people missing, found or seen** [Imran et al. 2013b]; **self reports** [Acar and Muraki 2011]; **injured, missing, killed** [Vieweg et al. 2010]; looking for missing people [Qu et al. 2011]
Infrastructure/utilities	**infrastructure damage** [Imran et al. 2013b]; **collapsed structure** [Caragea et al. 2011]; **built environment** [Vieweg et al. 2010]; **closure and services** [Hughes et al. 2014b
Needs and donations	**donation of money, goods, services** [Imran et al. 2013b]; **food/water shortage** [Caragea et al. 2011]; **donations or volunteering** [Olteanu et al. 2014]; **help requests, relief coordination** [Qu et al. 2011]; **relief, donations, resources** [Hughes et al. 2014b]; **help and fundraising** [Bruns 2014]
Needs Services and Other useful information	**hospital/clinic service, water sanitation** [Caragea et al. 2011]; **help requests, reports about environmen**t [Acar and Muraki 2011]; **consequences** [Olteanu et al. 2014]
By information source	
Eyewitnesses/Bystanders	**members of public** [Metaxas and Mustafaraj 2013], **victims, citizen reporters, eyewitnesses** [Diakopoulos et al. 2012; Olteanu et al. 2014; Bruns 2014]
Government	**administration/government** [Olteanu et al. 2014]; **police and fire services** [Hughes et al. 2014b]; **governmen**t [Bruns 2014]; **news organization and authorities** [Metaxas and Mustafaraj 2013]
NGOs	non-government organizations [De Choudhury et al. 2012]
News Media	**news organizations and authorities, blogs** [Metaxas and Mustafaraj 2013], **journalists, media, bloggers** [De Choudhury et al. 2012]; **news organizations** [Olteanu et al. 2014]; **professional news reports** [Leavitt and Clark 2014]; **media** [Bruns 2014]
By credibility	
Credible information	**newsworthy topics, credibility** [Castillo et al. 2013]; **credible topics** [Canini et al. 2011]; **content credibility** [Gupta and Kumaraguru 2012]; **users and content credibility** [Gupta et al. 2014]; **source credibility** [Thomson et al. 2012]; **fake photos** [Gupta et al. 2013]
Rumors	**rumor** [Hughes et al. 2014b; Castillo et al. 2013]
By time	
Pre-phase/preparedness	posted before an actual event occurs, helpful for the preparedness phase of emergency management [Petak 1985]; pre-disaster, early information [Iyengar et al. 2011; Chowdhury et al. 2013]
Impact-phase/response	**posted during the impact phase of an event, helpful for the response phase of emergency management** [Petak 1985]; **during-disaster** [Iyengar et al. 2011; Chowdhury et al. 2013]
Post-phase/recovery	**posted after the impact of an event, helpful during the recovery phase** [Petak 1985]; **post-disaster information** [Chowdhury et al. 2013; Iyengar et al. 2011]
By location	
Ground-zero	**information from ground zero** (victims reports, bystanders) [De Longueville et al. 2009; Ao et al. 2014]
Near-by areas	**information originating close to the affected areas** [De Longueville et al. 2009]
Outsiders	**information coming from other parts of world, sympathizers** [Kumar et al. 2013]; distant **witness** (in the sense of [Carvin 2013]); **not on the ground** [Starbird et al. 2012]; **location inference** [Ikawa et al. 2012]

3 Research Design and Methodology

The conceptual background section described in the previous chapter focused on three main domains: risk, local knowledge and new e-participation tools.

This chapter illustrates: i) how the three topics previously described relate to each other; and ii) the methodology used to analyze these relationships. As for the latter, the description focuses on the applied literature for risk in post-disasters domain.

Fig. 3.1 shows the methodological framework of the present Thesis. The figure highlights 3 sections. Each section comprises of a definition of risk and its declinations and of a non-structured dataset. The latter serves to study local knowledge which provides for decision support systems in risk domains.

The understanding of local knowledge in disaster risk is based on a combination of social sensing and machine learning approaches. The former includes both structured information from public participation retrieved from web 2.0 platforms such as PPGIS, Ushahidi, or VGI to cite a few and unstructured data from social networks as Facebook, Twitter and many others. These form the social representation containing useful information from people's perception about type, extent, intensity, impacts and emergencies in disaster response. The latter includes a machine learning approach which is based on information extraction to obtain the final dataset to compute the predictive model. The results obtained by the predictive model feedback onto the social sensing context to form the SDI and enrich both the knowledge of the public and that of the expert. Next, a detailed description of the conceptual model is offered.

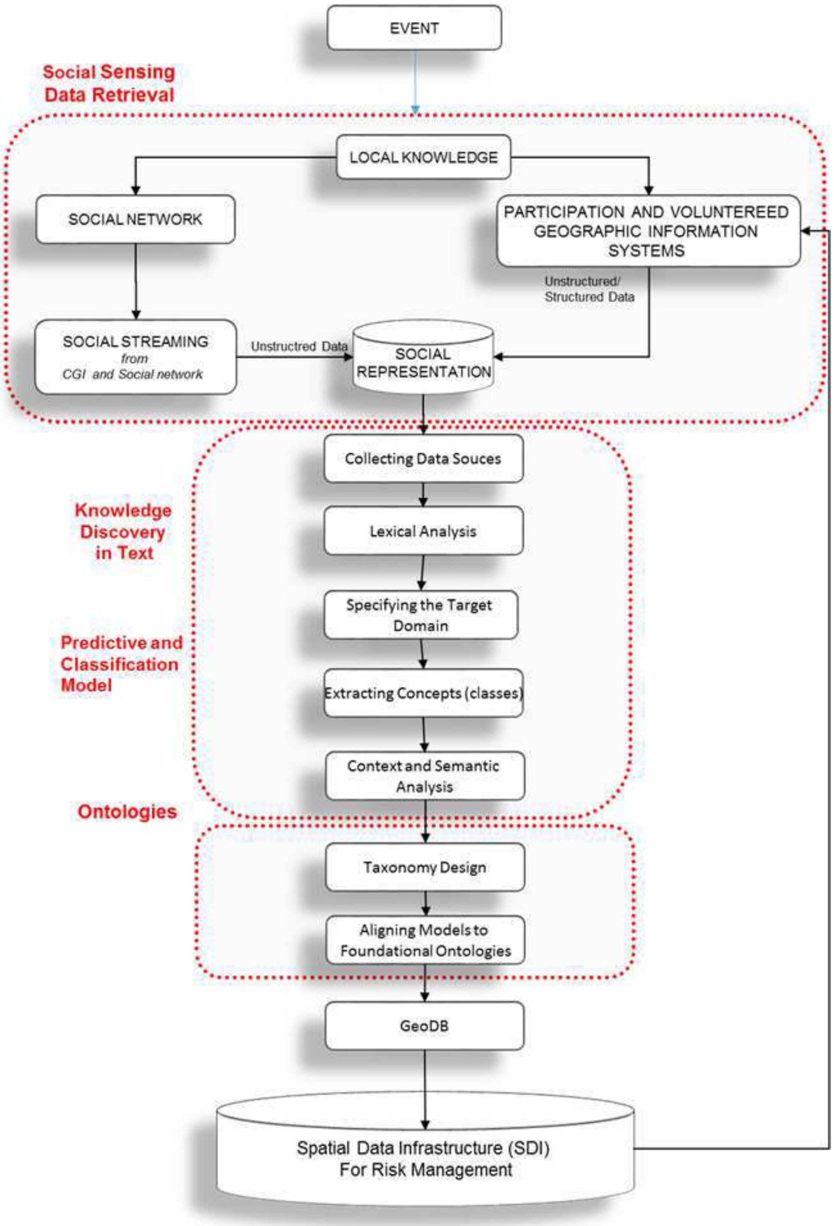


Fig. 3.1 Methodological framework

3.1 Application of the Methodology to Disaster Events

The present section illustrates the relevance of data acquisition as a tool to detect local knowledge and set up the modelling framework in disaster response domain.

The international community responds to a disaster event according to four main steps as identified in **Section 2.1.3** These steps involve Response, Recovery, Mitigation and Preparedness (Platt, 1999; Poser et al, 2010; R. Kirillov et al. 2012) **Tab. 3.1** adapts, from a paper by Field (2012), the main actions undertaken during the above phases[26].

Tab. 3.1 Activities associated with the impacts of disasters in pre, post and recovery phases

A - Pre-Event - Preparedness	B - Post Event - Response	C – Recovery
1. Public education 2. Awareness raising 3. Warning and evacuation plans 4. Pre-positioning of resources and supplies 5. Last minute alleviation and preparedness measures	1. Search and rescue 2. Emergency medical treatment 3. Damage and Needs Assessment 4. Provision of services – water, food, health, shelter, sanitation, social services, security 5. Resumption of critical infrastructure 6. Coordination of response 7. Coordination/ Management of development partner support	1. Transitional shelter in form of temporary housing or longterm shelter 2. Demolition of critically damaged structures 3. Repair of less seriously damaged structures 4. Clearance, removal, and disposal of debris 5. Rehabilitation of infrastructure 6. New construction 7. Social rehabilitation 8. ‘Building back better’ to reduce future risk 9. Employment schemes 10. Reimbursement for losses 11. Reassessment of risks

Source: adapted from (Field, 2012)

In addition, **Tab. 3.2** shows the activities associated to an event based on the indications provided in **Tab. 3.1**. For example, for a natural event it is possible to combine preparedness actions such as public education (A1) and awareness raising (A2); and response actions such as search and rescue (B1), emergency medical treatment (B2), damage and needs assessment (B3) and provision of services (B4).

[26] The mitigation phase is not considered in the analysis of Tab. 3.1.

Tab. 3.2 Activities associated with the type of event investigated in this study

Event	Activities Associated
Geopolitical Conflict	A1 – A2 –B1 – B2 – B3 – B4
Political Violence	A1 – A2 – B1 – B2
Natural Event	A1 – A2 – B1 – B2 – B3 – B4
Climatic Event	A1 – A2 – B1 – B2 – B3 – B4
Environmental Event	A1 – A2 – B1 – B2 – B3 – B4
Technological Event	A1 – A2 – B1 – B2 – B4
Humanitarian Crisis	A1 – A2 – B1 – B2 – B3 – B4

As shown in **Tab. 3.2**, the proposed methodologies are applied to two event phases such as Preparedness and Response and to six activities such as A1 and A2 (Preparedness) and B1, B2, B3, B4 (Response). Finally, for Natural Event and Technological Event no B3 and B4 activities are identified.

3.2 Social Sensing - Data Retrieval

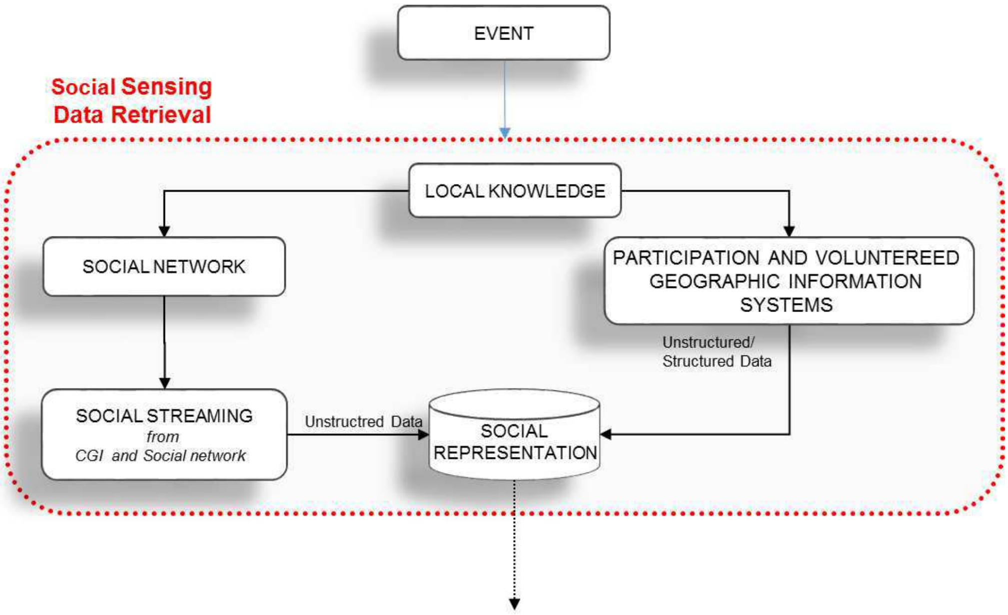


Fig. 3.2 Methodological framework: Data Retrieval

As already mentioned in **Section 2.4.1**, recent developments of e-participation through VGI technologies and crowdsourcing initiatives have witnessed data retrieval for numerous applications through mobile devices. Several platforms such as Ushahidi are active in supporting disaster response and are seen as best practices for e-participation.

Information retrieved through a social network can be stored in a database. However, some limitations exist due to memo's fields in which information is stored as text with no input-constraints. A further limitation of the social representation is that the user should get to know with the application and install it on his/her mobile device. To solve this problem several platforms add new modules and link these ones to social network, to capture further information and data. This process is attainable through application programming interface

(API)[27] which are dedicated libraries between the platform and the social network (e.g Twitter, Facebook). The social streaming captures, saves and stores text messages containing keywords such as for example ‘earthquake’ with the corresponding indication of location.

The present Thesis assumes that the user is accustomed to at least one of the most common social networks to exchange information, including requests for and offers to help in disaster response. Nonetheless, there exists an ethical problem with social network streaming processing. Users may not want or may not be aware that their messages can undergo a streaming process and stored in databases. In disaster events, it is evident that the user is willing that his/her text reaches as many people as possible to increase the chances that his/her text may be read by a wider audience.

Social streaming can be considered the latest development to data and information retrieval. Should this be suitably contextualised, it opens new research opportunities to public participation.

How to treat data with no input constraints from social streaming? The next sections will deal with specific methodologies to retrieve structured knowledge from unstructured one.

A first aspect provides a simulation analysis with the use of PGIS technologies. These platforms (e.g. Ushahidi) retrieve geo-referenced data either directly (through specific functions of the platform) or indirectly (through social network streaming processing) about generic territorial information (Fig. 3.3).

[27] In Computer Science, an API is a set of available procedures and tools to execute a function or a set of funtions. Generally API are referred to software libraries available in specific programming language.

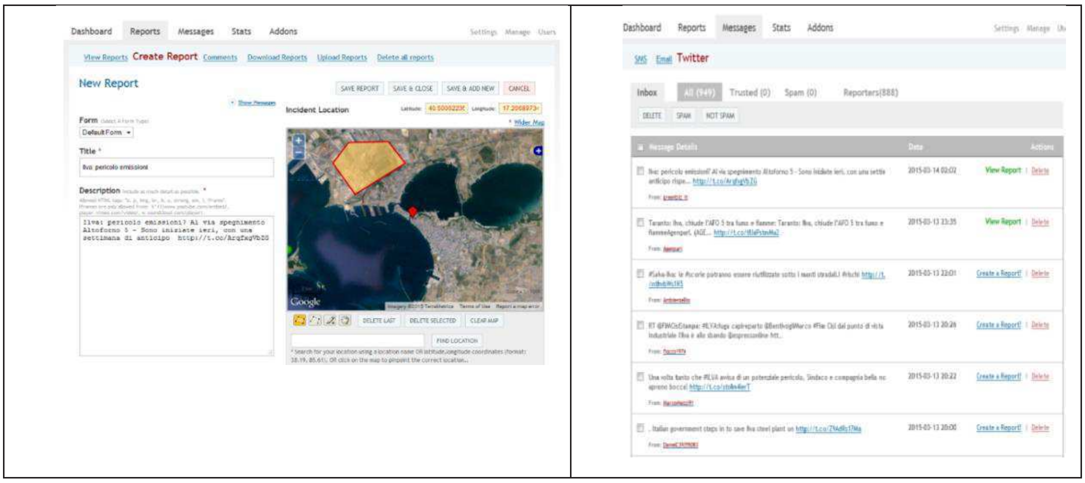
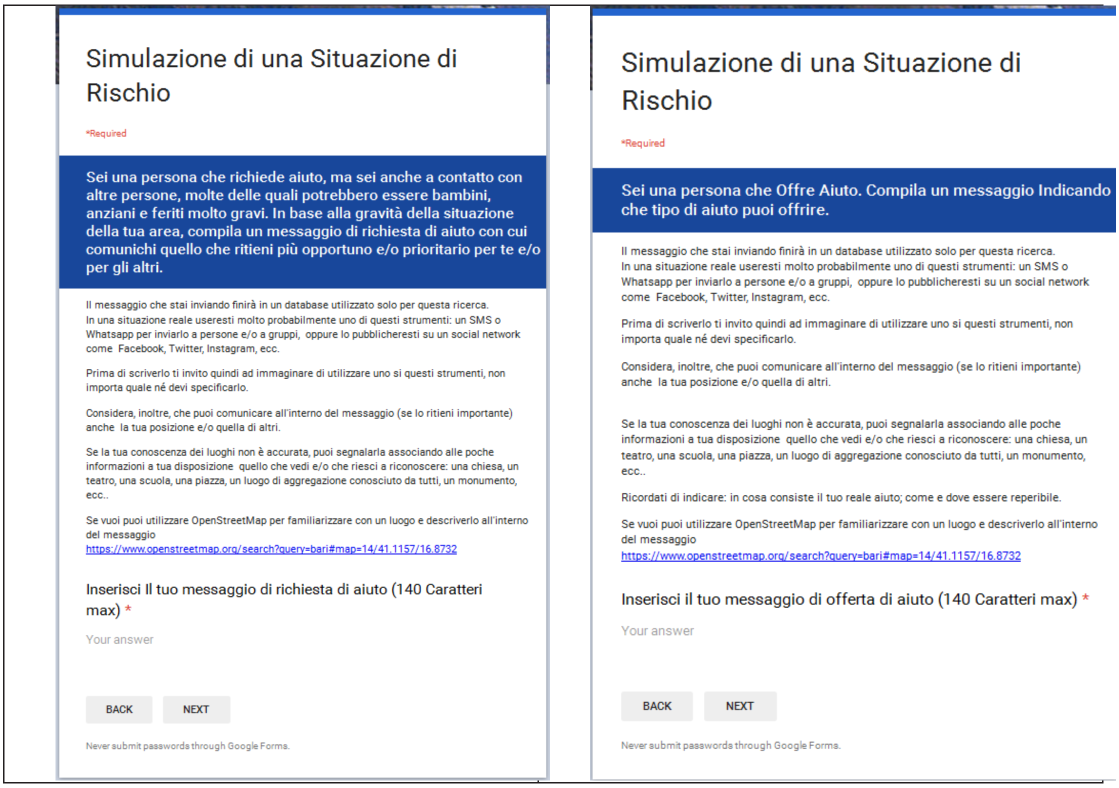


Fig. 3.3 Ushahidi platform

Simultaneously, an open-ended web survey questions is organized (**Fig. 3.4**). The on-line survey, after illustrating a hypothetical earthquake in the metropolitan area of Bari, asks participants to write request for or offer to help messages (max 140 characters). A subset of 314 observations contains messages from the post-earthquake domain.



Simulazione di una Situazione di Rischio

*Required

Sei una persona che richiede aiuto, ma sei anche a contatto con altre persone, molte delle quali potrebbero essere bambini, anziani e feriti molto gravi. In base alla gravità della situazione della tua area, compila un messaggio di richiesta di aiuto con cui comunichi quello che ritieni più opportuno e/o prioritario per te e/o per gli altri.

Il messaggio che stai inviando finirà in un database utilizzato solo per questa ricerca.
In una situazione reale useresti molto probabilmente uno di questi strumenti: un SMS o Whatsapp per inviarlo a persone e/o a gruppi, oppure lo pubblicheresti su un social network come Facebook, Twitter, Instagram, ecc.

Prima di scriverlo ti invito quindi ad immaginare di utilizzare uno si questi strumenti, non importa quale né devi specificarlo.

Considera, inoltre, che puoi comunicare all'interno del messaggio (se lo ritieni importante) anche la tua posizione e/o quella di altri.

Se la tua conoscenza dei luoghi non è accurata, puoi segnalarla associando alle poche informazioni a tua disposizione quello che vedi e/o che riesci a riconoscere: una chiesa, un teatro, una scuola, una piazza, un luogo di aggregazione conosciuto da tutti, un monumento, ecc..

Se vuoi puoi utilizzare OpenStreetMap per familiarizzare con un luogo e descriverlo all'interno del messaggio
https://www.openstreetmap.org/search?query=bari#map=14/41.1157/16.8732

Inserisci Il tuo messaggio di richiesta di aiuto (140 Caratteri max) *

Your answer

BACK NEXT

Never submit passwords through Google Forms.

Simulazione di una Situazione di Rischio

*Required

Sei una persona che Offre Aiuto. Compila un messaggio Indicando che tipo di aiuto puoi offrire.

Il messaggio che stai inviando finirà in un database utilizzato solo per questa ricerca.
In una situazione reale useresti molto probabilmente uno di questi strumenti: un SMS o Whatsapp per inviarlo a persone e/o a gruppi, oppure lo pubblicheresti su un social network come Facebook, Twitter, Instagram, ecc.

Prima di scriverlo ti invito quindi ad immaginare di utilizzare uno si questi strumenti, non importa quale né devi specificarlo.

Considera, inoltre, che puoi comunicare all'interno del messaggio (se lo ritieni importante) anche la tua posizione e/o quella di altri.

Se la tua conoscenza dei luoghi non è accurata, puoi segnalarla associando alle poche informazioni a tua disposizione quello che vedi e/o che riesci a riconoscere: una chiesa, un teatro, una scuola, una piazza, un luogo di aggregazione conosciuto da tutti, un monumento, ecc..

Ricordati di indicare: in cosa consiste il tuo reale aiuto; come e dove essere reperibile.

Se vuoi puoi utilizzare OpenStreetMap per familiarizzare con un luogo e descriverlo all'interno del messaggio
https://www.openstreetmap.org/search?query=bari#map=14/41.1157/16.8732

Inserisci il tuo messaggio di offerta di aiuto (140 Caratteri max) *

Your answer

BACK NEXT

Never submit passwords through Google Forms.

Fig. 3.4 Survey forms

a) Structured knowledge: It uses a specific app to collect information which is saved in a database. The app is designed through a dedicated web interface where the user inputs his/her own message, then it is possible, in real time, to geo-localize the post on an interactive map and classify it in a given category.

As for data analysis and machine learning, the present work is based on archived data available in the fields of 'Description' and 'Category'. The former contains the description of the message; the latter its classification.

b) Unstructured (see **Tab. 3.3** *and* **Tab.** 3.4). This type of knowledge uses information retrieval based on social network streaming which, through specific API functions, filters messages from Facebook and/or Twitter based on hash tags. The posts retrieved as such present some drawbacks. They should be validated and classified at a later stage by appointed experts. Therefore, the timing of these operations can last several days and suggest inefficiencies, should the community proceed with the elaboration of information in disaster response. The present Thesis attempts to overcome to this limitation. The assessment of social network streaming processing can save human and animal lives and the ecological system, should these assessed information be promptly read and /or observed by the rescue team.

Tab. 3.3 A sample dataset of the earthquake simulation

COD	DESCRIZIONE
Q	sono sotto ponte via cavour bari serve cibo e acqua
Q	Aiuto!Sono davanti la cattedrale di san sabino a Bari.Siamo 4 adulti di cui 2 anziani e 1 bambino di 3 anni.Abbiamo provviste per altri 2 gg
Q	sono rimasto bloccato nel sottopassaggio di Via Quintino Sella con la cinquecento, il livello dell'acqua sale e mia moglie mi picchia
Q	ho degli hamburger col ketchup nel tre ruote, li vendo a soli 10 euro l'uno, vista la situazione, servizio a domicilio
Q	Ci sono due donne (di cui una anziana) con tre bambini vicino la chiesa di Santa Lucia che hanno bisogno di viveri
Q	Sono nel borgo antico, posso offrire cibo, acqua e coperte
Q	Disponibile a soccorrervi: datemi coordinate gps
Q	Posso offrire sostegno per fornitura e consegna di beni di prima necessità
Q	Offro il mio supporto tecnico
Q	vi prego corrette ho due bambini con me nella palestra palacarrassi
Q	Sono a bari, San Girolamo, vicino canalone, ci sono persone bloccate sotto le macerie, si stimano 12 adulti e 3 bambini,si richiede aiuto.
Q	Una bambina col suo fratellino sono sotto le macerie affianco alla Cattedrale. Stanno bene ma hanno tanta paura e tanto freddo !!
Q	Ci stiamo Attivando per aiutarvi, per cercare di risolvere questa tragedia
Q	Siamo isolati all'interno di un palazzo crollato nei pressi dello stadio. Non c'è luce e fa molto freddo.
Q	Mi trovo in auto in corso Cavour davanti a me c'è un'altra macchina schiacciata dalle macerie dell'edificio adiacente sento lamenti
Q	Siamo circa una cinquantina di persone di tutte le età, siamo negli edifici vicino l' ospedale militare Via Petraglione e ci sono feriti.
Q	sono vicino alla Coop Via Fanelli, posso aggregarmi ad altri soccorritori in zona
Q	IMPORTANTE!!! All'angolo tra corso Cavour e via Putignani c'è un amico con una gamba bloccata da una maceria. Chi può venire ad aiutarci?
Q	Sono con altre persone vicino ad una chiesa
Q	Mi trovo vicino il Castello e attorno a me sento dei lamenti, io riesco a muovermi

Tab. 3.4 A sample dataset of general territorial information

COD	DESCRIZIONE
S	tutto il patrimonio, di valore storico ed artistico, della città vecchia reso pubblico
S	un porto turistico e commerciale e non più industriale
S	un recupero identitario che miri a potenziare l'attravità del centro storico.
S	una città vecchia ripoplata con una forte vocazione turistica e artistica basatasui mestieri legati al mare
S	una città vecchia sperimentale nelle pratiche legate alla sostenibilità e i centri storici
S	utilizzare i palazzi antichi come laboratori culturali per divulgazione della cultura antica, le tradizioni ,il mare e il paesaggio
S	utilizzare viabilità via mare per raggiungere la città vecchia
S	valorizzare centro storico (ristrutturare creare luoghi di aggregazione
S	valorizzare i siti archeologici come laboratori universitari per le materie idonee alla storia del territorio
S	valorizzare la città vecchia organizzando eventi e attrazioni riguardanti la cultura del territorio e la storia
S	valorizzazione della rete di ipogei e dei beni archeologici della città vecchia. apertura di un museo diffuso dell'acropoli
S	zona porta napoli, cementir... costruire una piastra intermodale, punto di convergenza sia per un grande snodo stradale per tradizioni fe
S	conversione dell'economia del territorio da industriale ad agricola tecnologica, turistica, energia rinnovabile
S	abbattimento mura arsenale e restituzione dlle zone prospicienti alla citta' (ospedale militare, villa ammiraglio)
S	acqua potabile
S	adeguamento sistema e programmazione sanitaria in linea con i problemi derivanti dall'inquinamento industriale

3.3 Tools for Data Analysis

The subsequent data analysis is performed with the use of RapidMiner Studio v. 7.0[28]. This is an open source software that implements models such as: segmentation, association, correlation and prediction. It is a framework which supports various types of data including text, that can parse through numerous text mining functions. It finds application in recent literature (Verna, 2014) and presents the following advantages: i) an immediate graphical user interface for input and output processes; ii) the handling of data from several formats; iii) a comprehensive text mining module; iv) the ability to apply several model predictions.

Thanks to its advanced graphical user interface, data mining can intuitively be built (**Fig. 3.5**) However, as it is freely implemented in Java, the software can integrate plug-in processes developed by other users. Also, it offers an API that can be used by other software as Java library.

RapidMiner has performed data cleaning operations for texts analysis.

The use of RapidMiner favours the simulation of several operations that could be integrated, in the near future, in PGIS and VGI platforms and more generally on decision support systems to develop, simultaneously, data streaming of risk events.

[28] http://www.rapidminer.com

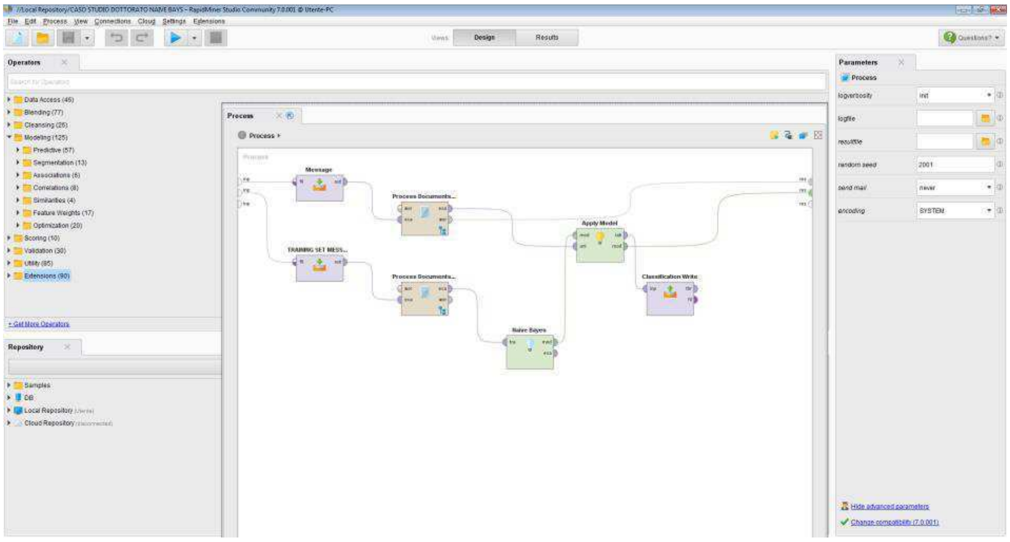


Fig. 3.5 Rapidminer interface

3.4 Text Analysis

As already described in **section 3.4.3** the content of messages, although quite heterogeneous, includes common characteristics (e.g. **Tab. 2.7**) that are suitable for clustering and classification analyses. Straightforward, two steps and two methods are taken into account.

Step 1: Identification of two types of methods which allow to reorganize the information in emergency situations, through supervised algorithms instructed by a training set.

Step 2. Processing of Information Extraction aiming at limiting the text format due to conceptual entities (needs, geo-location, people) (Liu et al., 2011; Ritter et al. 2011, Imran, 2015). The extracted concepts will be structured by an ontological analysis; and second by a shared spatial data infrastructure. For example, a typical post-event message would be: 'We need blankets, we are 4, 2 kids, we are in Garibaldi Garden (Abbiamo bisogno di coperte, siamo in 4, ci sono 2 bambini, siamo in Giardino Garibaldi).

The identified concepts are the following:

- affected/exposed = 2 adults and 2 children;
- report-type = help request;
- spatial location = Garibaldi Garden (Giardino Garibaldi);
- needs = blankets.

The simple structure information as presented above is easily managed by information systems for decision support.

Message extraction including a conceptual approach is carried out with different types of textual analysis such as lexical, syntactic and semantic analysis.

The present Thesis identifies four conceptual domains generally included in a message in disaster response events: needs, spatial location, actors, timing.

Each of these concepts requires an in-depth analysis for the construction of linguistic patterns that take into account knowledge and common sense related to the places where the event takes place. A particular concern is the use of natural language with reference to the spatial location to understand local knowledge. Natural language uses terms and combinations of terms often unknown outside certain local/spatial contexts.

The existence of a natural language which creates information and supports local knowledge in text analysis is one of the focal points of this Thesis. The sharing and understanding of local knowledge is the main requirement of an information system at a global level (e.g. Humanitarian crises). On this premises, local knowledge should require and deal with ontological models.

3.4.1 Knowledge Discovery in Text

Section 2.2.1 considers the Ackoff's (1989) pyramid to illustrate the concept of knowledge. This is used in several computer science studies. Mathematical modelling and other scientific disciplines refer to *Data* as the main level to introduce and define knowledge. If, on the one hand, web data availability favours the spread of information, on the other, new drawbacks arise in terms of management of unstructured data and its reliability.

In this complex scenario, automated intelligent systems finds a place. These systems, based on certain filters, search and select the information needed; in a second stage, the collected information is stored in different formats and is available for any purpose (e.g. statistical, ontological). The main feature of these systems, namely Knowledge Discovery in Text (KDT), is sharing and interoperability of data availability.

KDT aims at detecting and dismissing data (noise) which is not useful to the purposes of the platform. It provides 'extraction' of latent knowledge (Swanson,1991).

To understand latent knowledge it is useful to shed light on Text Mining (TM), Text Data Mining (TDM) and KDT.

There is not substantial difference across these terms. TM and TDM are generally used in marketing and business; KDT is used in academics (Dulli, 2004). Therefore, KDT will be used in the present Thesis.

It is also important to note the difference between KDT and Natural Language Processing (NLP) often used improperly. In particular, NLP refers to 'Language Understanding' in automated methods and can be used for a single occurrence; KDT is corpus-based and oriented to analyse several documents to provide interpretation and knowledge.

Prior to KDT, Knowledge Discovery in Database (KDD) was largely used. KDD is defined as "*the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*" (Fayyad et al.,1996:30).

Finally, automated intelligent systems appear valid tools to structure unstructured text and understand local knowledge, no matter which system a platform builds-in.

KDT Phases

KDT o TM (**Fig. 3.6**) is applied to any corpus of documents and is mainly designed to:

- Identify thematic groups
- Extract concepts for taxonomies and ontologies
- Classification
- Discovering hidden associations
- Extract specific information (i.e. addresses)

Usually it consists of four main phases: Information Retrieval (IR), Information Extraction (IE), Information Mining (IM), Interpretation (I).

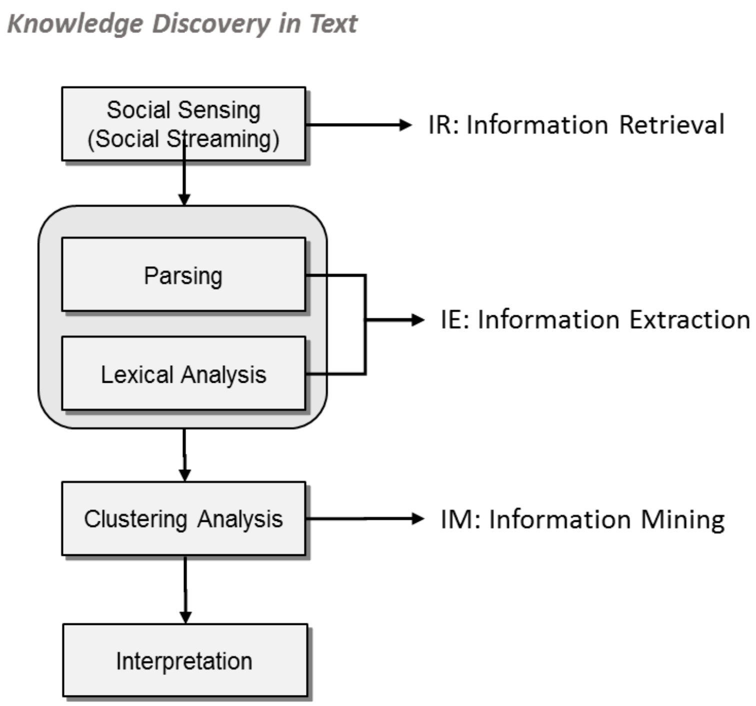


Fig. 3.6 Text mining phases

Information Retrieval

It is the first phase in which the texts are identified and from which it is possible to extract information. This process has been already introduced in **Section 3.2**

Information Extraction (IE)

It extracts information from a text and generally encodes it into vectors or matrices to be processed for further analysis. The main steps for text encoding are as follows:

- Lexical Analysis:
  - Language identification
  - *Tokenization,* operation by which the text is divided into tokens (words, dates, numbers, symbols, punctuation, etc.).
  - Morphological analysis;
  - *Part-of-speech tagging*. Words are marked by assigning a grammatical category;
- Recognition of lexical structures: Idioms, names, numbers, dates, phrases. Their identification is performed by a series of patterns (regular expressions), orthographic and syntactic features (i.e. a capital letter);
- Pruning: It is used to drop non-interesting elements (i.e. ‘empty’ words) and thus lighten the processing by means of filters;
- Syntactic analysis: It is the identification of syntactic links between the different elements of the phrase. This analysis recognises noun phrases or verbal groups;
- Pattern matching: It extracs terms from the context of important relations.

Information Mining (IM)

This phase extracts knowledge from texts with several different methods, some of which are shown in **Sections 3.4.**

As already described in the previous sections, a simple operation that can save lives in a disaster response domain is to extract high priority posts leaving, at a second stage, other messages containing the request for minor emergencies. To do so, structured and unstructured knowledge is retrieved from social networks and the Ushahidi platform.

The main aim of machine learning in the present Thesis is to observe, learn and classify local knowledge (Bishop, 2006) through proper algorithms and make this knowledge available on an SDI as well as an expert knowledge system.

Pre-processing activities (e.g. text-mining) are required to obtain an unbiased data entry matrix, run the predictive model and extract adequate entities to set up the ontological analysis.

3.4.2 Information Extraction and Lexical Analysis

'*The phase of lexical analysis provides a paradigmatic representation of the corpus: the study of its vocabulary, i.e the language*' Bolasco (2005; pp. 9).
The analysis of a lexicon of a 'corpus' is the extraction and the construction of a lexicon of a 'corpus' without considering the merits of the speech articulation. The approach, usually based on statistical analysis, aims to observe a text which is then split into parts (words) containing nouns, verbs and adjectives, and parts containing junctions, prepositions, connectives and punctuation. Then, descriptive statistics (e.g. absolute and relative frequencies) is performed on both parts as well as other indicators (e.g. TF/IDF) which serve to support the existence of important details and significant words within the speech.

As already mentioned in the previous section, machine learning techniques are used to perform a lexical analysis as follows: i) *Tokenization* allows to select the main words (token) included in a document; ii) *Stopwords* allow to drop all irrelevant words listed in the stopword dictionaries (i.e. Italian dictionary); iii) *Replacing tokens* replaces compound words with single words; and iv) *Stemming* reduces the number of words that have in common the same root in a single token (Verma, 2014) or more tokens (n-gram) (see **Fig. 3.7**).

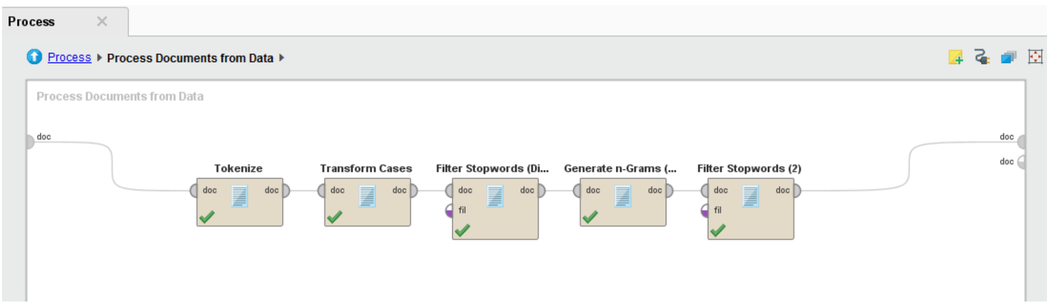


Fig. 3.7 Rapid Miner information extraction pipeline

The output of the text-mining approach is of two types.

1) The first one is a vector of lexical elements. The elements of the vector could be formed by a single word or several words. The latter depends on the data contained after setting the n-gram parameter. The analysis of an n-gram can help to better understand several aspects of a language in relation to a particular domain. To express the meaning of specific texts a single word (e.g. 'Help') could be enough. Problems arise for more complex documents containing indications relative to a specific spatial location which generally needs more than one word to express a suitable meaning.

The rationale of an additional lexical analysis is to deepen the results obtained with the descriptive statistics and text-mining approaches. The lexical analysis aims at understanding the forms of languages used in a disaster response domain.

After tokenizing operations, each token is looked up in a dictionary (which includes) to determine 'the part-of-the-speech' and assign a class to each token (nouns, verbs, adjectives, adverbs, pronouns, prepositions). The next step, namely POS tagging, assigns through the use of a dictionary, specific POS tags to each word and applies the disambiguation rules created and other features.

2) The second one is a data matrix to run the predictive model. The columns of the matrix represent the parts of the document with a specific meaning (i.e. nouns, verbs and adjectives); the rows (the observations) identify each text message in which 1 indicates the presence of the lemma and 0=otherwise.

3.4.3 Lexical Analysis: Understanding Local Knowledge of Spatial Locations from Natural Language

Suppose one wants to extract information relating to the spatial location of a person who asks for help by sending a message on a social network. Messages captured via streaming processing could potentially incorporate the geographic location [29]. Recent literature suggests that 0.42% of posts contain GPS information (Sui et al., 2014).

Information on the spatial or geographical location is provided directly (e.g. geographic coordinates, addresses) or indirectly by natural language (e.g. indication of known places). The latter case also arises when additional topological indications are provided such as a position relative to a known place or object or area: For example 'Help I am next to Saint Sabino church'. In this case 'church' (an object) defined as non-delimited space corresponds to thousands points on a map. 'Saint Sabino' is an instance of the church and its features make the church unique in a confined space.

The spatial location is not always reported in a text message. This is because the user, in most cases, assumes that the reader knows about an event and the place where it has happened.

Much more complex is the concept related to spatial relations, which, in the example above, is defined by the string 'next to'. The complexity arises because the word 'next to', often considered as an 'empty word' and dropped by the stopwords process, is central in disaster response.

A key aspect is the dictionary in which the stopwords are listed. Usually several software propose default dictionaries containing a list of empty words. However, there exists empty words that combined with others contribute to define a meaning.

In the Italian language, this is particularly true for the case of simple prepositions: 'di, a, da, in, con, su, per, tra, fra' and articulate prepositions such as: 'dal, dalla, dagli, dallo, del'. Similarly, for 'improper prepositions' such as: 'davanti (davanti o davanti a?), dietro (dietro o dietro a?), dopo, fuori, lontano, lungo, mediante, prima (prima che o prima di?), sopra

---

[29] This is true in the case in which the message is sent from a device equipped with GPS and the application used to sending the post is turned on. The users authorises the activation of the application.

(sopra o sopra a?), sotto (sotto o sotto a?)'. Prepositions can also be used with other grammatical roles (adjectives, verbs or adverbs).

In the example above, if one considers a sequence of words where the preposition 'next to' is placed between two nouns, he/she obtains: 'The house next to the church'. Should the word 'to the' be considered as non-important, the remaining word 'next' could be misinterpreted as an adjective of 'house'. This problem could cause loss of spatial location content in the text message.

Similarly, this is the case for simple prepositions such as 'da' (from) and 'a' (to) that from a spatial point of view indicates a direction.

The aspect of spatial location from natural language to understand local knowledge is attentively analysed in **Section 4.2**.

3.4.4 Knowledge Classification

With the growing development of large datasets stored in electronic form in various web platforms is difficult to retrieve information when data is unstructured. After performing pre-processing activities with the use of text-mining techniques to obtain a data entry matrix, the following machine learning models are generally used to classify text documents.

These techniques are required for those processes extracting, automatically or semi-automatically, unknown information from unstructured data which identify clusters and research topics trends (Delen and Crossland, 2008).

In the present Thesis machine learning classification models are used to classify the contents of the message as ‘emergency’ and ‘non-emergency’. The presence of ‘non-emergency’ content messages relates to queries of social content, with light restrictive filters [30].

The case study of the present Thesis uses a dataset containing several text messages not entirely related to a disaster event. One of the main purposes is to check whether the application of supervised classification techniques through a specific training set is statistically accurate.

Another aspect is to check the existence of situational and non-situational messages (Verma et al., 2011; Rudra et al., 2015):

- Tweets containing situational information are mainly of two types: (i) Updates on the event situation given by governmental agencies, organizations, mass-media (i.e. the number of casualties, extent of damage); (ii) Requests for help, offers of help and information that can help with the rescuing operations (i.e. hospitals phone numbers). Both types can be sent from affected people or people living outside the place of the event.
- Non-situational Tweets are generally of the following types: (i) Feelings, views, opinions; critics to the relief efforts, opinions on how similar accidents can be avoided

[30] The filtering operations are possible through the stream of keywords combined with each other (with logical ‘or’ or ‘and’). The use of very restrictive filters, through very complex keys, reduce this problem but increase the risk of excluding messages which nevertheless relate to the event.

over time; (ii) The analysis of events - post-analysis of how and why the disaster occurred; (iii) Charities-related information provided by specific bodies to help with affected populations. (Rudra et al., 2015).

The choice of supervised classification algorithms depends on the type of problem to explore. Adaptive Boosting, Naïve Bayes, Neural networks, Support Vector Machines (SVM), Random Forests are generally most used in current literature and also applied to the case study (see **Section 3.4.5**).

Classification methods with supervised training techniques instruct the model with the use of a training set.

If classification techniques relates to texts, the reference variables are provided by words. In large datasets the number of variables to be processed notably increase and the application of a feature selection method (Sheydaei et al., 2015) which drops non-significant variables is required.

The training phase builds the training set in order to achieve as accurately as possible the message classification. The size of the training set varies depending on the number of categories and occurrences to be classified.

In a recent study, Imran et al. (2015) show that a training set applied simultaneously to different types of events or with different timing produces less accurate results.

3.4.5 Classification Techniques

The present section contains a description of main text classification techniques used in the empirical part of this Thesis.

Adaptive Boosting or **AdaBoost** (Freund et al. 1999). AdaBoost is a machine learning algorithm (adaptive classifier) which, in combination with other learning algorithms usually referred to as “weak learners” can be used to improve their robustness. A weighted sum is employed to increase the accuracy of the boosted classifier. Compared to other learning algorithms, AdaBoost is less subject to the overfitting problem. The individual learners can be weak in the sense that they cannot achieve accurate classification on their own, but as long as their performance is better than random guessing (their accuracy is larger than 0.5 for binary classification), the final model can be proven to converge to a strong learner.
Often, AdaBoost (with decision trees as weak learners) is referred to as the best out-of-the-box classifier, while the majority of learning algorithms typically depends on many different parameter configurations. If used with decision tree learning, information gathered at each stage of the AdaBoost algorithm about the relative ‘hardness’ of each training sample is fed into the tree growing algorithm such that later trees tend to focus on harder-to-classify examples. The main limitation of AdaBoost is that it is sensitive to noisy data and outliers.

Random forest (Breiman, 2001) is a machine learning algorithm which uses a combination of tree classifiers. Each tree depends on the values of a randomly and independently sampled vector,for all trees in the forest (see **Fig. 3.8** for a typical tree representation). The error of a random forest classifier depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node, the error rates compare favorably to other machine learning techniques (**AdaBoost**), and are more robust with respect to noise. Internal estimates check for error, strength, and correlation and are used as response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable significance.

The training set is seen as a collection of *N* examples of *f* features. In subsequent steps, Random Forest experiments use a combination of bagging and random features selection. The process is iterative such that a number of new training sets are drawn from the original one, with replacement. Then a tree is grown on the new training set using random feature selection. The trees grown are not pruned. The case study considered a trained forest with 500 trees, where at each split $\sqrt{f}$ features were randomly selected and thus the forest grown.

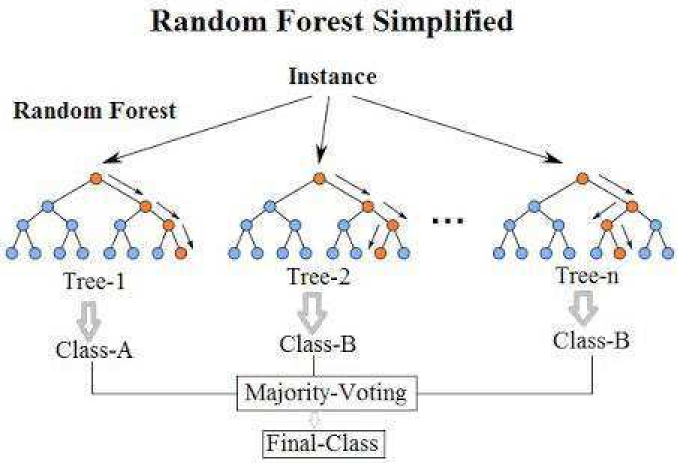


Fig. 3.8 The Random Forest algorithm

Support Vector Machine (SVM) (Suykens et al., 1999) is the state-of-the-art supervised machine learning method. For a given set of training examples a SVM model identifies the so-called support vectors, an hyperplane which best separates the classes, usually two, forming the training set. The dimension of the hyperplane can be generally different from the native space's one. **Fig. 3.10** defines the new examples that can be simply classified observing in which part of the hyperplane they are confined.

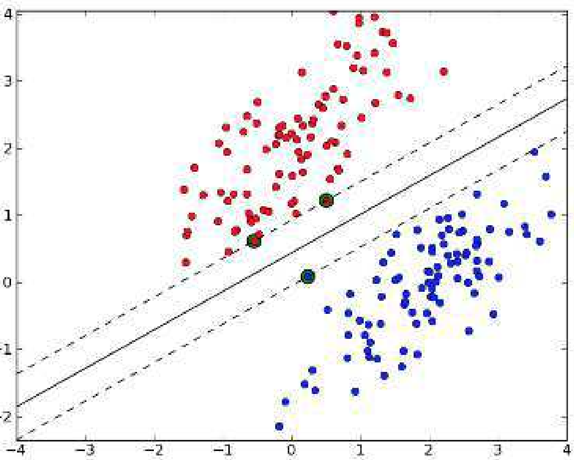


Fig. 3.9 The support vectors (red and blu dots) define the best separating hyperplane for classification

SVM is an effective model when dealing with dimensional spaces. It can also be effective in cases where the number of dimensions is greater than the number of samples (curse of dimensionality) as it uses a subset of training points in the decision function (called support vectors). Since it requires low memory space, it is a very efficient algorithm. The main disadvantage of SVN concerns the probability of its predictions. SVM do not directly provide probability estimates. These can be computed using a five-fold cross-validation. Moreover, if the number of features is much greater than the number of samples, the method is likely to give poor performances.

Neural Networks (Haykin, 2004) are a computational approach commonly adopted to fit a series of input variables with a defined outcome. When dealing with binary outcome, neural networks can be used to solve classification tasks. The rationale of neural networks mimics the way a biological brain attempts to solve problems with clusters of neurons connected by axons. Neural networks typically consist of multiple layers through which the signal is transmitted. According to a function cost, the internal weights of the network are adjusted in order to get maximum performance. Back propagation is one of the most adopted training configuration. In this case the signal moves back and training is iterated until the best configuration is reached. Like other machine learning methods neural networks have been used to solve a wide variety of tasks, like computer vision and speech recognition, that are hard to solve using ordinary rule-based programming.

Naïve Bayes (Rish, 2001) is a classification algorithm that uses density estimation to the data. The algorithm leverages Bayes theorem, and (naively) assumes that the predictors are conditionally independent, given the class. Naïve Bayes has been widely used for text classification and remains a benchmark methodology in the area of text analysis.
Naïve Bayes classifiers are highly scalable. They require a number of parameters which are linear in the number of variables (features/predictors) to solve a learning problem. Despite their naive design and simplified assumptions, comprehensive comparisons with other classification algorithms showed that Bayes classification can be outperformed by other approaches, such as boosted trees or random forests. An advantage of Naïve Bayes is that it only requires a small number of training data to estimate the parameters for classification analysis.

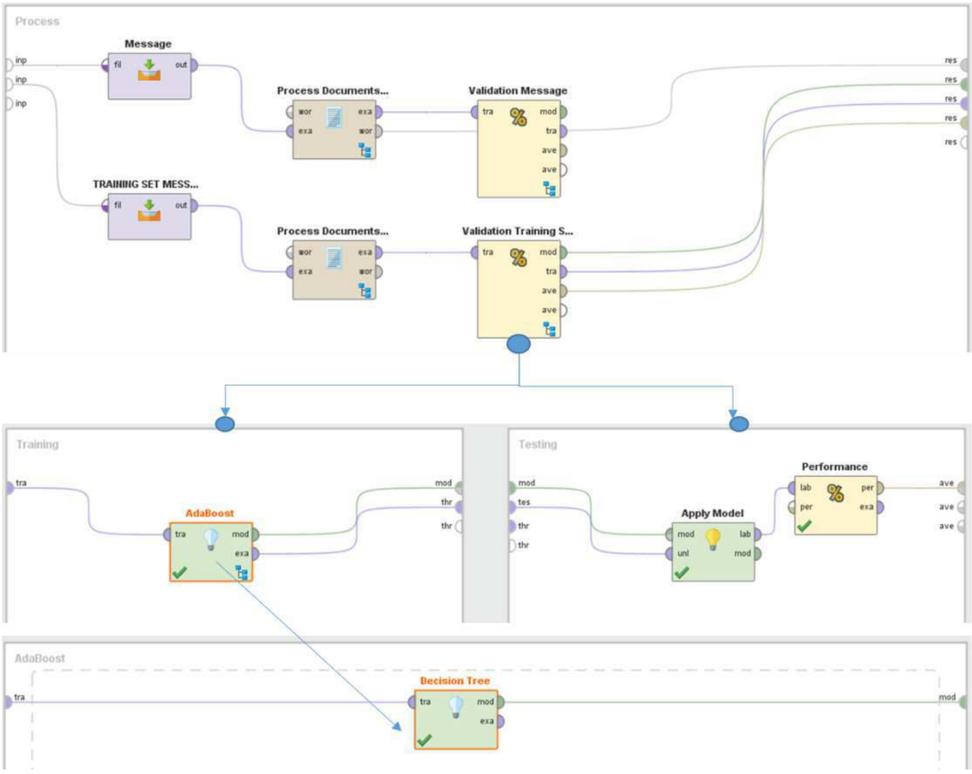


Fig. 3.10 Rapid Miner AdaBoost Algorihtm pipeline

3.4.6 Semantic Recognition

Disambiguation, in semantic analysis, is the automatic procedure attaching a suitable meaning to the linguistic expression (token).

Modern KTD platforms use a machine learning approach often running stochastic type models. This paradigm is different from classical rule-based approaches implementing a relevant number of rules for text analysis.

Automatic learning techniques define a set of rules through the analysis of a large number of texts (corpora) from a given domain. In the context of the present Thesis a corpus is a set of messages that should be learned. This set identifies different documents each with a probability value. Nonetheless, given the complexity of the process and the domain dimension, automatic learning techniques can be replaced by rule-based approaches.

In particular cases, the identification of certain forms of expression or the assignment of a meaning to previously determined syntactic structures, can easily be determined through rule-based approaches. These approaches determine a morfological-lexical-grammatical interpretation rule using appropriate glossaries and dictionaries.

Semantic recognition applies as follows:

- Named Entity Recognition: This is a recognition of entities' proper names (e.g. people, organizations, places), a recognition of temporal expressions (e.g. dates), addresses, and other particular information. The first step is defined as Named Entity Detection (NED) which identifies entities without knowing specific values (instances). For example, the instance 'Aiuto mi trovo davanti al Petruzzelli' (Help I am in front of Petruzzelli) indicates that the information 'Petruzzelli' is a person. The user has about knowledge on who this person is or what this person represents within the context of the instance. This problem is generally sorted out by Named Entity Recognition and Disambiguation (NERD) a technique which implements disambiguation methods.
- Coreference Resolution: This is a recognition of coreference. It is the set of referrals to the same referent. For example, 'I vigili del fuoco hanno comunicato che loro sono già sul posto' (Firemen said they already reached the place). Similarly, the case of anaphora where the coreference is within portions of text such as the following example 'ci

sono persone bloccate sotto il ponte ma nessuno li riesce ad aiutare' (There are people under the bridge but nobody is able to help).

- Relationship Extraction: This is the recognition of associative links between entities. For example, in the instance 'Ci sono persone infreddolite che hanno bisogno di coperte' (There are shivering people who need blankets) there exists a link between 'people' and objects ('blankets').

A semantic recognition often requires the recognition of all of the above aspects. The following example (see Fig. 3.11) 'Nell'area antistante Petruzzelli, volontari di Emergency distribuiscono cibo e vestiario, e coordinano tutti i volontari che devono interagire con loro' (In the area in front of Petruzzelli, Emergency volunteers distribuite food and clothing and coordinate all other volunteers which have to interact with them). A semantic recognition should recognise that there exists a place called Petruzzelli, two subjects named 'Emergency' and 'Volunteers', a link of 'distribution' between 'Emergency' and objects such as 'food and clothing', a link of 'coordination' between 'Emergency' and the 'volunteers', a link of 'interacting' between 'volunteers' and 'Emergency' (the latter also includes a coreference by the pronoun 'them').

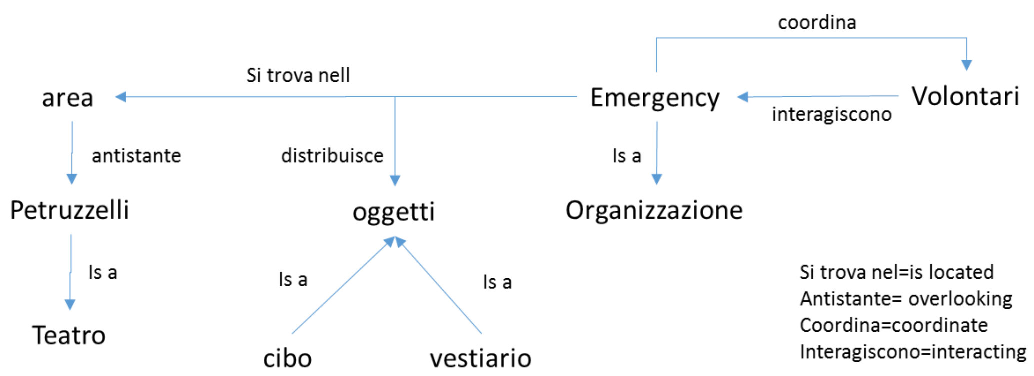


Fig. 3.11 Semantic recognition

Semantic recognition appears a complex task when instances include a relatively large number of subjects, expressions, and coreferences.

Nowadays, tools such as Stanford CoreNLP[31], UIMA (Unstructured Information Management Architecture) and GATE (General Architecture for Text Engineering) provide the management of semantic recognition.

Architectures developed by Stanford University[32] which implements a series of NLP tools[33] are particular known at the international level.

Stanford CoreNLP is a Java framework for the management of pipelines for document analysis which has found a notable expansion in the industrial and scientific community[34].

Stanford CoreNLP supplies Java API and implements a multi-thread management on a single machine minimizing learning and personalised computations on the basis of specific application needs. Stanford CoreNLP comprises of several models and annotated resources for English and other languages (Arabic, Chinese, French, German). As for Italian language there are several modules not yet supported by the Stanford CoreNLP architecture.

---

[31] Stanford CoreNLP is licensed under the GNU General Publi released by GNU General Public License

[32] http://nlp.stanford.edu/software/corenlp.shtml

[33] http://nlp.stanford.edu/software/

[34] http://nlp.stanford.edu/pubs/

3.4.7 Class Identify

Fig. 3.12 shows a flowchart of useful information to knowledge construction in disaster event.

After verifying the domain of the text message, this should be identified. The identification can refer to a request for or offer to help or a communication indicating a danger, a food and clothing collection point. Alternatively, it can be a text message which supplies a telephone number or a bank account for donations. These types of texts provide information on the role played by the user (e.g. affected person, person offering help). The content of the text message could also refer to other people which are different from the user or to spatial / geographical location. The latter is often linked to other spatio-temporal information of the surrounding environment. In Italian language this link is provided by specific words which can appear before or after the noun, can be adjectives, improper prepositions, adverb prepositions (i.e. ‘I am next to the theatre’ - ‘mi trovo vicino al teatro’, ‘the theatre is next to my house’- ‘il teatro è vicino casa mia’). Finally, class identify recognises the needs such as ‘needs for survival’ (e.g. food, water, shelter) and ‘secondary needs’ (e.g. mobile phones, toys).

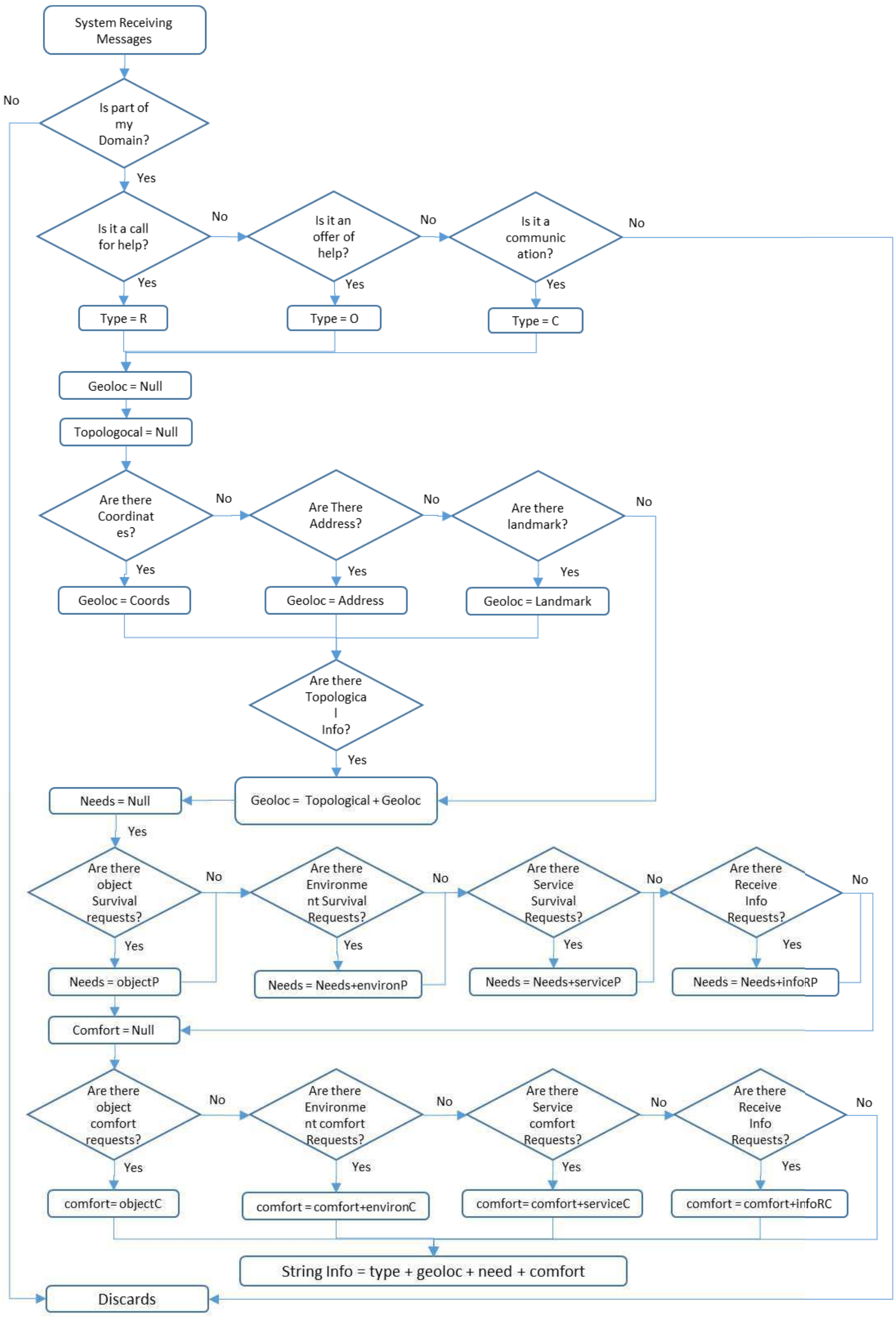


Fig. 3.12 Flow Chart showing Primary Information Extraction

3.4.8 Local Knowledge and Spatial Relationship in the Municipality of Bari

This section briefly describes a qualitative analysis to support some of the results obtained by the text mining approach. The text mining analysis revealed the presence of statements whose meaning could create ambiguities to determine the spatial location of users asking for/offering to help in risk situations. The ambiguities, arising from the use of 'improper' prepositions (as described in **section 3.4.3**) adopted in the natural language, provide biased information and affect local knowledge in disaster response. To cover this gap a mini survey is carried out with the following aims: i. To test local knowledge on given landmarks and help determine the spatial location of affected population under risk situations; ii. To provide further information from local knowledge and improve the services offered by the spatial information system; iii. To determine the existence of local/tacit knowledge in the natural language; iv. To demonstrate the existence of subjective knowledge in the perception of spaces; v. To support the usefulness of natural language to the understanding of local knowledge as effective tool for rescuing operations.

The mini-survey is performed on a random sample of 143 people living in Bari. The survey is structured to retrieve information on the use of natural language in emergency events and to respond to the above aims. The first five questions inquire on terms such as 'the entrance of', 'end of street', 'in front of', 'behind' which are ambiguous to determine a specific spatial location. These terms refer to landmarks in Bari such as the 'Fiera del Levante' (Levante's Fair), 'La Muraglia' (The Wall), 'Basilica di San Nicola' (Basilica of Saint Nicholas), Teatro Petruzzelli' (Petruzzelli Theatre), and 'Parco Largo 2 Giugno' (2 June Square). **Figures Fig.** 3.13**Fig. 3.17** show the questions related to the above landmarks.

Each question also comprises of a 3D-picture of the landmark taken by Google Map with the indication of the possible answers.

Fig. 3.13 shows the following question: 'A person that you know asks for your help and he/she says that he/she is at the ENTRANCE OF the Levante's Fair. Which of the marked entrances would you reach to help your friend?'. The question clearly indicates the spatial location of the entrances (B,G,R,M) on a geo-referenced map as well as shows a picture of each of the entrances.

Fig. 3.13 Question 1 of the mini-survey

Fig. 3.14 asks respondents an enquiry on the following aspect: 'A person that you know asks for your help and he/she says that he/she is at the END OF THE MURAGLIA. Which of the marked entrances would you reach to help your friend?' The question indicates the location of the entrances (X,J or do not know) and shows the pictures of each of the entrances.

Fig. 3.14 Question 2 of the mini-survey

Fig. 3.15 asks respondents an enquiry about the Cathedral of Saint Nicholas: 'A person that you know asks for your help and he/she says that he/she is IN FRONT OF the Cathedral of Saint Nicholas. Which of the three marked entrances would you reach to help your friend?'

The question indicates the location of the entrances (B,G,R) and shows the pictures of each of the entrances.

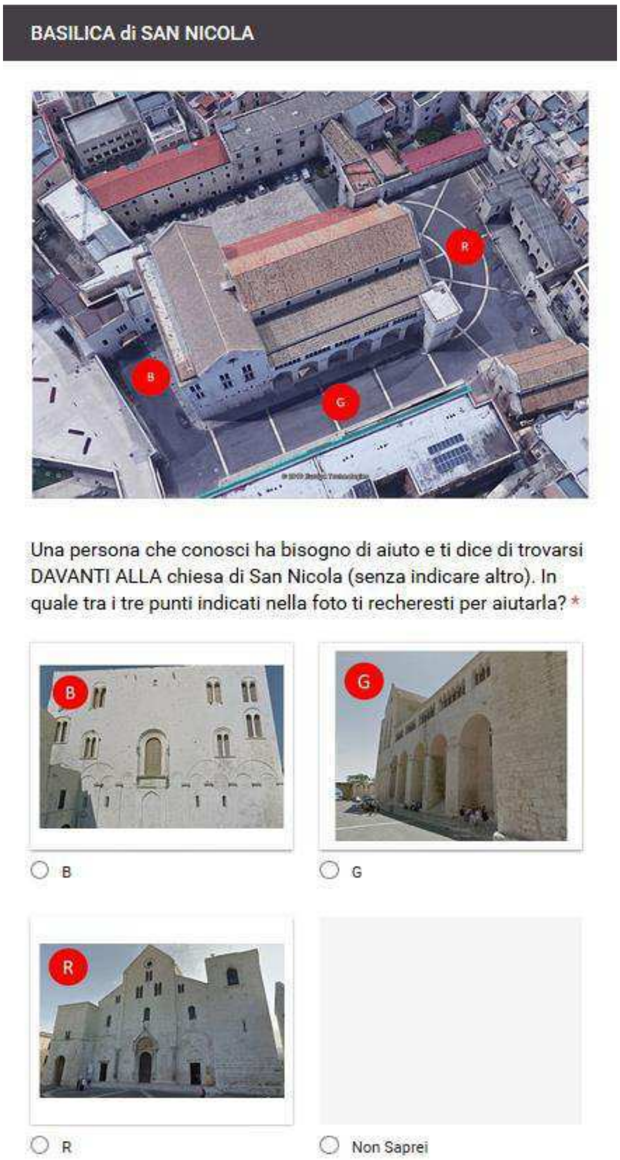


Fig. 3.15 Question 3 of the mini-survey

Fig. 3.16 asks respondents an enquiry about the Petruzzelli Theatre: 'A person that you know asks for your help and he/she says that he/she is BEHIND the Petruzzelli Theatre.

Which of the four marked entrances would you reach to help your friend?' The question indicates the location of the entrances (B,G,R, M) and shows the pictures of each entrance.

Fig. 3.16 Question 4 of the mini-survey

Fig. 3.17 asks respondents an enquiry about the Parco 2 Giugno: 'A person that you know asks for your help and he/she says that he/she is in the building IN FRONT OF the entrance of Parco 2 Giugno. Which of the three marked entrances would you reach to help your friend?' The question indicates the location of the entrances (B,G,R, do not know) and shows the pictures of each entrance.

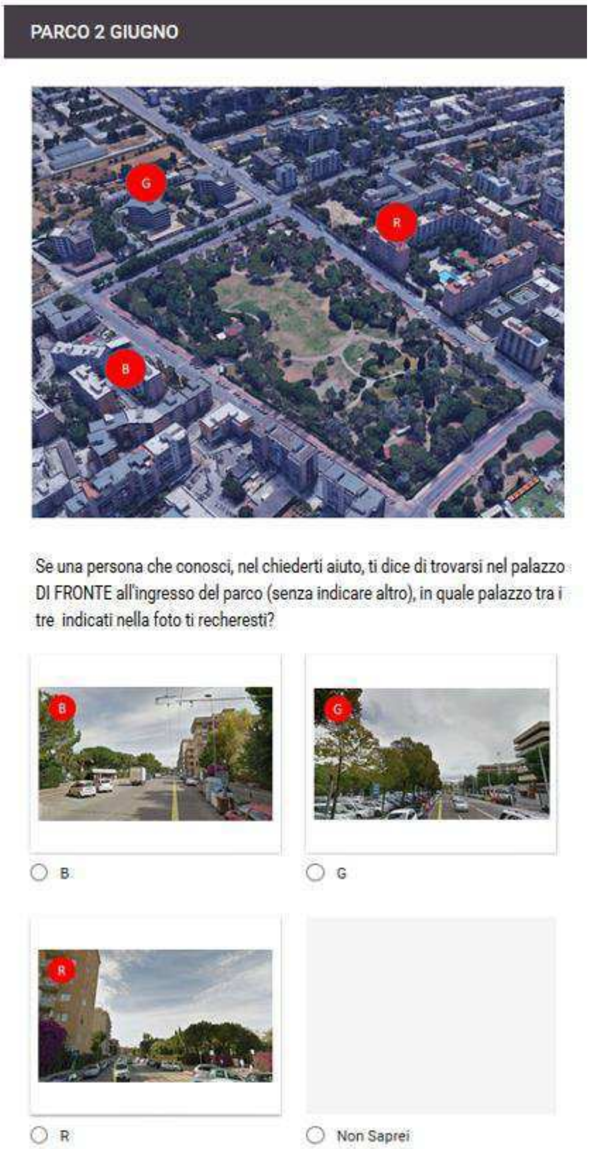


Fig. 3.17 Question 5 of the mini-survey

The last three questions of the survey retrieve useful information on the use of the words 'near to/far from' (Figures Fig. 3.18 - Fig. 3.19) and 'nearby' (Fig. 3.20). Figures Fig. 3.18 and Fig. 3.19 refer to 'Corso Vittorio Emanuele' (Vittorio Emanuele Street) one of the main streets in Bari Fig. 3.20 refers to the Central Station.

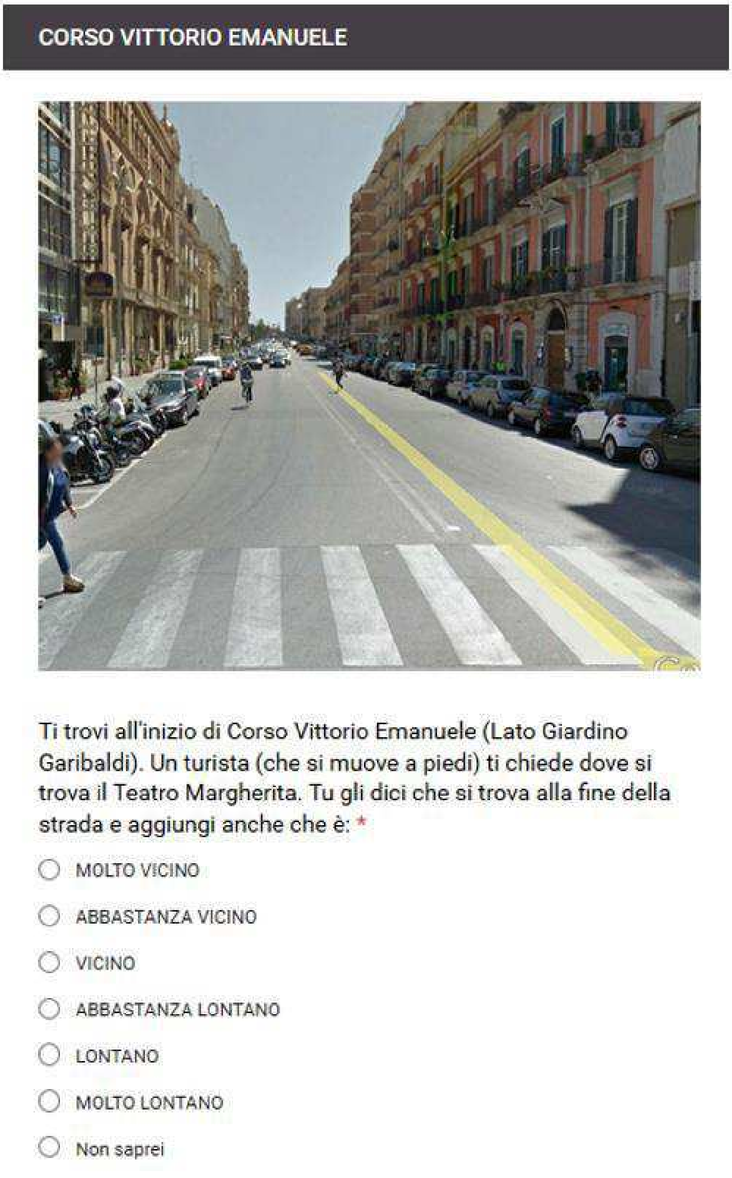
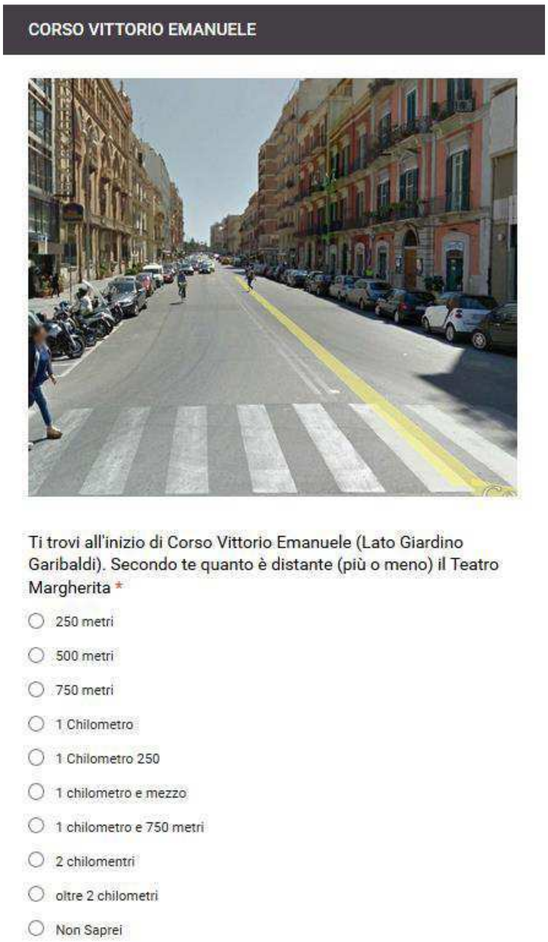


Fig. 3.18 Question 6 of the mini-survey

In particular, Fig. 3.18 shows a question asking respondents the following statement: ‘You are at the beginning of Vittorio Emanuele Street (next to Garibaldi Garden). A tourist asks you for the location of the Margherita Theatre. You tell him/her that he/she is at the end of the street and that the Theatre is: very near, quite near, near, quite far, far, very far, do not know’. The question that follows (Fig. 3.19) tests respondents on the perception of the distance between his/her/the tourist’s location and that of the Margherita Theatre. Respondents provide to answer on a scale 1-9(250m-more than 2km) and the ‘don’t know’ answer.


CORSO VITTORIO EMANUELE

Ti trovi all'inizio di Corso Vittorio Emanuele (Lato Giardino Garibaldi). Secondo te quanto è distante (più o meno) il Teatro Margherita *

- 250 metri
- 500 metri
- 750 metri
- 1 Chilometro
- 1 Chilometro 250
- 1 chilometro e mezzo
- 1 chilometro e 750 metri
- 2 chilomentri
- oltre 2 chilometri
- Non Saprei

Fig. 3.19 Question 7 of the mini-survey

Finally, Fig. 3.20 shows the last question which captures an understanding of the term 'nearby' by respondents. The figure shows a location of the Bari Central Station and ten ring zones. Each zone is 100 metres distant from the station. In particular, the question asks the following statement: 'Bari Central Station is located in A. Which area (identified by a letter) do you think can be defined as being NEARBY the station?'. The map also indicates some outlier landmarks such as Chiesa Redentore (Redentore Church), Giardino Garibaldi (Garibaldi Garden) and Campus Universitario (University Campus).

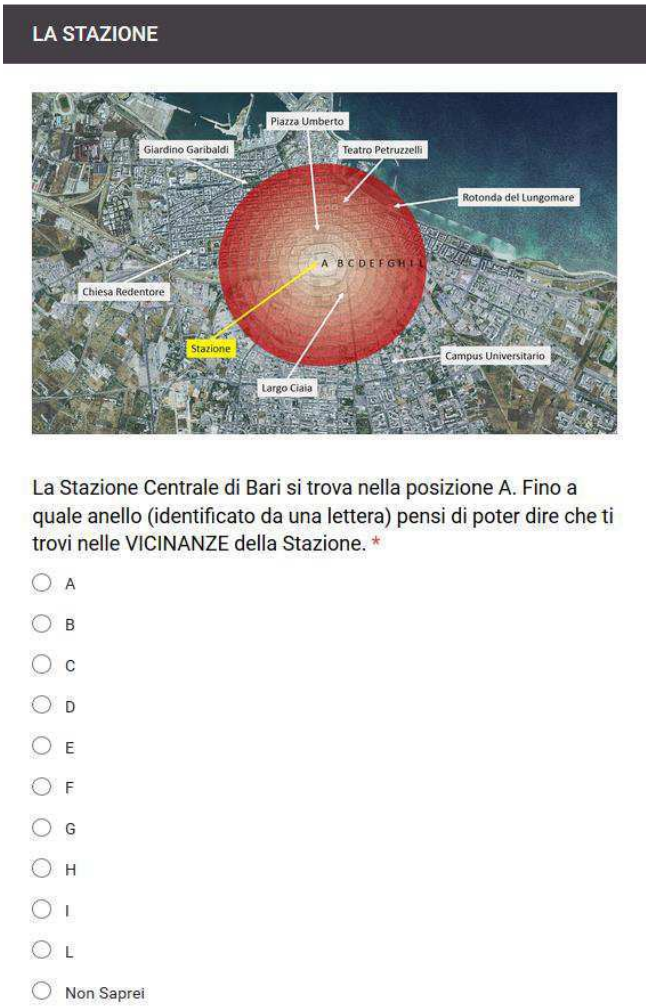


Fig. 3.20 Question 8 of the mini-survey

3.5 Understanding Knowledge Using Ontological Analysis

3.5.1 General Concept for Model Construction

From a methodological point of view the conceptual model takes into account both general and specific aspects of risk/emergency in disaster response domain. The model identifies classes and relationships organised according to ordinal hierarchies.

The ‘concept’ class describes the concepts/definitions of a domain. This can have subclasses containing specific aspects. Two main approaches for the development of a hierarchy of classes are defined by Uschold and Gruninger (1996):

- A top-down model identifies the general concepts (general classes) and specifies particular details (specific classes) about the concepts;
- A bottom-up model identifies specific concepts (specific classes) which represent the leaves of the hierarchical tree and then aggregates these concepts into more general ones (general classes).

The present Thesis adopts a bottom-up approach for the construction of the conceptual model. First, it analyses specific terms/documents taken from the data. Second, a membership entity class is established; finally, general relations are established between these entities on the basis of particular properties to describe the entire domain.

In contrast to the high degree of standardization of conceptual modeling, there are no widely accepted ontological standards to use as a basis for a common analysis which also captures cross-domain concepts.

In addition, there is a general trend to consider local knowledge as it appears in natural language and to represent them with a basic analytical effort. Since the semantics of natural language is not entirely defined and varies across cultures, it leads to certain degrees of heterogeneity in specific ontologies. This is particularly relevant when these ontologies are developed by independent organizations and are culturally diverse.

Communication tools represent a value added in the management of emergencies, but at the same time they create complexity due to the amount of data to manage.

Section 3.4 highlighted the steps needed to define the domain and isolate single words with the use of text mining approach.

The following sections will focus on the description of the tools necessary to create taxonomies and ontologies for the context of the present work. The creation of taxonomies and ontologies in a disaster response domain arises from the need to build conceptual models including all elements that somehow come into play. This is particularly relevant for the definition and interpretation of ‘needs’ to save lives.

The use of taxonomies and ontologies also arises with the aim of creating a strong reference for further developments such as application designs providing adequate indications to satisfy the needs which are universally shared and accepted about people and places.

3.5.2 Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)

This section illustrates the DOLCE ontology as this is used in the empirical work carried out in the context of the present dissertation. The choice of DOLCE that defines the emergency/risk domain may be traced to its specific descriptive properties, and to its features allowing to incorporate local knowledge.

The DOLCE ontology (Masolo et al., 2002) aims at both extending ontological libraries, which are built using a standard language for the web and a foundational ontology (Borgo et al., 2010); opening to the semantic integration and sharing of different ontologies through the development of specific methodologies (Gaio et al., 2010).

As for the use of foundational ontologies, the main idea was to develop several ontologies (assuming than one alone would be too difficult to manage) to face the diffusion of the Semantic Web – a research-oriented technology also based on the comparison between several foundational ontologies pertaining to diversified contents (Gaio et al. 2010).

Since 2002, DOLCE is part of the WonderWeb library and one of the most relevant references of the ontological literature. The contents of the ontology collect conceptualizations based on the structure of the natural language, human cognition and local knowledge (Borgo et al., 2010).

DOLCE is a descriptive ontology. It is centred on the natural language 'as is', in its daily use and according to common sense, assuming that both are ontologically identifiable. It is different from other ontologies that rely mainly on the use of expert and scientific knowledge (Gaius et al., 2010).

Tab. 3.5 shows DOLCE's categories as 'cognitive boxes' (Oltremari et al., 2003) affected by human perception, cultural traces and social codes (Gaio et al. 2010; Masolo et al., 2003).

Tab. 3.5 'Leaf' basic category

"Leaf" Basic Category	Examples
Abstract Quality	*the value of an asset*
Abstract Region	*the (conventional) value of 1 Euro*
Accomplishment	*a conference, an ascent, a performance*
Achievement	*reaching the summit of K2, a departure, a death*
Agentive Physical Object	*a human person (as opposed to legal person)*
Amount of Matter	*some air, some gold, some cement*
Arbitrary Sum	*my left foot and my car*
Feature	*a hole, a gulf, an opening, a boundary*
Mental Object	*a percept, a sense datum*
Non-agentive Physical Object	*a hammer, a house, a computer, a human body*
Non-agentive Social Object	*a law, an economic system, a currency, an asset*
Physical Quality	*the weight of a pen, the color of an apple*
Physical Region	*the physical space, an area in the color spectrum, 80Kg*
Process	*running, writing*
Social Agent	*a (legal) person, a contractant*
Society	*Fiat, Apple, the Bank of Italy*
State	*being sitting, being open, being happy, being red*
Temporal Quality	*the duration of World War I, the starting time of the 2000 Olympics*
Temporal Region	*the time axis, 22 june 2002, one second*

Source: Masolo et al. (2003).

DOLCE is also a *Multiplicative* ontology, since it has a general view of the real world and admits distinct entities that can share the same spatio-temporal area (i.e. entities can be co-localised) (Masolo et al., 2003). The use of relational instances in DOLCE is the link between universal and particular ontologies. The former assumes a primary (universal)/organizational role, and the latter a secondary (particular)/descriptive role.

DOLCE consists of four main types of entities: endurants (continuants), perdurant (occurrents), quality and abstract.

The first two, as already seen in **Section 2.3.2**, are linked to each other through a relation of participation – e.g. a person (endurant) participates in an event (perdurant). Quality and property-type ontologies do not describe the same concept. The former indicates a particular entity, the latter a universal one (Masolo et al., 2003).

Quality-type ontology illustrates a perception (such as colours, forms, sounds, odours) or a measure (such as weight, height, width). The life cycle of a quality ontology is strictly linked to the life cycle of the entity to which it belongs to: once the entity terminates, the quality terminates too (Masolo et al., 2003).

A further feature of the quality ontology is the existence of a specific position within a 'quality space' named 'qualia' such as, for example, between a quality colour of a rose, a quality cotton of a fabric (Gaio et al, 2010; Masolo et al., 2003; Oltremari et al. 2003).

Fig. 3.21 shows DOLCE's taxonomy of entities. The taxonomy adopts a logic of observation which differentiates objects, as distinct entities, in their diverse functions. The concept of 'plaza', for example, is considered as a physical object (PhysicalObject), a place (SpatialLocation) and a space region (SpaceRegion).

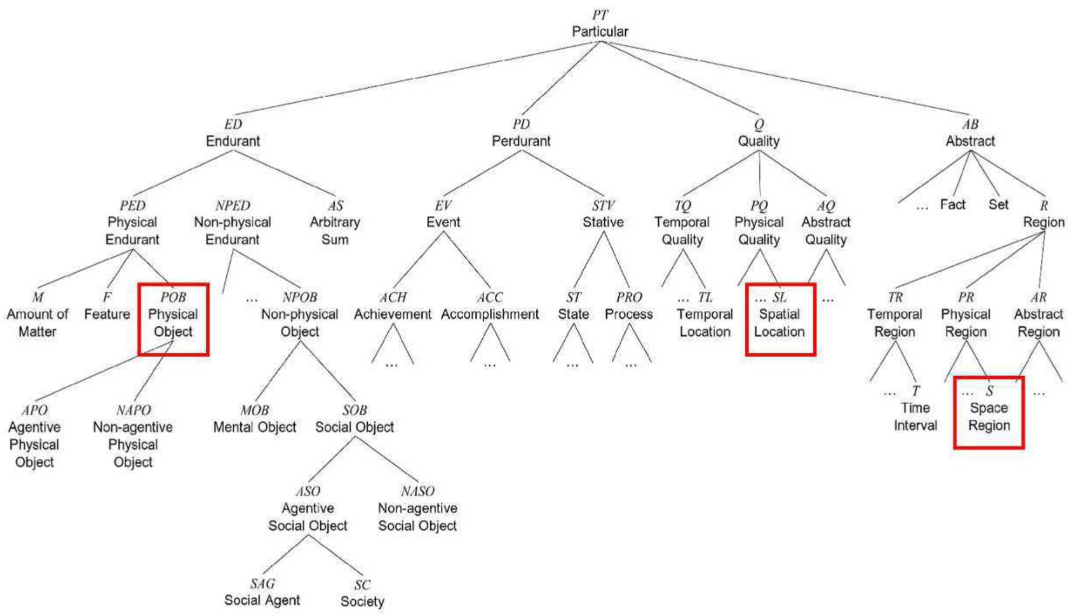


Fig. 3.21 DOLCE's taxonomy of entities – adapted from Masolo et al. (2003).

Finally, DOLCE is based on the OntoClean methodology developed by LOA-CNR[35] as validation tool aimed at assessing the robustness and adequacy of the taxonomy's relations (Guarino and Welty, 2009).

35 See http://www.loa.istc.cnr.it. Accessed: 15/12/2016.

3.5.3 Ontology and Event Response Knowledge

In the present Thesis, the creation of an ontology for a post-disaster event, aims to help solving complex problems that entail lack of understanding between actors and matching help requests with offers of assistance, through exchange of messages by the affected population. These forms of requests generally communicate a need for survival by including other information (not just basic information) – such as reference to places, other actors, objects, actions.

The above gaps may widen, as a result of demographic dynamics involving populations from heterogeneous ethnic groups with different languages, cultures and knowledge backgrounds. This diversity implies changes of social, cultural and language contexts and, while the transition is ongoing, it is complex to manage in terms of knowledge understanding. For this reason, the use of ontologies is predicated on the need to improve semantic interoperability of natural disaster domain models. Also, ontologies limit potential misunderstanding among rescue operators and strengthen the effectiveness and reliability of decision support systems.

The central role played by ICT in supporting disaster response requires cooperation between governmental bodies and other agencies, while operating on different software platforms to share the same type of information.

Regardless of the technologies used for the exchange and sharing of information, it is relevant to have a common understanding of concepts and meanings of shared entities, in order to facilitate the functioning of decision support systems and make them interoperable (Hiltz et al., 2011).

More precisely, on the semantic level, interoperability calls for specific requirements, including common vocabularies and links between machine-understandable concepts to facilitate the interpretation and sharing of knowledge (Imran et al., 2015).

The use of ontologies to improve emergency domain definitions is well established in the scientific literature. Wang et al. (2006; 2009) define an ontological model of events, processes and actions based on sharing a vocabulary to exchange information. Xu et al. (2014)

use specific geo-ontology libraries to describe an earthquake event. A geo-ontology is oriented to a geo-spatial hierarchy of information, and it offers a semantic interpretation of concepts (Li, 2007).

Murgante et al. (2009) address seismic risk in urban areas through the use of an ontology. The model is developed to share knowledge so that concepts are fully understandable and accessible to the intended stakeholders.

Lee et al. (2013) apply an ontological model to develop a smart-type approach through the use of a context-aware platform and address real-time emergency operations/situations.

Apisakmontri (2013) uses an ontological approach for Refugee Emergencies in Disaster Management. The study is similar to that adopted in the present Thesis, as it involves the construction of an ontology to define needs or the integration of a lightweight ontologies with four foundational ones (namely, DOLCE, SUMO, FOAF, and SWEET) (Fig. 3.22).

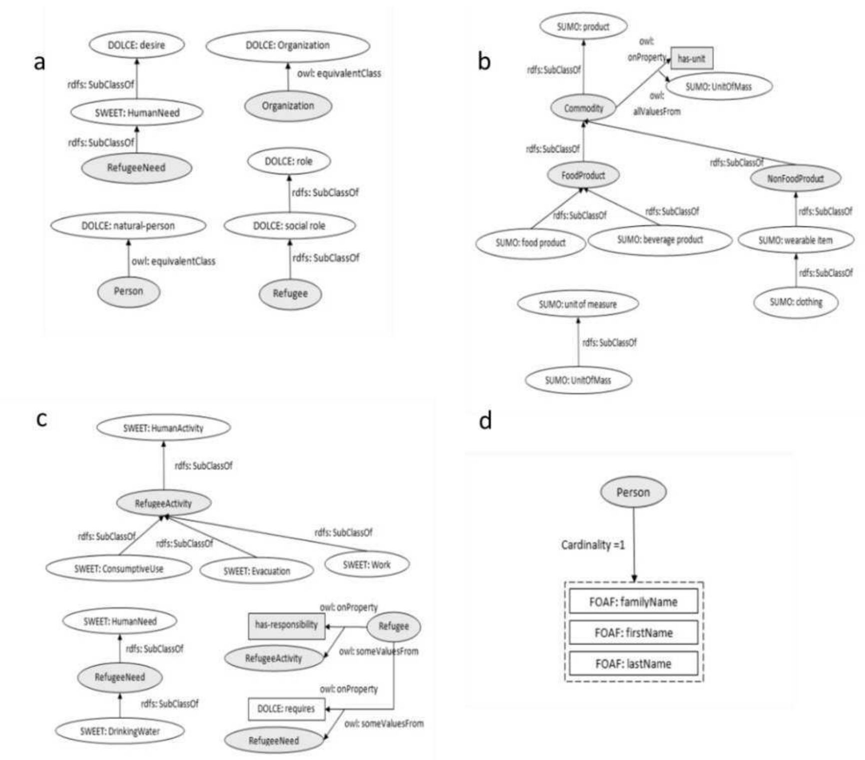


Fig. 3.22 (a) Integration with DOLCE, (b) integration with SUMO, (c) integration with SWEET, integration with (FOAF). Adapted from Apisakmontri (2013).

The W3C Incubator Group Report (Ianella, 2009) focuses on the main aspects of communication between rescue operators in a post-disaster situation, and uses foundational and non-foundational ontologies.

The ontological model (Fig. 3.23) is build by means of three sections: i) 'What' deals with needs issues; ii) 'Where' refers to spatial aspects; and iii) 'Who' deals with the actors involved in the post-disaster event. The model W3C does not focus on the time frame ('When') although it is an important aspect (e.g., 'I need food by tomorrow').

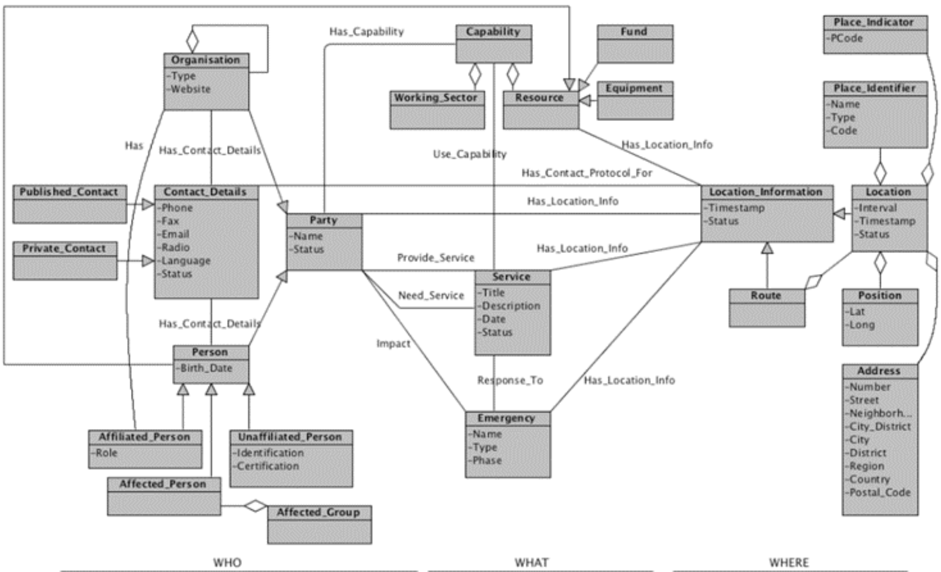


Fig. 3.23 Who, What, Where information models. Adapted after the W3C Incubator Group Report (Ianella, 2009).

The Humanitarian eXchange Language (HXL)[36] is a project run by the United Nations Office for the Coordination of Humanitarian Affairs that is geared towards improving the management and exchange of information during the disaster-response phases. HXL is designed as an action response to disasters like the Haiti earthquake of 2010, and is organized

[36] http://hxl.humanitarianresponse.info. Accessed: 15/12/2016.

on four main domains: disaster type, organization (actors, agencies, etc.), location (event location, people location), and damage (to humans and to infrastructures) (Keßler and Hendrix, 2015) (**Fig. 3.24**).

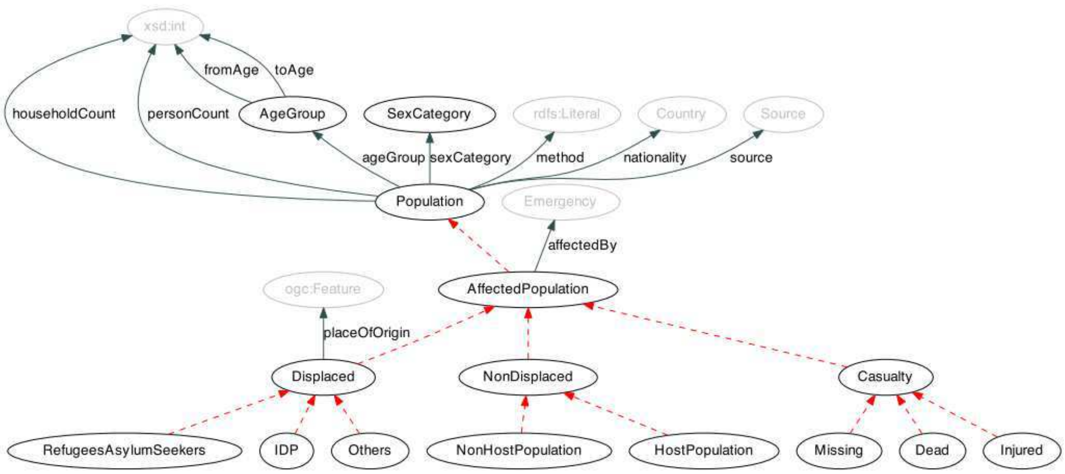


Fig. 3.24 Overview of the core classes and properties of the humanitarian profile section (Keßler and Hendrix, 2015).

Management of A Crisis Vocabulary (MOAC)[37] is a lightweight vocabulary that is designed to provide terminological references to agencies and emergency operators. Like HXL, MOAC aims to describe different aspects of a crisis, including its effects, the needs of affected people, and the response to the crisis. MOAC, in particular, deals with four areas: disaster, damage, processes (e.g. rescue, search, evacuation processes) and resources (such as services, vehicles, tents) (Imran et al., 2015).

[37] http://observedchange.com/moac/ns. Accessed: 14/12/2016.

The Service-oriented architectures supporting networks of public security (SoKNOS)[38] is an ontology to develop and evaluate concepts with a view to support governmental agencies, organizations and companies in their capacity as providers of public security in emergency situations. SoKNOS follows the principles of the DOLCE foundational ontology (Babitski et al., 2009). For example, a class 'Device' is defined in SoKNOS which specialises Dolce's taxonomy of 'Non – Agentive Physical Object'.

In recent years, the development of methodologies to implement these ontologies has generated a debate around the heterogeneity issue. In this regard, it should be remembered that one of the main objectives of an ontology is to facilitate knowledge sharing. The scientific community should aim to create a shared integration mechanism whereby ontologies that describe the same domain or have overlapping areas, adopt unambiguously the same concept.

Noy (2004) identifies two methods for tackling this issue. The first one, which has met wide consensus in the literature, turns to foundational or upper-level ontologies to identify the classes that serve as a link between specific ontologies. The second approach includes heuristics-based techniques or machine learning that take advantage of the different features of ontologies (structure, definitions of concepts, instances of classes) to work towards a shared mapping.

38 The English name of the ontology is a translation from the German 'Service-Orientierte Architekturen zur Unterstützung von Netzwerken im Rahmen Öffentlicher Sicherheit', see www.soknos.de. Accessed: 14/12/2016.

4 Results

4.1 Classification Results

4.1.1 Validation

To validate the robustness of the obtained results, a cross-validation approach is used. Generally, cross-validation procedures distinguish n-fold and leave-one-out cross-validation (Basari et al., 2013). The former is carried out with a nested approach and is the algorithm included in Rapidminer. Data is split into n-folds of equal size and trained and tested n-times. Of these n-subsets, a single subset is hold as input of the testing sub-procedure, and the rest of the n-1 subsets are then applied as training data in the subsequent reiteration (i.e. as input of the training sub-procedure). The cross-validation is repeated n-times treating the n-subsets as holdout sets each time. The cross-validation procedure then predicts how sensitive is the model (i.e. how well performs the model) to a hypothetical holdout dataset.

4.1.2 Models Comparisons

A set of classification experiments was performed using a 5-fold cross-validation framework. 50 cross-validation rounds were performed using several classification models (Adaboost, Random Forest, Support Vector Machine, Neural Networks and Naïve Bayes) on a training set consisting of $\sim$1000 observations. The trained model was then applied on an independent test to evaluate the robustness and the accuracy of the learned models (Tab. 4.1).

The datasets used for processing were collected according to the description of the previous section. Data retrieved from Ushahidi about territorial issues and the on-line survey are stored in the same database containing 1300 messages in total. The creation of the full database is useful to apply and validate text classification techniques in a subsequent stage of the Thesis's work.

For the simulation analysis a training set of about 200 observations was created. A sub-sample of 100 messages was taken from the on-line survey and a sub-sample of 100 messages was taken from the generic dataset about ‘territorial issues’.
Classification accuracy on the training and test sets has been compared. Training accuracy is obtained by averaging the cross-validation performances. Each trained model was also used on the test set.

Tab. 4.1 Cross-validation of classification models

Method	Training Accuracy	Test Accuracy
AdaBoost	92.1 ± 1.4 %	87.4 ± 1.3 %
Random Forest	94.1 ± 1.5 %	96.5 ± 0.8 %
Support Vector Machine	92.8 ± 2.1 %	94.4 ± 3.5 %
Neural Networks	93.2 ± 1.6 %	95.2 ± 1.1 %
Naïve Bayes	81.6 ± 2.3 %	73.3 ±1.8 %

4.2 Local Knowledge and Spatial Relations in The Municipality Of Bari: Results

This section describes the results of the qualitative analysis carried out in Bari. The analysis is based on a mini-survey:

- To test local knowledge on given landmarks and help determine the spatial location of affected population under risk situations;
- To provide further information from local knowledge and improve the services offered by the spatial information system;
- To determine the existence of local/tacit knowledge in the natural language;
- To demonstrate the existence of subjective knowledge in the perception of spaces;
- To support the usefulness of natural language to the understanding of local knowledge as effective tool for rescuing operations

As already illustrated in **Section 3.4.8**, the mini-survey is based on a random sample of 150 people living in Bari. The final sample comprises of 123 observations after removing outliers.

Tables Tab. 4.2 **-**Tab. 4.7 show the obtained results relative to questions 1-8 of the mini-survey. The obtained results suggest that the use of natural language to understand local knowledge is an important tool. In turn local knowledge appears a relevant factor in disaster response. In detail, **Tab. 4.2** shows that 65% of respondents argue that 'the' entrance of the Levante's Fair is that located at point G (which is the main entrance). Similarly, in **Tab. 4.3** the 'end of' the Wall (La Muraglia) is that corresponding to point J for almost 84% of respondents. In addition, the almost totality of the sample agree that 'In front of' Saint Nicholas Cathedral means the main entrance as indicated in point R (**Tab. 4.4**). For the majority of respondents being 'behind' The Petruzzelli Theatre means to be at point M (
Tab. 4.5). As for **Tables** Tab. 4.7 - Tab. 4.8 the responses seem somehow consistent. 38% of respondents think that Margherita's Theatre and Garibaldi Garden are 'Far' away from

each other. This perceived ‘distance’ corresponds to 750m for the majority of the sample (32.5%). In **Tab. 4.9** the ‘nearby’ distance from the Central Station coincides with 400m for 39% of respondents. Finally, as for question in **Tab. 4.6** the distribution of respondents is mostly bimodal (see kernel density function in **Tab. 4.6 b**)) on the entrances G (about 30%) and R (about 30%).

Tab. 4.2 Frequency table Q1

	Levante’s Fair (Fiera del Levante) entrance		
	Freq	%	Cum
B	5	4.07	4.07
G	**83**	**64.48**	71.54
M	22	17.89	89.43
N	10	8.13	79.56
R	3	2.44	100

Tab. 4.3 Frequency table Q2

	‘End of’ the Wall (La Muraglia)		
	Freq	%	Cum
J	**103**	**83.74**	83.74
N	8	6.5	90.24
X	12	9.76	100

Tab. 4.4 Frequency table Q3

	‘In front of’ Saint Nicholas Cathedral (Basilica di San Nicola)		
	Freq	%	Cum
B	2	1.63	1.63
G	4	3.25	4.88
N	1	0.81	5.69
R	**116**	**94.31**	100

Tab. 4.5 Frequency table Q4

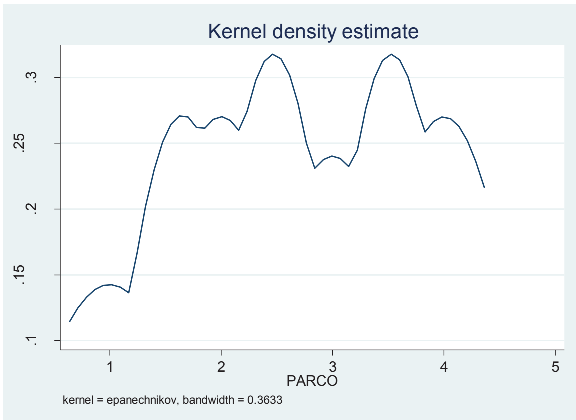
	'Behind' Petruzzelli Theatre (Teatro Petruzzelli)		
	Freq	%	Cum
B	9	7.32	7.32
M	**27**	**78.86**	86.18
N	2	1.63	87.80
R	15	12.20	100

Tab. 4.6 Frequency table Q5

a)

	'In front of' Largo 2 June Park (Parco Largo 2 Giugno)		
	Freq	%	Cum
B	19	15.45	15.45
G	**36**	**29.27**	44.72
N	32	26.02	70.73
R	**36**	**29.27**	100

b)



Tab. 4.7 Frequency table Q6

	How far is Margherita's Theatre from Garibaldi Garden		
	Freq	%	Cum
Quite far	18	14.63	14.63
Quite near	41	33.33	47.97
Far	7	5.69	53.66
Very far	2	1.63	55.28
Very near	5	4.07	59.35
Near	3	2.44	61.79
Far	**47**	**38.21**	100

Tab. 4.8 Frequency table Q6

	How far is Margherita's Theatre from Garibaldi Garden		
	Freq	%	Cum
Quite far	18	14.63	14.63
Quite near	41	33.33	47.97
Far	7	5.69	53.66
Very far	2	1.63	55.28
Very near	5	4.07	59.35
Near	3	2.44	61.79
Far	**47**	**38.21**	100

Tab. 4.9 Frequency table Q7

	The perceived distance between the Margherita's Theatre and Garibaldi Garden		
	Freq	%	Cum
250m	3	2.44	2.44
500m	19	15.45	17.89
750m	40	32.52	50.41
1000m	32	26.02	76.42
1250m	6	4.88	81.30
1500m	20	16.26	97.56
2000m	**3**	**2.44**	100

4.3 Ontological Analysis of What, Where, Who, When

The ontologies illustrated in **Section 3.5.3** are crisis-specific, but not social-media specific. Many concepts of these ontologies can be analysed with a different perspective, which focuses on the natural language used in event response, largely influenced by local knowledge.

Recently, specific ontologies have been developed to describe social media concepts as that of 'Semantically Interlinked Online Communities' (SIOC) ontology. This was originally developed to model websites such as blogs and online forums (Imran et al. 2015); while Meaning-Of-A-Tag (MOAT) implements an ontology with semantic tagging of social media data (Passant and Laublet, 2008).

The conceptual aspects of the ontology are inspired to the work by Mele and Sorgente (2011): The *Eventory project [WAM*]. This project takes its roots from the journalism field and adopts the model called 'W's and one H'. This models uses six fields to represent an event: 'Who', 'When', 'Where', 'What', 'Why', and 'How' (Atit and Sundaram, 2007). With respect to the design aspects, the ontology refers to the model proposed by the W3C Incubator Group Report 2009 (Ianella, 2009) (shown in **Section 3.5.3**), which is based on three fields: 'What', 'Where', and 'Who'.

The analysis of the simulation study illustrated in this section and by the recent literature (e.g. the Haiti earthquake[39], the Hurricane Sandy[40]) justify the ontological model structured into the following macro-fields: 'What', 'Where', 'Who' and 'When'.

Fig. 4.1 shows a first general taxonomy which considers the abovementioned four fields and the relationships occurring across the entities.

Before illustrating the entities attached to the fields of 'What', 'Where', 'Who' and 'When', it is relevant to briefly introduce the 'Why' field (which is not included in the above taxonomy). Under certain aspects, this field can be present in text messages and justifies the existence of the domain.

---

[39] Haiti crisis map https://datahub.io/dataset/ushahidi/resource/81d058a8-173a-49d9-8ce9-4edf5e7cafc9 https://github.com/unthinkingly/haiti.ushahidi.com-twitter-export

[40] Hurricane Sandy http://www.zubiaga.org/datasets/hurricane-sandy-tweets/

Generally, text messages belong to the disaster response domain. The contents related to the 'Why' field answer to the question of 'Why did the event happen?'. The present Thesis does not consider the reasons why an event happens or why a message is exchanged. The present Thesis deals with post disasters texts. It is worth mentioning about the possibility of exchanging messages before an event and this may contribute to mitigate disaster response.

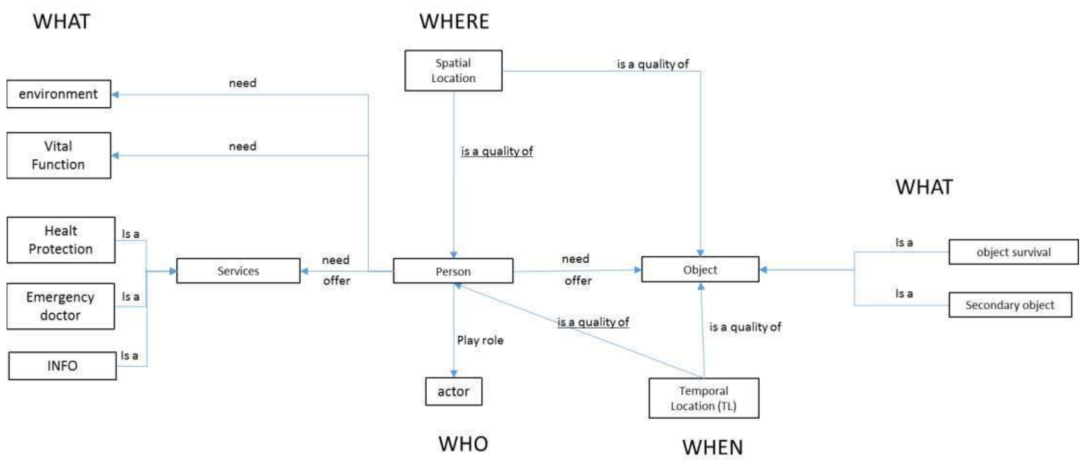


Fig. 4.1 Taxonomy of the Disaster Response Model

The lexical and sintactic forms obtained from the simulation study underline the existence of recurrent forms. These forms establish the rules of belonging to the fields of 'What', 'Where', 'Who', 'When' and 'How'.

4.3.1 Lexical Analysis

Tab. 4.10 Word frequency and cloud words - Haiti earthquake of 2010 Ushahidi Platform illustrates a brief descriptive statistics of the Haiti database retrieved from the Ushahidi platform. The dataset includes words expressed in three languages such as English, French and Creole (the local language of Haitian people). The table shows the word occurency and its relative cloud visualization. Words like food ('cibo'), help ('bisogno'), time ('tempo'), water ('acqua') are the most recurrent in the Haiti Earthquake post-disaster event.

Tab. 4.10 Word frequency and cloud words - Haiti earthquake of 2010 Ushahidi Platform[41]

WORD	OCCURENCY
food	2011
help	1602
time	1293
water	1261
please	1206
people	928
rue	849
manje (food)	607
bezwen (needs)	557
dlo (water)	535
distribution	417
street	416
port	403
live	397
jwenn (find)	328
area	320
tents	618
house	273
tanpri	267
hospital	245
family	226
message	223
shelter	220
children	215

[41] https://datahub.io/dataset/ushahidi/resource/81d058a8-173a-49d9-8ce9-4edf5e7cafc9. Accessed 11/12/2016

Similarly, **Tab. 4.11** shows the word frequency and the cloud visualization of the simulation analysis conducted in the Bari area. As it can be noted, words like help ('aiuto'), need ('bisogno'), water ('acqua'), food ('cibo'), present the highest frequency values. Therefore, the results of the simulation study in terms of word frequencies finds support from the evidence of the Haiti database.

Tab. 4.11 Word frequency and cloud words – Simulation in Bari Metropolitan Area

WORD	OCCURENCY	%
aiuto	79	6,98
bisogno	55	4,86
acqua	54	4,77
cibo	42	3,71
feriti	38	3,36
trovo	38	3,36
bambini	37	3,27
coperte	35	3,09
bari	32	2,83
chiesa	32	2,83
posso	30	2,65
macerie	27	2,39
soccorso	27	2,39
piazza	25	2,21
zona	24	2,12
situazione	23	2,03
pressi	19	1,68
viveri	19	1,68
freddo	18	1,59
casa	16	1,41
servono	16	1,41
soccorsi	16	1,41
giorni	15	1,33
anziani	14	1,24
necessità	14	1,24

Tab. 4.12 represents a correlation matrix between the first 17 most recurrent words in the database of the simulation study carried out in the Bari area. The table shows evidence of a positive correlation between the words food and water (nearly 40%) and the words children ('bambini') and elderly people ('anziani') (about 35%).

Tab. 4.12 Correlation matrix

Attributes	acqua	aiuto	anziani	bambini	bari	bisogno	casa	chiesa	cibo	coperte	feriti	freddo	giorni	macerie	necessi...	piazza	posso
acqua	1	-0.119	-0.016	-0.059	0.013	-0.031	0.009	0.067	0.399	0.154	-0.013	0.066	0.016	-0.047	-0.055	0.051	-0.030
aiuto	-0.119	1	0.085	-0.007	0.116	0.171	-0.033	-0.025	-0.074	-0.041	0.009	-0.016	0.041	0.031	-0.052	-0.060	-0.125
anziani	-0.016	0.085	1	0.352	0.029	0.225	-0.050	-0.022	-0.040	-0.027	0.014	0.079	0.024	-0.066	-0.047	-0.007	-0.066
bambini	-0.059	-0.007	0.352	1	-0.025	0.169	-0.085	0.007	-0.057	-0.004	0.045	0.165	0.057	-0.077	-0.031	0.075	-0.111
bari	0.013	0.116	0.029	-0.025	1	0.039	-0.030	-0.079	-0.071	-0.052	0.192	0.007	-0.026	0.084	0.080	0.018	-0.069
bisogno	-0.031	0.171	0.225	0.169	0.039	1	0.046	0.011	-0.058	0.076	-0.091	-0.078	0.015	-0.022	0.063	0.019	-0.140
casa	0.009	-0.033	-0.050	-0.085	-0.030	0.046	1	-0.030	-0.006	-0.036	-0.084	-0.057	0.084	0.136	-0.050	-0.068	0.068
chiesa	0.067	-0.025	-0.022	0.007	-0.079	0.011	-0.030	1	-0.040	0.081	-0.059	0.007	-0.075	0.009	-0.022	-0.060	-0.069
cibo	0.399	-0.074	-0.040	-0.057	-0.071	-0.058	-0.006	-0.040	1	0.188	0.026	0.024	-0.044	-0.087	0.006	0.023	0.089
coperte	0.154	-0.041	-0.027	-0.004	-0.052	0.076	-0.036	0.081	0.188	1	-0.037	0.043	-0.032	-0.109	0.022	0.083	-0.043
feriti	-0.013	0.009	0.014	0.045	0.192	-0.091	-0.084	-0.059	0.026	-0.037	1	0.034	0.008	0.025	0.014	-0.001	-0.109
freddo	0.066	-0.016	0.079	0.165	0.007	-0.078	-0.057	0.007	0.024	0.043	0.034	1	0.073	-0.027	0.013	0.130	-0.075
giorni	0.016	0.041	0.024	0.057	-0.026	0.015	0.084	-0.075	-0.044	-0.032	0.008	0.073	1	0.091	0.024	-0.011	-0.021
macerie	-0.047	0.031	-0.066	-0.077	0.084	-0.022	0.136	0.009	-0.087	-0.109	0.025	-0.027	0.091	1	-0.066	-0.090	-0.021
necessità	-0.055	-0.052	-0.047	-0.031	0.080	0.063	-0.050	-0.022	0.006	0.022	0.014	0.013	0.024	-0.066	1	-0.007	0.032
piazza	0.051	-0.060	-0.007	0.075	0.018	0.019	-0.068	-0.060	0.023	0.083	-0.001	0.130	-0.011	-0.090	-0.007	1	0.023
posso	-0.030	-0.125	-0.066	-0.111	-0.069	-0.140	0.068	-0.069	0.089	-0.043	-0.109	-0.075	-0.021	-0.021	0.032	0.023	1

4.3.2 Who – Subject

'Who' deals with the actors who, for different reasons, play a specific role in the context of the event.

- A first type of actor is the user who sends the message. He/she can represent an affected or exposed person and, hence is directly involved in the event.
- A second type of actor is an external observer who asks for help on behalf of or offers to help (e.g. volunteeting in rescuing operations) other people. He/she is not directly involved in the event.

The contents of the message is generally ascribed to information including the following sentences: 'mi trovo' ('I am'), or simply 'sono' ('am'). These provide an understanding about the role played by the sender of the message. The following examples also clarify the contents belonging to the 'Who' field.

- 'I am stuck in my house' ('Sono bloccata nella mia abitazione');
- 'I am a rescuer who can offer medical services' ('Sono un soccorritore che può offrire prestazioni mediche');
- 'I am injured in the rubble and surrounded by other people injured who need help' ('Sono ferito tra le macerie circondato da altre persone ferite che hanno bisogno di aiuto')
- 'I am a doctor and can provide assistance in Carrassi area' ('Sono un medico posso fornire il mio aiuto in zona Carrassi').

In the case that the abovementioned contents are reported as third-person plural (e.g.) ('ci troviamo' ('we are'); 'siamo' ('are')) these indicate that the user is not alone; there exists a group of people who is exposed to or affected by the event. The following examples illustrate the above comment:

- ‘We are 11 survivors. We are in the courtyard of a building at no. 5 of lapigia Avenue. The building is completely collapsed and do not see a way out’ (‘Siamo 11 sopravvissuti. Ci troviamo nel cortile di una palazzina al n. 5 di viale Japigia. Il palazzo è interamente crollato non vediamo vie d'uscita’).
- ‘We are nearby the castle, there are many people injured, including young children, food is little’ (‘Siamo in zona castello, ci sono molti feriti, anche bambini molto piccoli, il cibo è ormai scarso’).

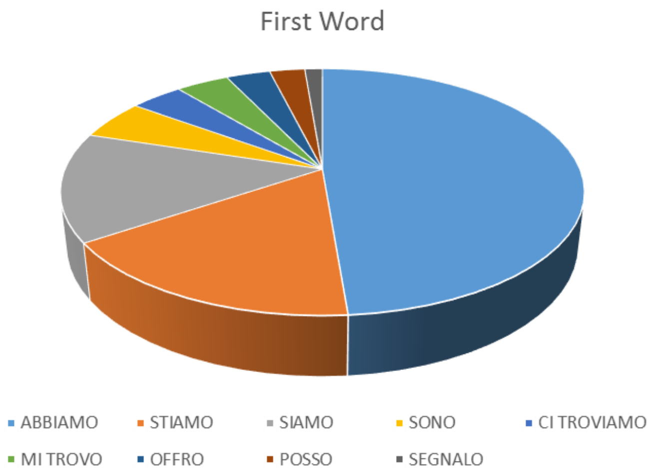
With reference to the actors of the message, the availability of further information such as age range (e.g. adults, children, seniors), their status (e.g. injured) can be seen as features of the actor.

The simulation analysis reveals that 50% of messages (158 by a total of 314) include one of the first words as reported in the table below. These refer to assertive words as indicated by the Speech Act Theory of Austin (1962) and Searle (1969) in Section 2.2.6.

Tab. 4.13 shows that verbal forms are treated as first-person singular or third-person plural and refer to the following verbs: avere (to have), essere (to be), stare (meant as synonim of to be), trovare (meant as a synonim of to be), offrire (to offer), potere (can), segnalare (to indicate/report).

Tab. 4.13 Relevance of the first word in simulation analysis

FIRST WORD	%
ABBIAMO	48,70
STIAMO	16,88
SIAMO	14,29
SONO	5,19
CI TROVIAMO	3,90
MI TROVO	3,90
OFFRO	3,25
POSSO	2,60
SEGNALO	1,30



The Table also shows that the last three first words (i.e. offro, posso, segnalo) indicate the role of the person who is able to offer his/her help. The type of help offered can be revealed by the subsequent terms to which the verbs are related to. The following examples clarify ‘offro’ (offer) and ‘posso’ (can) as commented above:

- ‘I offer help to search for missing people in the rubbles nearby the University campus. Poor medical skills but a lot of courage!’ (‘Offro aiuto di ricerca dei dispersi tra le macerie in zona campus universitario. Scarsa esperienza in ambito medico ma tanta forza d'animo!);
- ‘I offer help in S.Pasquale district of Bari. We are in the square of Saint Pasquale’s church. Basic necessities and blankets available’ (‘Offro aiuto nel quartiere S.Pasquale di Bari. Siamo nella piazzetta della chiesa di San Pasquale Beni di prima necessità e coperte disponibili);
- ‘I offer my technical support’ (‘Offro il mio supporto tecnico’);
- ‘I can offer help for the supply and delivery of basic necessities’ (‘Posso offrire sostegno per fornitura e consegna di beni di prima necessità’);
- ‘I can offer hospitality in my house. 4 people. 99 Dante street. Ring Mario Esposito’ (‘posso ospitare a casa mia 4 persone via dante 99 citofonare Mario Esposito’)

- ‘I can bring food and water if necessary I can help with temporary accomodation’ ‘Posso portare cibo e acqua e se necessario posso aiutarvi nel trovarvi una sistemazione temporanea’);
- ‘I can bring food and beverages..My number is 98989898’ (‘Posso portare viveri e bevande.. il mio numero è 98989898’).

4.3.3 What

In the project Eventory (see Section 4.4) the field ‘What’ refers to an action. It responds, for example, to the question ‘What is happening in this event?’. The message contents in disaster response domain often illustrate the event in general terms (e.g. ‘earthquake in Bari, hurry up’). Alternatively, it is possible to ascribe the event to other effects (e.g. ‘we are next to the train station, a building is crashed, hurry up there are people injured’).

The need to obtain a general information on the event or other information related to the event as mentioned above would respond to the hypothetical question of ‘what is happening in this event?. The answer to this question goes beyond the purpose of the present Thesis’s work. It is relevant for this study to capture and understand the needs required in a message. Therefore, the hypothetical question would be ‘What are the needs in disaster response event?’ The term ‘need’ could refer to an object or a service.

The lexical and semantic analysis of text messages (**Fig. 4.2** and **Fig. 4.3**) shows that the natural language, to express the abovementioed needs, makes use of terms such as ‘need of’ (‘bisogno di’), ‘it needs’ (‘serve’), and ‘they need’ (‘servono’).

N-GRAM = 3

Extract n_gram(3)	TYPE
bisogno_di_soccorso	S
bisogno_di_soccorsi	S
bisogno_di_medici	S
bisogno_di_cure	S
bisogno_di_assistenza	S
bisogno_di_aiuto	S
bisogno_di_viveri	O
bisogno_di_vestiti	O
bisogno_di_medicine	O
bisogno_di_coperte	O
bisogno_di_cibo	O
bisogno_di_beni	O
bisogno_di_alimenti	O
bisogno_di_acqua	O
bisogno_di_un	A
bisogno_di_essere	A

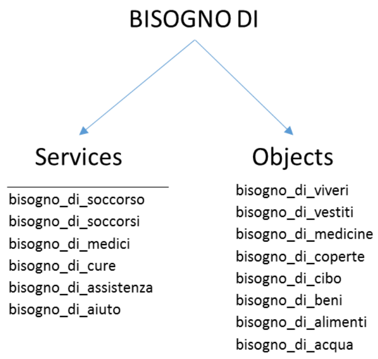


Fig. 4.2 n-grams = 3 analysis - Word = 'Bisogno di' Source - Sample n-grams from the simulation (eartquake survey) analysis

Extract n_gram(4)	TYPE
serve_acqua	O
serve_acqua_assistenza	OS
serve_acqua_cibo	O
serve_aiuto	S
serve_aiuto_subito	SW
serve_cibo	O
serve_cibo_tende	O
servono_acqua_cibo	O
servono_acqua_viveri	O
servono_alimentatori	O
servono_cibo	O
servono_cibo_acqua	O
servono_cibo_acqua_medicinali	O
servono_coperte	O
servono_coperte_per_proteggerci	O
servono_fiammiferi	O
servono_fiammiferi_per_accendere	O
servono_garze	O
servono_garze_dispositivi	O
servono_garze_dispositivi_medici	O
servono_medicinali_centro	O
servono_medicinali_centro_raccolta	O
servono_medicinali_per_curare	O
servono_vestiti	O
servono_vestiti_coperte	O
servono_vestiti_coperte_tende	O
servono_viveri	O
servono_viveri_coperte	O
servono_cure	S
servono_cure_entro	S
servono_cure_entro_un	S
servono_soccorsi	S
servono_soccorsi_medici	S
servono_soccorsi_vi_prego	S
servono_immediatamente	W
servono_immediatamente_soccorsi	W
servono_immediatamente_soccorsi_peucetia	W

Fig. 4.3 n-grams = 3 analysis - Word = 'Serve - Servono' Source - Sample n-grams from the simulation (eartquake survey) analysis

As is often the case with needs, these are not espressed with the natural language. Hence, needs are considered as lantent needs. In the above example ‘we are next to the train station, a building is crashed, hurry up there are people injured’ needs are tacit or latent. It is evident that the existence of people injured means that needs are intended as medical relief, doctors, objects (e.g. medicines, oxigen, water). Also, there exist other latent needs: the need of vital functions such as to be able to breathing and drinking. These would appear more relevant than the needs to receive oxigen or water. The ability of breathing or drinking requires adequate features to let vital functions work. The following examples define the need of vital functions:

- ‘We are under the rubble’ (‘siamo sotto le macerie’);
- ‘The temperature is below zero’ (‘c’è una temperatura sotto lo zero’);
- ‘There is fire and the smoke is saturating the environment’ (‘c’è un incendio e il fumo sta saturando l’ambiente’).

The relevance to understand local knowledge beyond the contents of a message appears clear. Cognitive models should be refined to capture latent knowledge in social network and integrate these models into information systems used in disaster response.

4.3.4 Where – Location

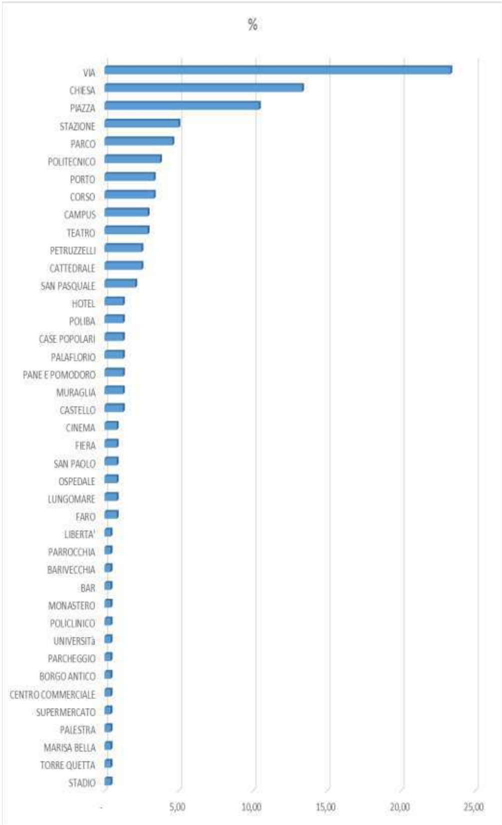
The field ‘Where’ responds to the questions ‘Where the event is located’, ‘Where the person who asks for or offers to help is located’. This field is among the most important and complex to analyse.

Spatial location can be retrieved in terms of geographical coordinations with the use of GPS systems available on the mobile device; by reading a map; and by expert knowledge. Alternatively, spatial location can be inferred by the indication of known places or objects.

Tab. 4.14 Landmarks located in the city of Baribelow provide a useful snapshot of known places (landmarks) retrieved in the simulation analysis. These landmarks are retrieved according to their frequencies and are labelled as ‘generic’ and ‘well-known’.

Tab. 4.14 Landmarks located in the city of Bari

LOCATION	KNOWN	%
VIA	GENERIC	23,33
CHIESA	GENERIC	13,33
PIAZZA	GENERIC	10,42
STAZIONE	GENERIC	5,00
PARCO	GENERIC	4,58
POLITECNICO	GENERIC	3,75
CORSO	GENERIC	3,33
PORTO	GENERIC	3,33
TEATRO	GENERIC	2,92
CAMPUS	GENERIC	2,92
CATTEDRALE	WELL KNOWN	2,50
PETRUZZELLI	WELL KNOWN	2,50
SAN PASQUALE	GENERIC	2,08
CASTELLO	WELL KNOWN	1,25
MURAGLIA	WELL KNOWN	1,25
PANE E POMODORO	WELL KNOWN	1,25
PALAFLORIO	WELL KNOWN	1,25
CASE POPOLARI	GENERIC	1,25
POLIBA	GENERIC	1,25
HOTEL	GENERIC	1,25
FARO	WELL KNOWN	0,83
LUNGOMARE	GENERIC	0,83
OSPEDALE	GENERIC	0,83
SAN PAOLO	GENERIC	0,83
FIERA	GENERIC	0,83
CINEMA	GENERIC	0,83
STADIO	WELL KNOWN	0,42
TORRE QUETTA	WELL KNOWN	0,42
MARISA BELLA	WELL KNOWN	0,42
PALESTRA	GENERIC	0,42
SUPERMERCATO	GENERIC	0,42
CENTRO COMMERCIALE	GENERIC	0,42
BORGO ANTICO	GENERIC	0,42
PARCHEGGIO	GENERIC	0,42
UNIVERSITà	GENERIC	0,42
POLICLINICO	GENERIC	0,42
MONASTERO	GENERIC	0,42
BAR	GENERIC	0,42
BARIVECCHIA	GENERIC	0,42
PARROCCHIA	GENERIC	0,42
LIBERTA'	GENERIC	0,42



Tab. 4.15 Token (2 words) (Generic = indicates a general location. Need to extract token with 3 words)

DESCR	KNOWN
via_de	GENERIC
via_giulio	GENERIC
via_guido	GENERIC
via_re	GENERIC
viale_giuseppe	GENERIC
via_amendola	WELL KNOWN
via_argiro	WELL KNOWN
via_bari	WELL KNOWN
via_caldarola	WELL KNOWN
via_capruzzi	WELL KNOWN
via_cardassi	WELL KNOWN
via_cavour	WELL KNOWN
via_celso	WELL KNOWN
via_crispi	WELL KNOWN
via_dante	WELL KNOWN
via_fanelli	WELL KNOWN
via_fortunato	WELL KNOWN
via_fuga	WELL KNOWN
via_gobetti	WELL KNOWN
via_gramsci	WELL KNOWN
via_grecia	WELL KNOWN
via_japigia	WELL KNOWN
via_kennedy	WELL KNOWN
via_lembo	WELL KNOWN
via_leonida	WELL KNOWN
via_manzoni	WELL KNOWN

DESCR	KNOWN
piazza_alberata	GENERIC
piazza_aldo	GENERIC
piazza_cattedrale	WELL KNOWN
piazza_centrale	GENERIC
piazza_cesare	WELL KNOWN
piazza_ferrarese	WELL KNOWN
piazza_garibaldi	WELL KNOWN
piazza_giulio	GENERIC
piazza_massari	WELL KNOWN
piazza_mercantile	WELL KNOWN
piazza_mercato	GENERIC
piazza_moro	WELL KNOWN
piazza_prefettura	WELL KNOWN
piazza_umberto	WELL KNOWN

DESCR	KNOWN
chiesa_ferdinando	WELL KNOWN
chiesa_lamaddalena	WELL KNOWN
chiesa_moderna	GENERIC
chiesa_paese	GENERIC
chiesa_pasquale	WELL KNOWN
chiesa_redentore	WELL KNOWN
chiesa_russa	WELL KNOWN
chiesa_san	GENERIC
chiesa_sant	GENERIC
chiesa_santa	GENERIC
chiesa_sparano	WELL KNOWN

DESCR	KNOWN
teatro_kurssal	WELL KNOWN
teatro_margherita	WELL KNOWN
teatro_petruzzelli	WELL KNOWN

DESCR	KNOWN
hotel_ambasciatori	WELL KNOWN
hotel_baltico	WELL KNOWN
nicolaus_hotel	WELL KNOWN

DESCR	KNOWN
stazione_capruzzi	WELL KNOWN
stazione_centrale	WELL KNOWN

DESCR	KNOWN
cinema_armenise	WELL KNOWN
cinema_odeon	WELL KNOWN

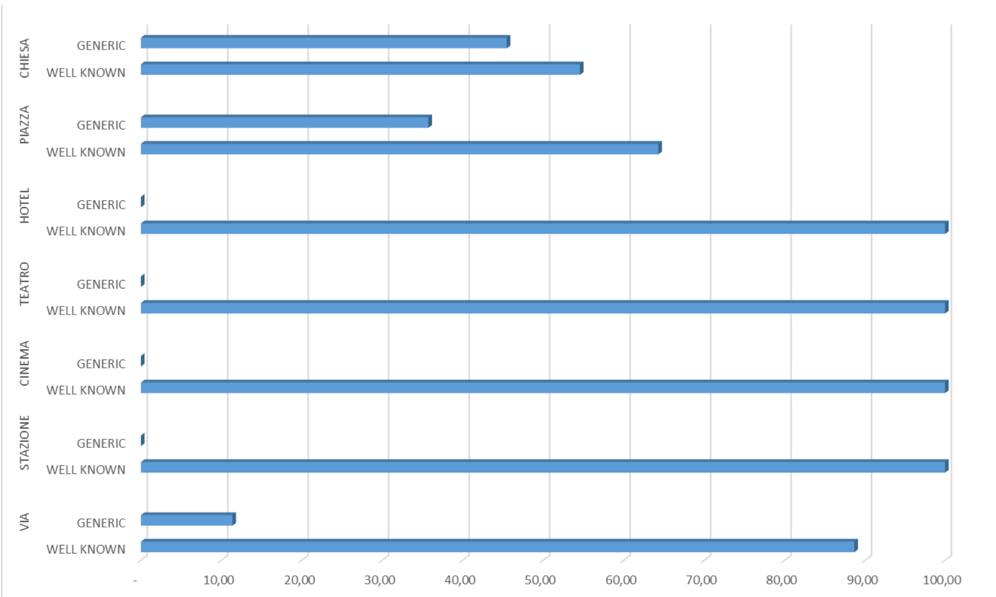


Fig. 4.4 Comparison of places by generic and well known locations

Fig. 4.4 shows that places such as 'cinema', 'theatre' and 'hotel' can be referenced by 2 words n-gram. As for 'Church' (Chiesa) and 'Plaza' (Piazza), 40% of respondents do not need further information to identify them; as for 'Via', a 3 words n-gram is sufficient to identify the location of the street (the place) to which it refers to.

Fig. 4.5 shows the entities extracted by employing a 3 words n-gram. In this case there is a large use of prepositions such as 'next to' ('vicino'), 'nearby' ('vicino/vicino a'), 'in front of' ('di fronte') indicating the spatial relation between the geographical position of people and places. The latter may be represented by addresses or landmarks known at local or global level.

Extract n_gram(3)	TYPE
trovo_a_japigia	P
trovo_al_quartiere	P
trovo_auto	P
trovo_auto_corso	P
trovo_nel_campus	P
trovo_per_strada	P
trovo_piazza	P
trovo_piazza_ferrarese	P
trovo_un_supermercato	P
trovo_via_dante	P
trovo_via_gramsci	P
trovo_via_guido	P
trovo_via_re	P
trovo_via_sparano	P
trovo_viale_capruzzi	P
trovo_vicino_castello	P
trovo_zona_cinema	P
trovo_zona_sheraton	P
trovo_sulla_mia	R
troviamo_di_fronte	R
troviamo_nei_pressi	R
troviamo_nel_cortile	R
troviamo_vicino_al	R
troviamo_vicino_una	R
trovo_nei_pressi	R
trovo_nel_centro	R
trovo_nell_atrio	R
trovo_nelle_vicinanze	R
trovo_prossimità_del	R
trovo_tra_case	R
trovo_vicino_all	R
trovo_vicino_una	R
trovo_zona	R
troviamo_genitori	S
troviamo_genitori_ci	S
troviamo_una_situazione	S
trovo_mio_figlio	S
trovo_insieme_altri	W

R= SPATIAL RELATIONSHIP
S=SEARCH SOMEONE
P=PLACE
W=WITH

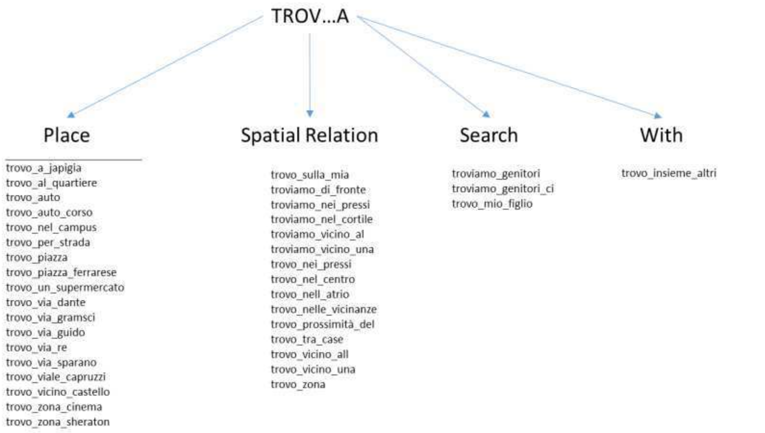


Fig. 4.5 n-grams = 3 analysis - Word = Trov...iamo - Trovo

N-GRAM = 2

Extract n_gram(1-4)	TYPE 1	TYPE 2	COMPLETE
siamo_adulti	E		Y
siamo_famiglie	E		Y
siamo_persone	E		Y
siamo_un	E		N
siamo_una	E		N
siamo_al	L	S	N
siamo_all	L	S	N
siamo_alle	L	S	N
siamo_allo	L	S	N
siamo_davanti	L		N
siamo_giardino	L		Y
siamo_negli	L	S	N
siamo_nei	L	S	N
siamo_nel	L	S	N
siamo_nella	L	S	N
siamo_nelle	L	S	N
siamo_piazza	L		Y
siamo_situati	S		Y
siamo_sotto	L	S	N
siamo_sulla	L	S	N
siamo_via	L		Y
siamo_vicino	L		N
siamo_zona	L		Y
siamo_circa	Q		N
siamo_dieci	Q		Y
siamo_bloccate	S		Y
siamo_bloccati	S		Y
siamo_condizioni	S		N
siamo_feriti	S		Y
siamo_intrappolati	S		Y
siamo_isolati	S		Y
siamo_pronti	S		Y
siamo_radunati	S		Y
siamo_rimasti	S		N
siamo_riuniti	S		Y
siamo_senza	S		N
siamo_sopravvissuti	S		Y

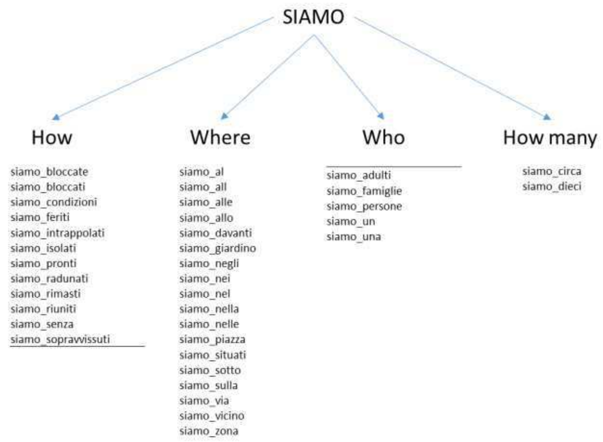


Fig. 4.6 n-grams = 2 analysis - Word = 'Siamo' words extracted by the simulation study (survey)

4.4 Taxonomies

The following sections shows the taxonomies for needs, actors and spatial location.

- Needs are declined into 5 categories such as objects, services, communications, environmental and functional needs. Each of the above categories is split into two sub-fields: 'needs for survival' and 'comfort'.
- Actors are seen as single individual and group of people. For each of these actors specific roles are indentified in post-disaster domain (i.e. exposed people, affected people, assistants, or information providers).
- Spatial locations are defined according to a dual approach: the first one indentifies the elements of a location relative to a reference system; the second one, based on natural language, indicates a location (e.g. address, landmark, meeting places). The abovementioned details will be addressed in the sub-sequent sections.

4.4.1 Spatial Location

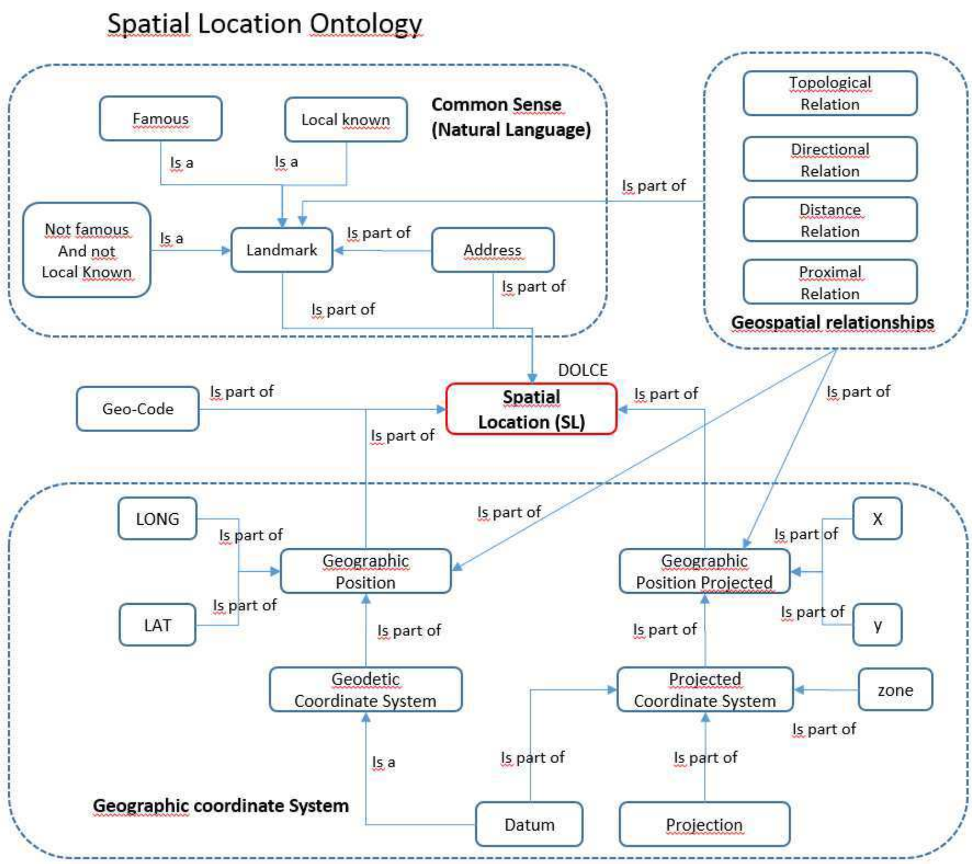


Fig. 4.7 Spatial Location taxonomy

Spatial Location represents the location of an object, an event or an agent.

Fig. 4.7 shows two ways such as *Common Sense* and *Geographic Information System* framework in which an agent uses a message to communicate his/her own location:

1. *Geographic Coordinate System* includes two scenarios:

A spatial location can be obtained through absolute and relative coordinate systems. These can use different reference systems.

- The Geographic Position through a *Geodetic Coordinate System* (identified by a specific type of *Datum* with Latitude and Longitude)
- The *Geographic Postion Projected* through a Projected Coordinate System (identified by a specific type of *Datum* and a Projection with its relative *Zone*) with X (East) and Y(North) coordinates.

The spatial location can automatically be detected by the system if the application allows to do so or if the GPS is turned on and records the location of the user at any time.

2. Common Sense. This is achieved by writing a text message in a *Natural Language*. The user supplies as much information as possible about his/her own location as follows:

- *Landmark. The user refers to and describes a generic place* (e.g., 'red building'). He/she also supplies further elements such as an address (should the location contain one) useful to determine his/her location.
- *Address*. The user indicates the address. This alone is an *istance of Spatial Location.*

3. *GeoCode* (see next sections).

Famous and *Well Known* are specific modes allowing to use universal locations which are indexed in databases containing information about known places (e.g. geonames[42]) or places which are known by local knowledge. Should the latter be the case, local knowledge is determined as an important element to retrieve information.

For this purpose, the creation of databases of local places which can be linked to each other through crowdsourcing and shared are relevant in a disaster response domain.

---

[42] The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge - http://www.geonames.org/

An example of crowdsourcing is the what3words[43] (Fig. 4.8). This is defined as 'link shortening services' which allows the union of three words separated by a dot with no logical meaning. These three words identify a location (*Geocode*) which is included in a geometrical space of dimensions 3 x 3 m. A simple service which opens new useful scenarios in risk/emergency domains (Gould et al. 2016).

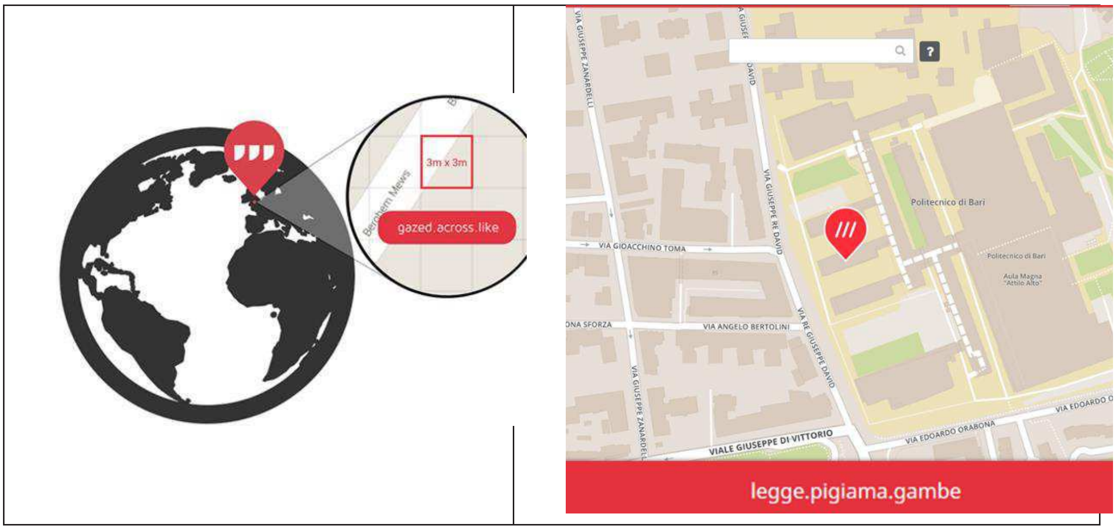


Fig. 4.8 what3words example

According to what3words, the world area is divided in 57 trillion squares. Each *GeoCode* identifies a square with 3 pre-assigned unrelated words. This geocode is then the spatial location of a given 3 x 3 m area.

The innovative aspect of What3words is that it is easy to memorize by non-expert agents in risk/emergency situations and it is also simpler compared to a latitude and longitude system. What3words is available on mobile devices or personal computers. It is also open to be integrated with other apps through API libraries. It has a database of about 25.000-40.000 headwords (the algorithm is less than 10 MB) for every language. To favour the

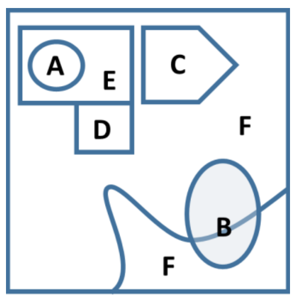
[43] http://map.what3words.com/research.hard.work

retention of accustomed locations, simple headwords are associated with urban areas; while complex headwords are associated to sub-urban areas.

Another framework for spatial location is the *Geospatial RelationShip*. It contains information of a spatial location according to Geographic Position, *Geographic Postion Projected, Landmark* and Address. The use of *Geospatial RelationShip* enriches actual information with further spatial elements (Longley et al. 2011, Xu, 2014).

Tab. 4.16 shows Geospatial RelationShip such as Topological, Directional, Distance, Proximal.

Tab. 4.16 Geospatial RelationShip example



Topological	Directional	Distance	Proximal
A inside E	C north of F	C at 100 m from E	C near E
D connected To E	A east C		F far from A
C disjoint E			
B overlaps E			

Topological, Directional, Distance, Proximal express spatial relations between geometric primitives (points, polylines and polygons). A regional space can be modelled by the use of these geometries. A spatial location can be represented by a polygon (e.g. a plaza), by a point (e.g. bus stop) and by a line (e.g. a street). The relationships occurring between these objects identifies useful information on spatial location between two or more objects. Therefore, expressions like ‘I am outside the train station’ is similar to ‘C disjoint E’ in Tab. 4.16; ‘I am nearby the church’ is ‘C near E’; ‘I am at 500 m from the University’ is ‘C at 500 m from E’ or also ‘We moved to North compared to the point 723000, 4523000 – WGS84 UTM 33N’.

4.4.2 Needs

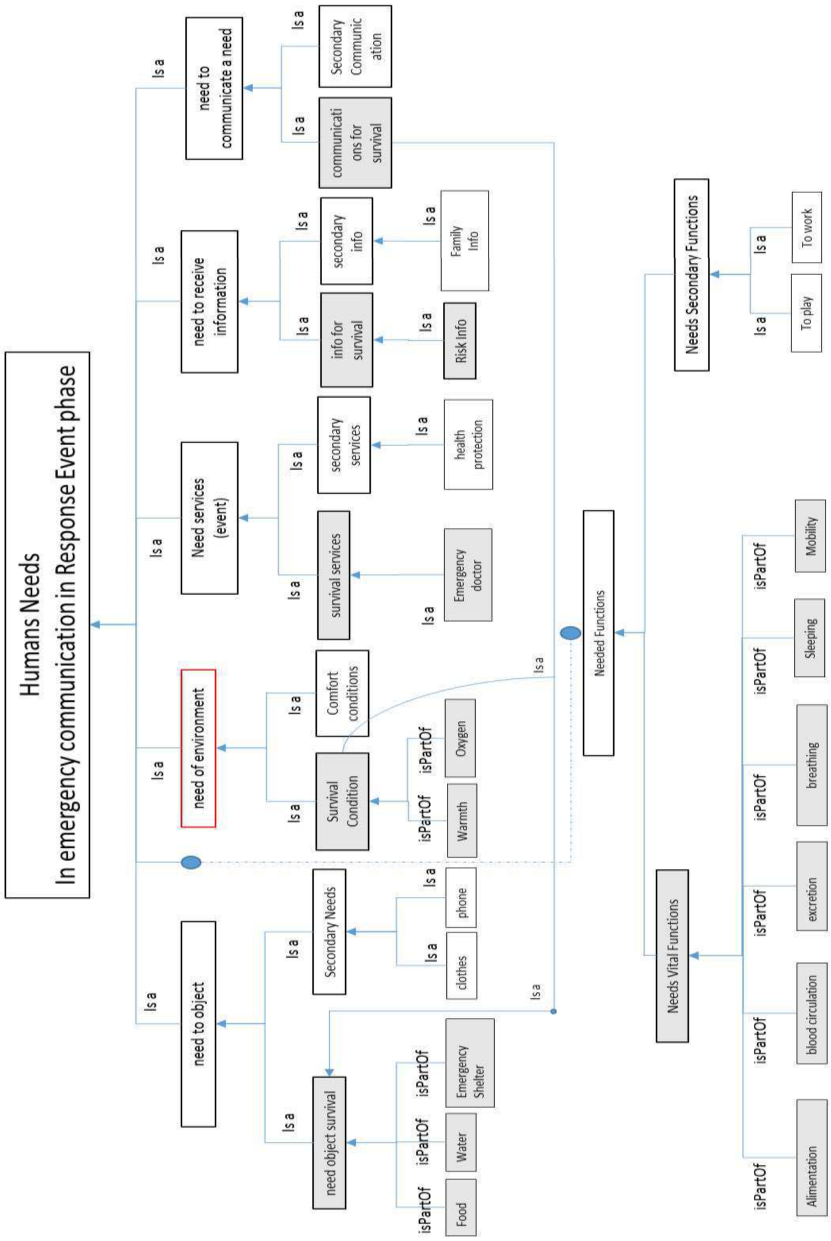


Fig. 4.9 Needs taxonomy

The taxonomy of ‘needs’ defines a domain which comprises of the following categories: a) *need Object*; b) *need Event (services)*; c) *need Environmental condition*; d) *Vital Functions*; e) *need to Receive Information;* f) *need to Communicate Need*. These elements constitute main needs in disaster response.

The taxonomy is supported by the international literature such as the World Health Organization and the King’s College London (2011) which have analysed and defined the needs for specific events.

For each category above, two types of needs have been considered: ‘needs for survival’ and ‘comfort’. The category ‘need Object’ explains the satisfaction of a material need. Although an object is necessary for survival this does not ensure the survival of the individual.

From an ontological point of view, if a need is located in the categories of ‘*need environmental conditions*’ and ‘*vital functions*’, these locations do not fully capture their conceptual meanings. For example, water is a primary need for survival. This condition is necessary (water availability) but not sufficient to ensure survival. At least two other conditions should be met: i) the water is drinking water (*need Environmental Condition*); and ii) the agent who asks for water is able to drink (*Vital Function*).

The categories ‘*need to Receive Information*’ and ‘*need to Communicate Need*’ refer to requesting information (e.g. information on available foothpaths to reach a specific location) or to share information (e.g. information on available foothpaths to reach a group of affected people).

The remaining type of need defined as ‘comfort’, refers to objects and activities which are not essential for the survival of people. Nonetheless, comfort needs can supply additional information which can be useful to trace a request for or an offer to help.

4.4.3 Actors

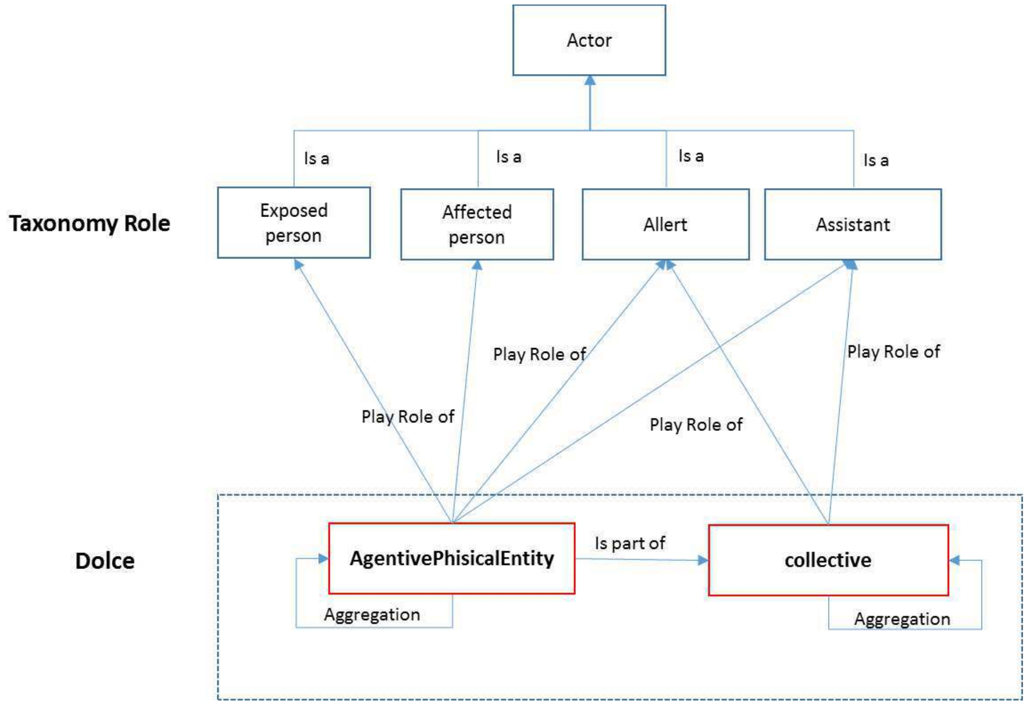


Fig. 4.10 Taxonomy of actors and role

The ontology *Actor* is used to define the actors of a domain. The description of an agent/person can be unbiased (subjective) compared to other elements such a spatial location. In a specific domain, actors refers to multiple scenarios and can cover multiple roles (Ye et al., 2007).

In the domain of the present Thesis (i.e. disaster response), actors are defined as:

AgentivePhisicalEntity and *Collective* or private, public, non-profit organizations which can offer a service to satisfy a need (e.g. medical or transport services) or can request a service from an individual or group of people.

AgentivePhisicalEntity and *Collective* actors can cover one or more roles. These actors can be exposed or affected people (*exposed person*), can offer collaboration and information

(*Allertant*) or hold speficic skills (*Assitant)*. *Collective* actors can be both *Assistant* and *Allertant.*

5 Discussion and Conclusions

The general aim of the present doctoral research was to contribute to tapping into local knowledge to support disaster response. To that purpose, this contribution was based on the application of cognitive, predictive and ontological models to the study of text message exchange in relevant Social Networks.

The lack of studies on local knowledge in risk domains and the growing attention of scholars and decision makers alike to distributed governance and citizen participation (taking place before, during and after a disaster event) have moved the present work to address the following three research questions:

1) *How relevant is local knowledge to the governance of disaster response?*
2) *How useful is the interpretation of text messages to grasp local knowledge in disaster response?*
3) Under what conditions, and with what limitations, *may cognitive, predictive and ontological models facilitate the interpretation of text messages in the context of disaster response?*

How relevant is local knowledge to the governance of disaster response?

The conceptual background to the first research question has been extensively discussed in **Section 2.2**. Despite remarkable advances in scientific knowledge and ever faster technological innovations, natural hazard-related risk and vulnerability seem to have increased in both industrialized and developing countries (Gardner 2002; Van Aalst e Burton 2002; Dekens, 2007). Morevoer, while there exist several studies on risk and expert knowledge (especially in computer science literature), the relationships between risk and local knowledge are still erratically investigated and poorly understood. Following a long-standing distrust of local knowledge, scholars are now turning to it with ever-increasing interest: while acknowledging the difficulties in bridging the divide between local and expert knowledge (Gaillard, J. C., & Mercer, 2013), there is now a growing awareness of the wealth of information being embedded into local knowledge, all the more so when they are strictly linked to behaviours, beliefs, and lifestyles, pertain to highly sensitive domains (like risk), or concern spatial knowledge and territoriality (Smith, 2011).

The empirical work carried out for the present doctoral research seem to confirm the as promising as untapped potential of local knowledge to support the governance of disaster response. The analysis of hundreds of messages collected in real or simulated disaster response contexts has highlighted two main aspects:

- firstly, even in emergency situations, human communication seems to be imbued with culturally mediated and place-based understanding of spatiality, relationality and actions;
- secondly, natural language appears to be locally articulated (in vernacular and dialectal forms) and thus using terms that can only be understood within specific networks or communities.

Hence, the accuracy of any text mining methodology to assist disaster response is strongly dependent upon its ability to detect and decode the peculiarities inherent to these personalized, informal and context-laden languages.

How useful is the interpretation of text messages to grasp local knowledge in disaster response?

As for the second research question, this study helped advance the conceptual understanding of local knowledge's role in disaster response, and to support the relevance of *speech act* theory. The latter has, in turn, shed light on the understanding of communication by focusing on the intentions of speakers to express certain situations, feelings, states of mind. The investigation of text messages within relevant social networks has highlighted that these messages – even when exchanged in the framework of disaster response activities – seem to have the same features of Austin's speech act.

A dedicated simulation proved key to addressing whether and how the interpretation of text messages could help integrate local knowledge into disaster response. The simulation was carried out by administering an open-ended on-line questionnaire, in which participants were asked to write requests for, or offers of, help. Both structured and un-structured data were collected and stored in a single database containing a total of 1300 text messages, of which 314 observations concerned the responses to the on-line simulation survey, the rest originating from e-participation initiatives in the metropolitan area of Bari (Italy).

A relevant issue that the author was confronted with over the whole empirical work cycle consisted in the pre-treatment of data in terms of text-mining approach, which involved *information retrieval*, *information extraction* and *information mining*. When applied to the earthquake simulation study in the metropolitan area of Bari (Italy), the analysis singled out a few highly important terms of the like of ‘water’, ‘help’, ‘need’, ‘food’, and ‘children’. It is interesting to note that, when comparing such outcomes to those of a parallel analysis targeting the Haiti 2010 earthquake messages dataset, a very similar lexicon emerged which focussed on primary needs. The analogy is likely to hinge upon a powerful hierarchy of needs under emergency conditions, which is apparently able to offset the striking differences in both socio-economic (a highly industrialized urban region vs one of the least developed country) and linguistic (the Haiti messages were exchanged in three different languages – Creole, French and English; the simulation’s in Italian only) terms. It should not, however, go unnoticed that local-and tacit-knowledge-related contents, embodied in natural languages, were detected and investigated only in the Bari simulation survey, because of cultural barriers that proved insuperable in the framework of the present research. Not surprisingly, this drawback should be addressed by future efforts, for instance by teaming up with culturally diverse investigators to replicate similar research designs in different contexts or, in the longer term, to strengthen international networks of scholars and practitioners who appreciate the manifold global/local links (on both methodological and operational ground) in disaster response.

As presented in **Section 4.2,** 50% of the messages collected over the simulation analysis (158 out of 314), verbs as “stare” (meant as a synonym of *to be*), “trovarsi” (*to find onself*, *to be*), “offrire” (*to offer*), “potere” (*to can*), “segnalare” (to indicate, to report), may all be framed following the above mentioned *speech act theory* (Austin, 1962; Searle, 1969), along the lines of what Ogasawara and Ginsburg (2015) observed about the 2014 Japanese tsunami. In most cases, these verbal forms introduce the description of either a need or a geographical location, and hence have proved a useful cognitive tool to identify and mark the boundaries of relevant phrases. Adding to the existing literature, the present study tries to grasp the exclusively local linguistic subtleties, and to put their inherent ambiguities to the test – as it was discussed with regard to the direction of the Muraglia promenade in Bari (**Section 4.2**), according to which a start and an end of the walls could be identified. This is a crucial issue, since not only in disaster response, but in everyday communication alike, terms are being used which do not fall into a universally defined domain, but rather shift form one context-specific domain to another and their meaning is fully and univocally understandable only when the involved agents share the same relevant knowledge base.

The correct interpretation of local knowledge-laden natural language becomes an even more challenging problem, when the role of tacit or implicit knowledge is taken into account, as agents restrain themselves from declaring those concepts or information which they consider would be taken for granted within their group or community. Therefore, and based on the research work that has been carried out, the present dissertation argues in favour of acknowledging the importance of local knowledge in disaster response – consistently with the position of many scholars (Nonaka, 2001; Raymond at al. 2010; Olaide and Omolere, 2013) – and calls for implementing spatial data science tools (databases, vocabularies, maps, models, *etc.*) in such a way that local and tacit knowledge (including vernacular forms) for primary needs and landmarks may be adequately understood and modelled.

To advance in this direction, it appears that distributed crowdsourcing methods, combined with collaborative hypertext editors (wikis), may be effective in many fields (Imran *et al.* 2014). A number of recent trends have been dealt with in the conceptual background (**Section 2.4**), such as VGI and PPGIS, as well as the use of all-purpose social networks (Twitter, Facebook, *etc.*) to send out disaster-related messages that could reach well beyond the institutional body in charge of traditional top-down frameworks, to the wider internet community.

One of the empirical activity included in this research's design (the earthquake simulation study in the metropolitan area of Bari) adopted the Ushahidi 3.0 platform (a PPGIS-oriented tool), to collect 314 messages in a limited time frame. (Participants were not trained in the use of the tool). It was observed that only 2% of the messages were geotagged, by including the spatial coordinates of the places they referred to. Hence, without text analysis it would have been impossible to extract the necessary place-related information to localize people and objects.

To reflect on such a low share of geotagged messages, it should be first noted that geotagging entails a series of steps:

- sending a message from a GPS-equipped terminal
- using a platform, tool or social network that provides for geotagging, and
- activating and/or authorising the use of GPS by the application.

Privacy-related issues are clearly at stake when exploring the drivers behind the low share of spatially explicit messages being shared. It could be argued that, the many contentious issues that ensue from the rise of the information society notwithstanding, at least in the context of disaster response (and emergency situations in general) consensus could easily build up around the need to activate positioning services to increase the likelihood of timely and effective aid.

Another reflection triggered by this research work concerns the need to raise awareness on effective text message writing in emergency situations, given the format limitations imposed by most web and mobile services (e.g. the maximum number of characters). A possible policy recommendation would then point to including effective writing – for instance, according to the W3C Incubator Group Report model (Who, What, Where and When) – into emergency drill programmes (Atit and Sundaram 2007; Ianella, 2009). Likewise, simplified techniques to implement geocoding without any prior training in GIS are being put forward to open up new perspectives in risk/emergency domains (Gould *et al.* 2016). This is the case of *What3words*, a "link shortening service" that allows the mix of three dot-separated words with no logical meaning to identify a specific location (*Geocode*), i.e. a 3m square cell size.

Under what conditions, and with what limitations, may cognitive, predictive and ontological models facilitate the interpretation of text messages in the context of disaster response?

With respect to the third research question, the underlying cognitive, predictive and ontological models have been extensively discussed in **Sections 2.2**, **2.3** and **3.4**. The resulting, integrated, modelling approach mixed machine learning and ontologies, and led to four main observations.

Firstly, as disaster response is likely to happen in unpredictable conditions where different types of threats interact to cause an array of direct and indirect, complex effects (Coburn *et al.*, 2014), the relationships among these factors should be carefully conceptualized and fully integrated into the chosen model. With climatic and environmental catastrophes being the threats most likely to trigger further socio-economic and ecological effects on the affected population, this study pointed to text analysis as being key to disentangle relevant information from messages that – though blurred with local and implicit knowledge – are often the only available direct source of messages sent by survivors and aid workers alike.

Secondly, and following the demand for innovative methods that ensues from the above reflection, the present research added to the evidence in favour of ontologies as being an adequate approach to disaster response. In such framework, the effective interpretation of text messages was attempted at by building a shared conceptualization of risk. To this purpose, three separate taxonomies were developed (regarding Spatial Location, Needs and Actors – see **Section 4.4**), each being linked to a terminal entity in DOLCE foundational ontology (Masolo *et al.*, 2002):

- 'PhysicalObject', for *needs*,
- "SpatialLocation', for localizations and spatial relationships,
- 'AgentivePhisicalEntity', for actors.

Thus, the ontological framework was aligned to that of DOLCE's foundational ontology. This internationally-renowned ontology may represent the coordination apparatus among locally-differentiated knowledge information systems, with a view to enhancing knowledge sharing. Such effort should however be complemented by the development at international level of shared disaster- and risk-related ontologies – in finer-grained details, so as to refine the linkages between the ontological entities and the concepts embodied in the natural language forms extracted from text messages.

A third aspect that may be singled out by reflecting on the empirical analysis is the effectiveness of predictive models to classify information from text messages. Through a 5-fold cross-validation framework of 50 iterations, the following predictive models have been used: Adaboost, Random Forest, SVM, Neural Networks, and Naïve Bayes. The Random Forest and the Neural Network emerged as the two most successful classifiers. The former yielded a training and test accuracy of, respectively, 94.1 (± 1.5)% and 96.5 (± 0.8)%,. For the latter, the figures were 93.2 (± 1.6) %and 95.2 (± 1.1)% for training and test sets, respectively.

In line with the outcomes of other studies, the chosen models proved to be reliable enough in sorting data over two domains, when a matrix detecting the presence or absence of a set of terms was applied. The resulting lexicon is relatively specific, as most messages converge on the use of few terms ('water', 'food', 'shelter', 'to be', 'to need', *etc.*), which however convey universal concepts that are largely independent from specific places and knowledge.

Concluding remarks

This research work is not without limitations. First, the machine learning and ontological models have only partly been integrated to each other. While machine learning is an application-oriented approach, the ontological framework is a higher-level conceptual construct. Although the analysis of the context (i.e. the disaster response interactions) has been studied through an integrated approach between these two modelling approaches, the integration was only tested at a conceptual level. The integration of an ontological framework in actual web platforms falls outside the scope of this thesis.

An important improvement that may be made to the present work in the near future is to consider some form of empirical integration between the ontological and machine learning approaches in VGI/PPGIS technologies. In other words, future research could focus on the design and implementation of integrated platforms to collect, retrieve and analyse unstructured data, and communicate structured knowledge to policy makers and citizens.

Another drawback of this work is information reliability when resorting to crowdsourcing. The identification of *false negatives*, most notably, by text classification systems may undermine the robustness of classification models. As a result, these types of messages are neither included in disaster response domains nor considered as local knowledge. Future developments of the present work should therefore be geared towards advancing text classification models and improving the analysis of latent knowledge in the disaster response domain.

Based on the above considerations, it may be concluded that the use of natural language should be explored case-by-case, with due consideration of the specific place-based, socio-cultural settings where disaster events occur. On a parallel track, special attention should be paid to sharing advances in the foundational cognitive structures that underpin sense-making and speech acts in disaster response. Therefore, the prospects for a better integration of local knowledge into the governance of disaster response depend on the ability, within the wider community of scholars and practitioners, to proceed along two apparently diverging paths – and yet boost mutual learning and coordination.

Acknowledgements

I just realised that 12 years have passed, since I took the decision to start with degree and post-degree studies.

If I am where I am now, is because I was truly blessed by those I have met along the way who, for their affection before their professional expertise, I intend to express my gratitude.

I thank my supervisors, Dino Borri and Caterina De Lucia, who, although from different perspectives, have taught me to be a curious and rigorous researcher. Their advices have been crucial for the development of the present work.

A further appreciation is to Dino Borri for being a constant source of knowledge and inspiration (often without realising it) and for the chance to participate in research activities long before the doctoral studies; to Caterina De Lucia for her patience and diligence who guided me through each step of this work.

I thank Angela Barbanente and Domenico Camarda, Claudia Ceppi, Francesco Mancini, Giulia Mastrodonato, Mauro Patano, Rocco Rubino, Rosella Stufano and Eufemia Tarantino for the time spent to share views and opinions on various research issues covered in this work.

I am greatly indebted to Alessandro Bonifazi and Carmelo Torre, true friends above all, who provided the foundation for my research career.

My sincere thank goes to Francesco Amoroso, Roberto Bellotti and Stefano Borgo who contributed, with their advices, to improve the methodological part of the present work.

A particular thank I reserve to my family:

To my mother and my father who, with their simplicity and kindness, have taught me to put the right value upon things.

To Marcella for her unconditional support to the choices I have made, for her encouragement during tough times and to have “beared” my absences…

To Giulia and Claudia for starting the ‘countdown’ over the last year of studies, and to remind me, day by day, that the next years will be dedicated to spend time with them and have fun together.

To my sister and my brothers and my brothers and sisters-in-law, and my aunt Rosa for their love and affection throughout these years.

References

Abbasi, M.-A., Kumar, S., Filho, J.A.A., Liu, H. (2012). *Lessons learned in using social media for disaster relief - ASU crisis response game, in Social Computing, Behavioral - Cultural Modeling and Prediction*. Springer Nature, pp. 282–289.

Acar, A., Muraki, Y. (2011) Twitter for crisis communication: lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, 7(3), 392-402.

Ackoff RL. (1989) From data to wisdom. *Journal of Applied Systems Analysis*, 16(1):3-9.

Alavi, M., Leidner, D. E. (2001) Knowledge Management and Knowledge Systems : Conceptual Foundations and Research Issue. *MIS Quarterly*, 25(1), 107–136.

Albtoush, R., Dobrescu, R., Ionescou, F. (2011) A Hierarchical Model for Emergency Management Systems. *University' Politehnica of Bucharest Scientific Bulletin, Series C: Electrical Engineering*, 73(2), 53-62.

Alexander, D.E. (2002) *Principles of Emergency Planning and Management* (New York: Oxford University Press).

Ambrosini, V., Bowman, C. (2001) Tacit knowledge: Some suggestions for operationalization. *Journal of Management studies*, 38(6), 811-829.

Carvin, A. (2013). *Distant witness*. CUNY Journalism Press, 204-226.

Apisakmontri, P., Nantajeewarawat, E., Buranarach, M., Ikeda, M. (2013) Towards the Use of Upper. *2nd Asian Conference on Information Systems* (ACIS2013).

Argyris, C., Schön, D.A. (1979*), Organizational Learning: A Theory of Action Perspective*. AddisonWesley, Reading, MA.

Austin, John L. (1962). *How to do things with words*, Oxford, Oxford U.P., II ed. riv. 1975.

Aven, T. (2013) A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy. *Reliability Engineering and System Safety*, 111, 30–36. http://doi.org/10.1016/j.ress.2012.09.014

Awad, E. M., Ghaziri, H. M. (2004). *Knowledge management*, 2004. ed: Prentice-Hall, Upper Saddle River, New Jersey.

Basari, A. S. H., Hussin, B., Ananta, I. G. P., Zeniarja, J. (2013) Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, 53, 453–462. http://doi.org/10.1016/j.proeng.2013.02.059

Babitski, G., Bergweiler, S., Grebner, O., Oberle, D., Paulheim, H., Probst, F. (2011) SoKNOS – Using Semantic Technologies in Disaster Management Software. *8th Extended Semantic Web Conference* (ESWC 2011), 183–197. http://doi.org/10.1007/978-3-642-21064-8_13

Baharin, S., Shibghatullah, A., Othman, Z. (2009) Disaster management in malaysia: An application framework of integrated routing application for emergency response management system. In *International Conference of Soft Computing and Pattern Recognition (SOCPAR)*. (pp. 716 –719)

Barnes, S. (2002). *Knowledge management systems: Theory and practice*. Cengage Learning EMEA.

Baškarada, S., Koronios, A. (2013) Data, information, knowledge, wisdom (DIKW): A semiotic theoretical and empirical exploration of the hierarchy and its quality dimension. *Australasian Journal of Information Systems*, 18(1), 5–24. http://doi.org/10.3127/ajis.v18i1.748

Bayes T. (1763) An essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.

Berkes, F., Folke, C., (2002) Back to the future: ecosystem dynamics and local knowledge. In: Gunderson, L.H., Holling, C.S. (Eds.), *Panarchy: Understanding Transformations in Human and Natural Systems*. Island Press, Washington, DC, pp. 121–146. Berkhout,

Bernstein, J. H. (1989). The data-information-knowledge-wisdom hierarchy and its antithesis. *NASKO*, 2(1), 68-75.

Biagetti, M. T. (2010). Le ontologie come strumenti per l 'organizzazione della conoscenza in rete , AIDA Informazioni.

Bishr, M., Janowicz, K. (2010). Can we Trust Information? - The Case of Volunteered Geographic Information. In *Workshop at Future Internet Symposium Towards Digital Earth: Search, Discover and Share Geospatial Data*

Blair, D. C. (2002) Knowledge management: Hype, hope, or help? *Journal of the American Society for Information Science and Technology*, 53(12), 1019-1028.

Blum, J. R., Eichhorn, A., Smith, S., Sterle-Contala, M., Cooperstock, J. R. (2014) Real-time emergency response: improved management of real-time information during crisis situations. *Journal on Multimodal User Interfaces*, 8(2), 161–173. http://doi.org/10.1007/s12193-013-0139-7

David Boddy, Albert Bonnstra & Kennedy, G., (2008). Managing Information Systems: Strategy and Organisation. In *Managing Information Systems.*

Bolasco S., Canzonetti A., Capo F. (2005). *Text Mining - Uno strumento strategico per imprese e istituzioni*. Cisu Editore, Roma.

Bolasco, S. (2005) Statistica testuale e text mining: alcuni paradigmi applicativi. *Quaderni Di Statistica,* 7, 17–53.

Bond, A. J., Viegas, C. V., Coelho de Souza Reinisch Coelho, C., Selig, P. M. (2010) Informal knowledge processes: the underpinning for sustainability outcomes in EIA? *Journal of Cleaner Production*, 18(1), 6–13. http://doi.org/10.1016/j.jclepro.2009.09.002

Borgo, S., Masolo, C. (2010) Ontological foundations of DOLCE. *In Theory and applications of ontology: Computer applications*, 279-295. Springer Netherlands.

Borst, W. N. (1997) Construction of engineeriing ontologies for knowledge sharing and reuse. Technology. PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands

Breiman, L. (2001) Random forests. *Machine learning*, 45(1), 5-32.

Brown, G., Kyttä, M. (2014) 'Key issues and research priorities for public participation GIS (PPGIS): a synthesis based on empirical research'. *Applied Geography*, 46, 122–136. http://doi.org/10. 1016/j.apgeog.2013.11.004.

Brown, G., Kelly, M., Whitall, D. (2014) Which 'public'? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. *Journal of Environmental Planning and Management*, 57(2), 190–214. http://doi.org/10.1080/09640568.2012.741045

Bruns, A. (2014) Crisis communication. *The media and communications in Australia*, 351-355.

Bruns, A., Stieglitz, S. (2012) Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services*, 30(3-4), 160-185.

Budzynska, K. ., Reed, C. (2011) Speech acts of argumentation: Inference anchors and peripheral cues in dialogue. AAAI Workshop - Technical Report, WS-11-10, 3–10.

Budzynska, K., Janier, M., Reed, C., Saint-Dizier, P. (2016) Theoretical foundations for illocutionary structure parsing. *Argument & Computation*, 7(June 2013), 91–108. http://doi.org/10.3233/AAC-160005

Bulkeley, H. (2000) Common knowledge? Public understanding of climate change in Newcastle, Australia. *Public Understanding of Science*, 9, 313–333.

Canini, K. R., Suh, B., Pirolli, P. L. (2011) Finding credible information sources in social networks based on content and social structure. *In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)* (pp. 1-8).

Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H. W., Mitra, P., Yen, J. (2011) Classifying text messages for the haiti earthquake. In *Proceedings of the 8th international conference on information systems for crisis response and management* (ISCRAM2011).

Carlisle, J. P. (2015) Continuing the DIKW Hierarchy Conversation. *MWAIS 2015 Proceedings*.

Castillo C., Mendoza M., Poblete B. (2013) Predicting information credibility in timesensitive social media. *Internet Research,* 23, 5 (2013), 560–588

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D. S. (2014) Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers and Graphics (Pergamon)*, 38(1), 51–60. http://doi.org/10.1016/j.cag.2013.10.008

Clarke-Guarnizo, C. (1992) 'Living with Hazards. Communities' Adjustment Mechanisms in Developing Countries'. In Kreimer, A. Munansinghe, Y.M. (eds*) Environmental Management and Urban Vulnerability*, pp 93-106. Washington DC: World Bank.

Cobb, C., McCarthy, T., Perkins, A., Bharadwaj, A., Comis, J., Do, B., Starbird, K. (2014). Designing for the deluge: understanding & supporting the distributed, collaborative work of crisis volunteers. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 888-899). ACM.

Coburn, A. W., Bowman, G., Ruffle, S. J., Foulser-Piggott, R., Ralph, D., Tuveson, M. (2014) *A Taxonomy of Threats for Complex Risk Management*. Cambridge Risk Framework series.

Cohn, J. P. (2008). Citizen science: Can volunteers do real research? *Bioscience*, 58 (3), 192–197.

Coleman, D. J., Georgiadou, Y.,Labonte, J. (2009) 'Volunteered geographic information: the nature and motivation of produsers', *International Journal of Spatial Data Infrastructures Research*, 4, pp. 332–358. http://doi.org/10.2902/1725-0463.2009.04.art16

Cronin, S. J., Gaylord, D. R., Charley, D., Alloway, B. V., Wallez, S., Esau, J. W. (2004) Participatory methods of incorporating scientific with traditional knowledge for volcanic hazard management on Ambae Island, Vanuatu. *Bulletin of Volcanology*, 66(7), 652–668. http://doi.org/10.1007/s00445-004-0347-9

Crozier, M. J., T. Glade (2005). Landslide hazard and risk: issues, concept and approach. Chichester, England, John Wiley and sons.

Cutler, A. (1983) Lexical complexity and sentence processing. *The Process of Language Understanding*, 43–79.

Cutter, S. L. (2001). *American Hazardscapes: The Regionalization of Hazards and Disasters*. Washington DC: The Joseph Henry Press.

Chaffey, D., White, G. (2010). *Business information management: improving performance using information systems*. Pearson Education.

Davenport, T. H., Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.

De Longueville, B., Smith, R. S., Luraschi, G. (2009). Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 international workshop on location based social networks (pp. 73-80)*. ACM.

Dekens, J. (2007). *Local Knowledge for Disaster Preparedness: A literature Review.* International Centre for Integrated Mountain Development (ICIMOD), 97. Retrieved from http://lib.icimod.org/record/22470\nhttp://lib.icimod.org/record/22470/files/DPL.pdf

Delen, D., Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining*. Expert Systems with Applications*, 34(3), 1707–1720. http://doi.org/10.1016/j.eswa.2007.01.035

Di Lorenzo Ajello, F. (2001). *Mente, azione e linguaggio nel pensiero di John R. Searle*.

Diakopoulos, N., De Choudhury, M., Naaman, M. (2012, May). Finding and assessing social media information sources in the context of journalism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2451-2460). ACM.

Drabek, T. (2002) *The Social Dimensions of Disaster (FEMA Emergency Management Higher Education Project College Course Instructor Guide)*. Emmitsburg, MD: Emergency Management Institute.

Drupsteen, L., Guldenmund, F. W. (2014) What is learning? a review of the safety literature to define learning from incidents, accidents and disasters*. Journal of Contingencies and Crisis Management*, 22(2), 81–96. http://doi.org/10.1111/1468-5973.12039

Dulli, S., Polpettini, P., Trotta, M. (2004) *Text mining: teoria e applicazioni* (Vol. 19). Franco Angeli.

Dynes, R. R. (1998). Coming to Terms with Community Disaster and on the Concept of Disaster: A Response, a Slight Recovery and Less Reconstruction. University of Delaware, Disaster Research Center.

Elwood, S. (2008). Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, 72(3-4), 173–183.

Elwood, S. (2010). Geographic information science: emerging research on the societal implications of the geospatial web, *Progress in Human Geography*, 34(3), pp. 349–357.

Fabricius, C., Sholes, R., Cundill, G., (2006) Mobilizing knowledge for integrated ecosystem assessments. In: Reid, W.V., Berkes, F., Wilbanks, T., Capistrano, D. (Eds.), *Bridging Scales and Knowledge Systems: Concepts and Applications in Ecosystem Assessment*. Island Press, Washington, pp. 165e182.

Fayyad, U. Piatetsky-Shapiro, G., P. Smyth, (1996) . From data mining to knowledge discovery: an overview. In *Advances in Knowledge Discovery and Data Mining*, pages 1-34. AAAI Press

Farazmand, Ali. (2001). Introduction – Crisis and Emergency Management. Chapter 1 in Ali Farazmand (ed.), *Handbook of Crisis and Emergency Management*. New York and Basel, Marcel Dekker, Inc.

Fast, V.,Rinner, C. (2014). A systems perspective on volunteered geographic information, ISPRS *International Journal of Geo- Information*, 3(4), pp. 1278–1292.

Fazey , I. R. A. , Proust , K. , Newell , B. , Johnson , B. Fazey , J. A. (2006). Eliciting the implicit knowledge and perceptions of on-ground conservation managers of the Macquarie Marshes. *Ecology and Society*, 11(1), pp. 25 .

Field, C. B. (Ed.) (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change*. Cambridge University Press.

Fiori, S., H. A. (2009). *Simon and Contemporary Theories of Bounded Rationality*. University of Torino, Department of Economics Research Paper No. 2/2008-CESMEP. Available at SSRN: http://ssrn.com/abstract=1367091 or http://dx.doi.org/10.2139/ssrn.1367091

Flanagin, A. J.,Metzger, M. J. (2008) The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), 137–148

Frappaolo, C., (2008) Implicit knowledge. *Knowledge Management Research & Practice*, 6(2008), 23–25.

Frege, G. (1923) Logische Untersuchungen. Drittel Teil: Gedankengefuge, «Beitrage zur Philosophie des deutschen Idealismus»; trad. *Le connessioni di pensieri. Ricerche logiche. Terza parte* in M. Di Francesco (a cura di), Ricerche logiche, Guerini, Milano, (1988), 99-125.

Freund, Y., Schapire, R., Abe, N. (1999) A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780), 1612.

Fricke, M., (2008). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, 35(2), pp.131–142. Available at: http://jis.sagepub.com/cgi/doi/10.1177/0165551508094050.

Friedmann, J. (1987). *Planning in the public domain: From knowledge to action*. Princeton University Press.

Frommberger, L., Schmid, F. (2013). *Crowdsourced bi-directional disaster reporting and alerting on smartphones in Lao PDR*. Retrieved from http://arxiv.org/abs/1312.6036

Gaillard, J. C., Mercer, J. (2013) From knowledge to action: Bridging gaps in disaster risk reduction. *Progress in Human Geography*, 37(1), 93–114. http://doi.org/10.1177/0309132512446717

Gaio, S., Borgo, S., Masolo, C., Oltramari, A., Guarino, N. (2010*). Un'introduzione all'ontologia DOLCE*. AIDAinformazioni, 1–2, 107–125.

GAO (Government Accountability Office). *Emergency Preparedness and Response – Some Issues and Challenges Associated with Major Emergency Incidents: Statement of William O. Jenkins, Jr., Director Homeland Security and Justice Issues; Testimony before the Little Hoover Commission, State of California*. Washington, DC: GAO (GAO-06-467T), 22 pages, February 23, 2006.

Gardner, J. (2002) Natural Hazards Risk in the Kullu District, Himachal Pradesh, India. *The Geographical Review*, 92(2): 282-306.

Gill, A.,Bunker, D. (2012) Crowd Sourcing Challenges Assessment Index for Disaster Management. In *AMCIS 2012 Proceedings*.

Global Network of Civil Society Organizations for Disaster Reduction (GNDR) (2011*). If we do not join hands: Views from the frontline – local reports of progress on implementing the Hyogo Framework for Action, with strategic recommendations for more effective implementation*. Teddington: Global Network of Civil Society Organizations for Disaster Reduction.

Godschalk, David R. (1991) Disaster Mitigation and Hazard Management. Pp. 131-160 in Thomas E. Drabek and Gerard J. Hoetmer (eds.), *Emergency Management: Principles and Practice for Local Government*, Washington, DC: International City Management Association.

Gomez, E. A., Elliot, N. (2013) Research article measuring mobile ICT literacy: Short-message performance assessment in emergency response settings. IEEE Transactions on Professional Communication, 56(1), 16–32. http://doi.org/10.1109/TPC.2012.2208394

Goodchild, M. F. (2007) Citizens as sensors: the world of volunteered geography. In *Geojournal*, 1–12.

Goodchild, M. F.,Glennon, J. A. (2010) Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3), 231–241.

Gorman, D., (2001). The Use and Abuse of Speech-act Theory in Criticism, *Poetics Today*, 20, 93–119.

Gould, S. J. J., Cox, A. L., Brumby, D. P., Wiseman, S. (2016) Short links and tiny keyboards: A systematic exploration of design trade-offs in link shortening services. *International Journal of Human Computer Studies*, 96, 38–53. http://doi.org/10.1016/j.ijhcs.2016.07.009

Gruber, T. R. (1993) Toward principles for the design of ontologies used for knowledge sharing. In N. Guarino, P. R. (Eds.), *Formal Ontology in Conceptual Analysis and Knowledge Representation (pp. 907–928)*. Padova - Italy. http://doi.org/10.1006/ijhc.1995.1081

Guarino N. (2008). *Ontologies and ontological analysis: an introduction* - FOIS 2008, Saarbrücken, October 31st, 2008 - Available in date in date 18/03/2016: http://fois08.dfki.de/presentations/FOIS2008-Tutorial1-Guarino.pdf

Guarino, N. (1998). Formal Ontology in Information Systems. *Formal Ontology in Information Systems. Proceedings (FOIS)*, Italy.

Guarino, N., Welty, C. a. (2009) An overview of OntoClean*. Handbook on Ontologies*, 1–20. http://doi.org/10.1007/978-3-540-92673-3

Guarino, N., Oberle, D., Staab, S. (2009). *What Is an Ontology*?, 1–17. http://doi.org/10.1007/978-3-540-92673-3

Guizzardi, G., Pires, L. F., Van Sinderen, M. (2005). An ontology-based approach for evaluating the domain appropriateness and comprehensibility appropriateness of modeling languages. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3713 LNCS, 691–705. http://doi.org/10.1007/11557432_51

Gulnerman, A. G., Karaman, H. (2015). PPGIS Case Studies Comparison and Future Questioning. Proceedings - 15th International Conference on Computational Science and Its Applications, ICCSA 2015, 104–107. http://doi.org/10.1109/ICCSA.2015.15

Gupta, A., Kumaraguru, P. (2012). Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media* (p. 2). ACM.

Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A. (2013). *Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy*.

Haklay, M. (2013) 'Citizen science and volunteered geographic information: overview and typology of participation',in Sui, D., Elwood, S.,Goodchild, M. (eds). *Crowdsourcing Geographic Knowledge*,., pp. 105–122, Springer, Dordrecht, Netherlands. http://doi.org/10.1007/978-94-007-4587-2_7

Hall, N.; Davis, I. (1999) *The Perception of Risk at Local Levels, and Ways to Measure Community Vulnerability*. London: Benfi eld Grieg Hazard Research Centre, University College London (UCL).

Hansen, J., Holm, L., Frewer, L., Robinson, P., Sandøe, P. (2003) Beyond the knowledge deficit: Recent research into lay and expert attitudes to food risks. *Appetite*, 41(2), 111–121. http://doi.org/10.1016/S0195-6663(03)00079-5

Hart-Davidson W., 'Content management: Beyond single source,' in R. Spikla (Ed). *Digital Literacy for Technical Communication: 21st Century Theory and Practice*,. New York: Routledge, 2010, pp. 128–143.

Harvey, F. (2013). To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information, In: Sui, D., Elwood, S.,Goodchild, M., pp. 31–41*Crowdsourcing Geographic Knowledge*, ed. by, Springer, Dordrecht, Netherlands. http://doi.org/10.1007/978-94-007-4587-2_3

Harwood, A. R., Lovett, A. A., Turner, J. A. (2015) Landscape and Urban Planning Customising virtual globe tours to enhance community awareness of local landscape benefits. *Landscape and Urban Planning*, 142, 106–119. http://doi.org/10.1016/j.landurbplan.2015.08.008

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Ullah Khan, S. (2015). The rise of 'big data' on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. http://doi.org/10.1016/j.is.2014.07.006

Haworth, B., Bruce, E. (2015) A Review of Volunteered Geographic Information for Disaster Management. *Geography Compass*, 9(5), 237–250. http://doi.org/10.1111/gec3.12213

Haykin, Simon, and Neural Network. 'A comprehensive foundation.' *Neural Networks* 2.2004 (2004).

Hermida, A., Thurman, N. (2008) a Clash of Cultures. *Journalism Practice*, 2(3), 343–356. http://doi.org/10.1080/17512780802054538

Hewitt, K. (1983) Interpretations of Calamity from the Viewpoint of Human Ecology, *The Risks and Hazards Series: 1*. New York: Allen and Unwin

Hiltz Starr Roxanne, Diaz Paloma, and Mark Gloria. 2011. Introduction: Social media and collaborative systems for crisis management. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 4 (2011)

Hiwasaki, L., Luna, E., Syamsidik, Shaw, R. (2014) Process for integrating local and indigenous knowledge with science for hydro-meteorological disaster risk reduction and climate change adaptation in coastal and small island communities*. International Journal of Disaster Risk Reduction*, 10, 15–27. http://doi.org/10.1016/j.ijdrr.2014.07.007

Hoppe, A., Seising, R., Nürnberger, A., Wenzel, C. (2011) Wisdom - the blurry top of human cognition in the DIKW-model *? Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2011) and LFA-2011*, (July), 584–591.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. F. G., Zipf, A. (2013) The use of Volunteered Geographic Information and Crowdsourcing. VGI in Disaster Management: Systematic Literature Review, 1–10.

Howden, K. (2011) Indigenous traditional knowledge and native title. *UNSW Law Journal*, 24(1), 60–84.

Hughes, A. L., St Denis, L. A., Palen, L., Anderson, K. M. (2014) Online public communications by police & fire services during the 2012 Hurricane Sandy. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems (pp. 1505-1514)*. ACM.

Hughes, A.L. Palen, L., (2009) Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(May), p.248.

Hui, L. H. D., Tsang, K. E. (2016). Everyday Knowledge and Disaster Management: The Role of Social Media, 30, 71–88. http://doi.org/10.1007/978-981-10-0216-8

Ianella, R. (2009) *Emergency Information Interoperability Frameworks*, (August), 1–25. Retrieved from https://www.w3.org/2005/Incubator/eiif/XGR-Framework-20090806/

Ifrcs. (2011) *Public awareness and public education for disaster risk reduction: A guide, 102*. Retrieved from http://www.ifrc.org/Global/Publications/disasters/reducing_risks/302200-Public-awareness-DDR-guide-EN.pdf Accessed 12 - September-2016)

Ikawa, Y., Enoki, M., Tatsubori, M. (2012) Location inference using microblog messages. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 687-690). ACM.

Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. Proc. of ISCRAM, Baden-Baden, Germany.

Imran, M., Castillo, C., Diaz, F., Vieweg, S. (2015). A Processing Social Media Messages in Mass Emergency: A Survey. ACM Computing Surveys (Vol. 47). http://doi.org/10.1145/2771588

International Federation of Red Cross and Red Crescent Societies (IFRC) (2011). *Desk Review on Trends in the Promotion of Community-Based Disaster Risk Reduction Through Legislation*. Geneva: International Federation of Red Cross and Red Crescent Societies

Iyengar, A., Finin, T., Joshi, A. (2011) Content-based prediction of temporal boundaries for events in Twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)* (pp. 186-191).

James Surowiecki, J. (2004): Wisdom of the crowds – Doubleday: Anchor

Jankowski, A.,Skowron, A. (2007) A wistech paradigm for intelligent system. In Peters, J. F. et al. (Eds.). Transactions on Rough Sets VI, LNCS 4374, 94–132, Springer-Verlag Berlin Heidelberg

Kamp S. (2016) Digital in 2016. Available in date 21/04/2016: http://wearesocial.com/it/blog/2016/01/report-digital-social-mobile-in-2016

Kar, B., Sieber, R., Haklay, M., Ghose, R. (2016). Public Participation GIS and Participatory GIS in the Era of GeoWeb Editorial Public Participation GIS and Participatory GIS in the Era of GeoWeb. *The Cartographic Journal*, 53(4), 296–299. http://doi.org/10.1080/00087041.2016.1256963

Kates, R.W. (1971). Natural Hazard in Human Ecological Perspective: Hypothesis and Models'. *Economic Geography*, 47(3): 438-451

Keßler, C., Hendrix, C. (2015) The humanitarian exchange language: coordinating disaster response with semantic web technologies. *Semantic Web*, 6(1), 5-21.

Kettle, N. P., Dow, K., Tuler, S., Webler, T., Whitehead, J., Miller, K. M. (2014) Integrating scientific and local knowledge to inform risk-based management approaches for climate adaptation. *Climate Risk Management*, 4, 17–31.

Kibriya, A. M., Frank, E., Pfahringer, B., Holmes, G. (2005) Multinomial naive bayes for text categorization revisited. In: Australasian Joint Conference on Artificial Intelligence (pp. 488-499). Springer Berlin Heidelberg.

Kirillov, I. a., Metcherin, S. a., Klimenko, S. V. (2012) Metamodel of shared situation awareness for resilience management of built environment. *Proceedings of the 2012 International Conference on Cyberworlds, Cyberworlds 2012, 137–143*. http://doi.org/10.1109/CW.2012.26

Kobourov, S. G. (2012). *Spring Embedders and Force Directed Graph Drawing Algorithms*, 1–23.

Kovács, G., Spens, K. M. (2007). Humanitarian logistics in disaster relief operations. *International Journal of Physical Distribution & Logistics Management*, 37(2), 99–114. http://doi.org/10.1108/09600030710734820

Kreps GA (1991) Organizing for emergency manage- ment. In Drabek TS and Hoetmer GJ (eds) *Emergency Management: Principles and Practice for Local Government*. Washington, DC: International City/County Management Association, 30–54

Kuhnert, P. M., Martin, T. G., Mengersen, K., Possingham, H. P. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: A Bayesian approach using expert opinion. *Environmetrics*, 16(7), 717–747. http://doi.org/10.1002/env.732

Kumar, S., Morstatter, F., Zafarani, R., Liu, H. (2013) Whom should i follow?: identifying relevant users during crises. In *Proceedings of the 24th ACM conference on hypertext and social media (pp. 139-147).* ACM.

Kumar, A.J. Chakrabarti, A., (2012). Bounded awareness and tacit knowledge: Revisiting Challenger disaster. *Journal of Knowledge Management*, 16(6), pp.934–949..

Laituri, M. Kodrich, K., (2008) On line disaster response community: People as sensors of high magnitude disasters using internet GIS. *Sensors*, 8(5), pp.3037–3055.

Laudon, J. P., Laudon, K. C. (2006). *Essentials of business information systems*. Prentice-Hall, Inc.

Leavitt Alex and Clark Joshua A. (2014). Upvoting Hurricane Sandy: Event-based News Production Processes on a Social News Site. In Proc. of CHI. ACM, 1495–1504.

Leszczynski, A. (2014). On the neo in neogeography. *Annals of the Association of American Geographers*. 104(1), pp. 60–79.

Lewis, D. D. (1998). *Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning (pp. 4-15)*. Springer: Berlin Heidelberg.

Liew, A. (2013) DIKIW: Data, information, knowledge, intelligence, wisdom and their interrelationships. *Business Management Dynamics*, 2(10), 49-62.

Lindell, M. K. (2013) Disaster studies. *Current Sociology*, 61(5–6), 797–825. http://doi.org/10.1177/0011392113484456

Liu, S., Shaw, D. Brewster, C., (2013). Ontologies for crisis management: a review of state of the art in ontology design and usability. ISCRAM 2013 - 10th International Conference on Information Systems for Crisis Response and Management, (May), pp.349–359.

Lu, X., Brelsford, C. (2014). Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami. *Scientific Reports*, 4, 6773. doi:10.1038/srep06773

Mackay, E. (2009). Indigenous traditional knowledge, copyright and art - Shortcomings in protection and an alternative approach. *The University of South Wales Law Journal*, 32(1), 1–26.

Madhu C. Reddy, Abraham, J., Deflitch, C., Yen, J. (2008). Challenges to Effective Crisis Management : Using Information and Communication Technologies to Coordinate Emergency Medical Services and Emergency Department Teams Challenges to effective crisis management : Using. *Journal of Medical Informatics*, (November 2008). http://doi.org/10.1016/j.ijmedinf.2008.08.003

Masolo, C, Borgo, S., Gangemi, S., Guarino, N, Oltramari, A., and Schneider, L., (2002) *The wonderweb library of foundational ontologies*. Deliverable 17, EU Won-derWeb Project. Available at: http://wonderweb.man.ac.uk/deliverables.shtml

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. (2003). *WonderWeb Deliverable D17 Preliminary Report. Communities*, (August 11, 2008), 38. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Won-derWeb+Deliverable+D17#0

McAdoo, B. G., Moore, A., Baumwoll, J. (2009). Indigenous knowledge and the near field population response during the 2007 Solomon Islands tsunami. *Natural Hazards*, 48(1), 73–82. http://doi.org/10.1007/s11069-008-9249-z

McCall, M. K., Martinez, J.,Verplanke, J. (2015). Shifting boundaries of volunteered geographic information systems and modalities: learning from PGIS, *ACME An International E-Journal for Critical Geographies*, 14(3), pp. 791–826.

Mele, F., Sorgente, A. (2011). The temporal representation and reasoning of complex events. In CILC (pp. 385-399).

Mercer I. Kelman, L. Taranis et S. Suchet-Pearson, J. (2010). Framework for integrating indigenous and scientific knowledge for disaster risk reduction. *Disasters*, 34(1), 214–239. http://doi.org/10.1111/j.1467-7717.2009.01126.x

Metaxas, P., Mustafaraj, E. (2013). The rise and the fall of a citizen reporter. In Proceedings of the 5th Annual ACM Web Science Conference (pp. 248-257). ACM.

Moens, M. F., Boiy, E., Palau, R. M., Reed, C. (2007). Automatic detection of arguments in legal texts. In *Proceedings of the 11th international conference on Artificial intelligence and law (pp. 225-230)*. ACM.

Moens, M. F. (2013) Argumentation Mining: Where are we now, where do we want to be and how do we get there? In *Post-Proceedings of the 4th and 5th Workshops of the Forum for Information Retrieval Evaluation* (p. 2). ACM.

Murgante, B., Scardaccione, G., Las Casas, G. (2009) Building ontologies for disaster management: seismic risk domain. *Urban and Regional Data Management*, 259-269.

Natalya F. Noy. (2004) Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record* 33, 4 (2004), 65–70.

Neubig, G., Matsubayashi, Y., Hagiwara, M., Murakami, K. (2011). Safety Information Mining-What can NLP do in a disaster-. In *IJCNLP* (Vol. 11, pp. 965-973).

Newport, J. K., Jawahar, G. G. P. (2003) Community participation and public awareness in disaster mitigation. *Disaster Prevention and Management*, 12(1), 33–36. http://doi.org/10.1108/09653560310463838

Newport, J. K., Jawahar, G. G. P. (2003) Community participation and public awareness in disaster mitigation. *Disaster Prevention and Management*, 12(1), 33–36. http://doi.org/10.1108/09653560310463838

Nonaka, I., Takeuchi, H. (1995) The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford university press.

Nonaka, I Konno, N and Toyama, R (2001) Emergence of Ba. In Nonaka, I and Nishiguchi, T (Eds), *Knowledge emergence: Social, technical and evolutionary dimensions of knowledge creation* . Oxford: Oxford University press.

Nuzzaci, A. (2011). *Pratiche riflessive, riflessività e insegnamento*. Studi E Ricerche, 3.

O'Reilly, T. (2007) What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 1(First Quarter), 17. http://doi.org/10.2139/ssrn.1008839

Obermeyer, N. J. (1998). The evolution of public participation GIS. *Cartography and Geographic Information Science*, 25(2), pp. 65–66

Obrst, L., (2010). Ontological Architectures. In Roberto Poli, Michael Healy, and Achilles Kameas (eds). *Theory and Applications of Ontology: Computer Applications*, chapter 1, pages 27–66. Springer, 2010.

Office of the United Nations Disaster Relief Coordinator, (1979). *Expert Group Meeting on Vulnerability Analysis* (Geneva, Switzerland).

Ogasawara, N., Ginsburg, J. (2015). *Linguistic Analysis of Japanese Tsunami Evacuation Calls*.

Okolloh, O. (2009). Ushahidi or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action*, 59(1), 65–70.

Olaide, I. A., Omolere, O. W. (2013) Management of Indigenous Knowledge as a Catalyst towards Improved Information Accessibility to Local Communities: A Literature Review. Chinese Librarianship*: An International Electronic Journal*, 35(2005), 87–98. Retrieved from www.iclc.us/cliej/cl35IO.pdf

Olsson, P., Folke, C. (2001) Local ecological knowledge and institutional dynamics for ecosystem management: A study of Lake Racken watershed, Sweden. *Ecosystems*, 4(2), 85–104. http://doi.org/10.1007/s100210000061

Oltramari, A., Borgo, S., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N., Pisanelli, D. (2003). Il ruolo dell'ontologia nella disambiguazione del significato. *Networks*, (2003/2), 14–24.

Øhrstrøm, P., Andersen, J., Schärfe, H. (2005). *What has happened to ontology. In International Conference on Conceptual Structures* (pp. 425-438). Springer Berlin Heidelberg.

Pasquini, M., Alexander, M. (2005). Soil fertility management strategies on the Jos Plateau: the need for integrating 'empirical'and 'scientific' knowledge in agricultural development. *The Geographical Journal*, 171(2), 112–124.

Passant, A., Laublet, P. (2008). Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In *LDOW*.

Peldszus, A., Stede, M. (2013). From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence* (IJCINI), 7(1), 1-31.

Peng, F., Schuurmans, D., Wang, S. (2004) Augmenting naive bayes classifiers with statistical language models. *Information Retrieval*, 7(3-4), 317-345.

Perry, R.W.,Quarantelli, E.L. (Eds) (2005). What is a Disaster? New Answers to Old Questions , Xlibris, Philadelphia.

Petak William J. (1985). Emergency management: A challenge for public administration. *Public Administration Review*, 3–7.

Peters-Guarin, G., Mccall, M. K., van Westen, C. (2012). Coping strategies and risk manageability: Using participatory geographical information systems to represent local knowledge. *Disasters*, 36(1), 1–27. http://doi.org/10.1111/j.1467-7717.2011.01247.x

Picketts, I.M., Curry, J., Rapaport, E., (2012). Community adaptation to climate change: environmental planners' knowledge and experiences in British Columbia, Canada. *J. Environ. Planning Policy Manage*. 14 (2), 119–137

Polanyi, M. (1962). *Personal knowledge: towards a post-critical philosophy*.

Poli, R., Obrst, L., (2010). The Interplay Between Ontology as Categorial Analysis and Ontology as Technology In Roberto Poli, Michael Healy, and Achilles Kameas, editors, *Theory and Applications of Ontology: Computer Applications*, chapter 1, pages 1–26. Springer, 2010.

Pollock, M. L., Legg, C. J., Holland, J. P., Theobald, C. M. (2007). Assessment of expert opinion: seasonal sheep preference and plant response to grazing. *Rangeland Ecology and Management*, 60(December 2006), 125–135. http://doi.org/10.2111/06-032R2.1

Poole, H. W., Lambert, L., Woodford, C., Moschovitis, C. J. (Eds.) (2005). *The Internet: A historical encyclopaedia* (Vol. 1). Abc-Clio Inc.

Poser, K., Dransch, D. (2010) Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica*, 64(1), 89–98.

Prince, S. H. (1920). *Catastrophe and social change, based upon a sociological study of the Halifax disaster* (No. 212-214). Columbia university.

Qu, Y., Huang, C., Zhang, P., Zhang, J. (2011). Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 25-34). ACM.

Quarantelli, E. L. (2006). Catastrophes are different from disasters: some implications for crisis planning and managing drawn from Katrina. In: *The Social Science Research Council website: Understanding Katrina: Perspectives from the Social Sciences* . Available at http://understandingkatrina.ssrc.org/Quarantelli/. Accessed May 1, 2007.

Raskin, R. G., Pan, M. J. (2005). Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). *Computers & geosciences*, 31(9), 1119-1125.

Rambaldi, G., Chambers, R., Mccall, M.,Fox, J. (2006). Practical ethics for PGIS practitioners, facilitators, technology intermediaries and researchers. In Corbett, J., Rambaldi, G., Kyem, P., Weiner, D., Olson, R., Muchemi, J., McCall, M. K.,Chambers,Participatory Learning and Action 54, pp. 106–113, IIED, London

Raymond, C. M., Fazey, I., Reed, M. S., Stringer, L. C., Robinson, G. M., Evely, A. C. (2010). Integrating local and scientific knowledge for environmental management. *Journal of environmental management*, 91(8), 1766-1777.

Rennie, J. D., Shih, L., Teevan, J., Karger, D. R. (2003) Tackling the poor assumptions of naive bayes text classifiers. In *ICML* (Vol. 3, pp. 616-623).

Rish, I. (2001). *An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46). IBM New York.

Ritter, A., Clark, S., Etzioni, O. (2011). Named entity recognition in tweets: An experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1524-1534). Association for Computational Linguistics.

Robertson, H. A., McGee, T. K. (2003). Applying local knowledge: The contribution of oral history to wetland rehabilitation at Kanyapella Basin, Australia. *Journal of Environmental Management*, 69(3), 275–287.

Rowley, J. (2007) The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180. http://doi.org/10.1177/0165551506070706

Rudra, K., Ghosh, S., Ganguly, N., Goyal, P., Ghosh, S. (2015). Extracting Situational Information from Microblogs during Disaster Events: A Classification-Summarization Approach. *Communication Systems and Networks (COMSNETS), 2015 7th International Conference* http://doi.org/10.1109/COMSNETS.2015.7098720

Sadock, J. (2004) Speech Acts. *Handbook of Pragmatics*, (1962), 53–73.

Sakaki, T., Okazaki, M., Matsuo, Y. (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web (pp. 851-860)*. ACM.

Sbisà, M. (2011). *Teoria degli atti linguistici: appunti per una storia*. Università di Trieste.

Scharl, A., Tochtermann, K. (2007). *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. London, Springer-Verlag

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Schön, Donald. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Scorza, F., Casas, G. Las, Murgante, B. (2010). *Overcoming Interoperability Weaknesses in e-Government Processes: Organizing and Sharing Ontologies*. World Summit on the Knowledge Society, 112, 243–253

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), 1. http://doi.org/10.1017/S0047404500006837

Searle, J. R., Vanderveken, D. (1985) Speech acts and illocutionary logic. In *Logic, thought and action (pp. 109-132).* Springer Netherlands.

Senge, P.M. (1990), The Fifth Discipline; the Art and Practice of the Learning Organi-zation, Doubleday, NewYork.

Shaluf, I. M., Ahmadun, F., Said, A. M. (2003). A review of disaster and crisis. *Disaster Prevention and Management*, 12(1), 24–32. http://doi.org/10.1108/09653560310463829

Sheydaei, N., Saraee, M., Shahgholian, A. (2015) A novel feature selection method for text classification using association rules and clustering. *Journal of Information Science*, 41(1), 3–15. http://doi.org/10.1177/0165551514550143

Shklovski, I., Burke, M., Kiesler, S., Kraut, R. (2010) Technology adoption and use in the aftermath of Hurricane Katrina in New Orleans. *American Behavioral Scientist*, 53(8), 1228-1246.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, 24 (9), 467–471.

Sieber, R. (2006) Public participation geographic information systems: a literature review and framework. *Annals of the Association of American Geographers*,96 (3), 491–507.

Simon, H.A. (1972). Theories of bounded rationality. In: McGuire, C.B., Radner, R. (Eds.), *Decision and Organization: A Volume in Honor of Jacob Marschak*. North Holland Publishing Company, Amsterdam, pp. 161–176.

Simon, T., Goldberg, A. Adini, B. (2015). Socializing in emergencies—A review of the use of social media in emergency situations*. International Journal of Information Management*, 35(5), pp.609–619. Available at: http://www.sciencedirect.com/science/artle/pii/S0268401215000638.

Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., Adini, B. (2014). Twitter in the cross fire--the use of social media in the Westgate Mall terror attack in Kenya. *PloS One*, 9(8), e104136. http://doi.org/10.1371/journal.pone.0104136

Slinger, J., Muller, M., Hendriks, M. (2007) Exploring local knowledge of the flooding risk of the Scheldt Estuary. *Water Science and Technology*, 56(4), 79–86. http://doi.org/10.2166/wst.2007.539

Smith, E. a. (2001) The role of tacit and explicit knowledge in the workplace. *Journal of Knowledge Management*, 5(4), pp.311–321.

Smith, E., Wasiak, J., Sen, A., Archer, F., Burkle Jr., F. M. (2009). Three decades of disasters: a review of disaster-specific literature from 1977-2009. *Prehosp Disaster Med*, 24(4), 306–311. http://doi.org/10.1017/S1049023X00007020

Smith, T. A. (2011). Local knowledge in development (geography). *Geography Compass*, 5(8), 595–609. http://doi.org/10.1111/j.1749-8198.2011.00443.x

Smith, Van Den Eeckhaut, M., Hervás, J., Montanarella, L (2013). Landslide Databases in Europe: Analysis and Recommendations for Interoperability and Harmonisation. In Margottini, C., Canuti, P., Sassa, K., *Landslide Science and Practice*, pp 35-42.

Sprake, J., Rogers, P. (2014) Crowds, citizens and sensors: Process and practice for mobilising learning. *Personal and Ubiquitous Computing*, 18, 753–764. http://doi.org/10.1007/s00779-013-0715-6

Starbird, K., Stamberger, J.: (2010) Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In*: ISCRAM*.

Starbird, K., Palen, L.: Voluntweeters. (2011). *Self-organizing by digital volunteers in times of crisis*.

Starbird, K., Muzny, G., Palen, L. (2012) Learning from the crowd: collaborative filtering techniques for identifying on-the-ground Twitterers during mass disruptions. In *Proceedings of 9th International Conference on Information Systems for Crisis Response and Management, ISCRAM*.

Stenmark, D. (2002) Information vs. knowledge: The role of intranets in knowledge management. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2002–January(c), 928–937. http://doi.org/10.1109/HICSS.2002.994043

Studer, R., Benjamins, V. R., Fensel, D. (1998) Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1–2), 161–197. http://doi.org/10.1016/S0169-023X(97)00056-6

Sui, X., Chen, Z., Wu, K., Ren, P., Ma, J., Zhou, F. (2014) Social media as sensor in real world: Geolocate user with microblog. In *Natural Language Processing and Chinese Computing* (pp. 229-237). Springer Berlin Heidelberg.

Suykens, J. A., Vandewalle, J. (1999) Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300.

Swanson, D. R. (1991) Complementary structures in disjoint science literatures. In *Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 280-289)*. ACM. (pp. 280–289).

Tabler Jr., R. E. (2008). The social construction of a special needs program for hurricanes. ProQuest Dissertations and Theses, 469.

Tickner, J. A. (2015). Revisiting IR in a Time of Crisis: Learning from Indigenous knowledge. *International Feminist Journal of Politics*, 17(4), 536–553. http://doi.org/10.1080/14616742.2015.1071995

Tobin, G. A. (1997). *Natural hazards: explanation and integration*. Guilford Press.

Tronti L., (2014). *Economia della conoscenza, apprendimento e democrazia 'FOR – Tendenze Strumenti Strategie.* Franco Angeli, Milano, n. 98, 2014, pp. 12-17

Tuomi, I. (2000) Data Is More Than Knowledge: Of the Reversed Knowledge Implications for Knowledge Management Hierarchy and Organizational Memory. *Journal of Management Information Systems*, 16(3), 103–117.

Turban E., Rainer R.K. Pot,ter R.E. (2005). *Introduction to Information Technology* (Wiley, New York).

UK-NEA. (2011). *The UK National Ecosystem Assessment technical report*. Cambridge: UNEP-WCMC.

UNESCO. (2003). *Local and Indigenous Knowledge Systems. What is local knowledge?.* <http://portal.unesco.org/science/en/ev.php-URL_ID=2034&URL_DO=DO_TOPIC&URL_SECTION=201.html>(Accessed 23 - september - 2016)

UNISDR. (2009). *Terminology on Disaster Risk Reduction.* United Nations International Strategy for Disaster Risk Reduction, Geneva,

Van Camp, B. (2007) Keynote Address. L. Boon, E. K., Hens (Eds). *Indigenous Knowledge Systems and Sustainable Development: Relevance for Africa* (Vol. 1). Delhi, India: Kamla-Raj Enterprises.

Van Dijck, J., Poell, T. (2013). Understanding Social Media Logic. *Media and Communication*, 1(1), 2-14Uni. Press.

Van Dijk, T. A. (1977) Context and cognition: Knowledge frames and speech act comprehension. *Journal of Pragmatics*, 1(3), 211–231. http://doi.org/10.1016/0378-2166(77)90035-2

Van Herzele, A. (2004). Local knowledge in action - Valuing nonprofessional reasoning in the planning process. *Journal of Planning Education and Research*, 24(2), 197–212. http://doi.org/10.1177/0739456x04267723

Varnes, D. J.,IAEG. (1984). *Landslide Hazard Zonation – A Review of principles and practice. Commission on Landslides*. Paris, UNESCO pp. 60.

Verma, T. (2014). Tokenization and Filtering Process in RapidMiner. *International Journal of Applied Information Systems* (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA, 7(2), 16–18.

Verplanke, J., McCall, M. K., Uberhuaga, C., Rambaldi, G., Haklay, M. (2016). A Shared Perspective for PGIS and VGI. *The Cartographic Journal*, 7041(September), 1–10. http://doi.org/10.1080/00087041.2016.1227552

Vieweg, S., Hughes, A. L., Starbird, K., Palen, L. (2010) Microblogging during two natural hazards events: what twitter may contribute to situational awareness. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1079-1088). ACM.

Vosoughi, S. Roy, D. (2016) .Tweet acts: A speech act classifier for twitter. arXiv preprint arXiv:1605.05156.

Wang, H., Hovy, E., Dredze, M. (2015) The Hurricane Sandy Twitter Corpus. *AAAI Workshop on the World Wide Web and Public Health Intelligence*, 20–24.

Wang, W.J., Dong, C.X., and Yang, P. (2009). Ontology modeling of emergency plan systems. In: *Proceeding of 6th international conference on fuzzy systems and knowledge discovery*, FSKD 2009, Vol. 2, IEEE Press Piscataway, NJ, USA, 290–294.

Wang, X., Swathi, M., Atit, T., Aisling Kelliher, Hari, S. (2007). *Eventory – An Event Based Media Repository*.

Wang, W., Meng, F., Wang, Y., Luo, Y., & Xu, Z. (2006). Research on ontology-based emergency response plan template. *Jisuanji Gongcheng/ Computer Engineering*, 32(19), 170-172.

Weichselgartner, J., Pigeon, P. (2015). The Role of Knowledge in Disaster Risk Reduction. *International Journal of Disaster Risk Science*, 6(2), 107–116. http://doi.org/10.1007/s13753-015-0052-7

Wen, L., (2008). The study of Ontology for knowledge model based on emergency decision-making. Thesis (MS). DaLian University of Technology.

Wendling, C., Radisch, J., Jacobzone, S. (2013). The Use of Social Media in Risk and Crisis Communication. *OECD Working Papers on Public Governance* No. 24, (24), 1–42. http://doi.org/10.1787/5k3v01fskp9s-en

Whittaker, J., McLennan, B. Handmer, J., (2015) A review of informal volunteerism in emergencies and disasters: Definition, opportunities and challenges. International *Journal of Disaster Risk Reduction*, 13, pp.358–368. Available at: http://www.sciencedirect.com/science/article/pii/S2212420915300388.

Wierzbicka, A. (1987). *English speech act verbs: A semantic dictionary*. Academic Press Sydney.

World Health Organization King's College London (2011). *The Humanitarian Emergency Settings Perceived Needs Scale (HESPER): Manual with Scale*. Geneva: World Health Organization.

Xiaohua Liu, Shaodian Zhang, Furu Wei, and Ming Zhou. (2011). Recognizing named entities in tweets. In Proc. of ACL. *Association for Computational Linguistics*, 359–367

Xu Jinghai, Nyerges Timothy L. Nie Gaozhong (2014). Modeling and representation for earthquake emergency response knowledge: perspective for working with geo-ontology, *International Journal of Geographical Information Science*, 28:1, 185-205, DOI: 10.1080/13658816.2013.845893

Xu, J.,Nyerges, T.L. (2016). A framework for user-generated geographic content acquisition in an age of crowdsourcing. *Cartography and Geographic Information Science*. http://dx. doi.org/10.1080/15230406.2016.1163644

Xu, J., Nyerges, T. L. (2016). A framework for user-generated geographic content acquisition in an age of crowdsourcing. *Cartography and Geographic Information Science*, 406(August), 1–15. http://doi.org/10.1080/15230406.2016.1163644

Yang, Y., Liu, X. (1999). A re-examination of text categorization methods. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 42-49). ACM.

Yang, H.-C., Lee, C.-H. (2005). A text mining approach for automatic construction of hypertexts. *Expert Systems with Applications*, 29(4), 723–734. http://doi.org/10.1016/j.eswa.2005.05.003

Ye, J., Coyle, L., Dobson, S., Nixon, P. (2007) Ontology-based models in pervasive computing systems. *The Knowledge Engineering Review*, 22(4), 315–347. http://doi.org/10.1017/S0269888907001208

Zanasi A. (ed.) (2005), *Text mining and its applications to intelligence, CRM and knowledge management*. WIT Press: Southampton

Zeichner K.M., Liston D. (1990). *Traditions of reform and reflective teaching in US teacher education*. Michigan: National Centre for Research in Teacher Education, Michigan State University

Zeleny, M. (2005). Knowledge-information autopoietic cycle: towards the wisdom systems. *International Journal of Management and Decision Making*, 7(1), 3-18.

Zhang, R., Li, W., Gao, D., Ouyang, Y. (2013) Automatic twitter topic summarization with speech acts. *IEEE Transactions on Audio, Speech and Language Processing*, 21(3), 649–658. http://doi.org/10.1109/TASL.2012.2229984

Zhu, H. H., Madnick, S. (2006). *A lightweight ontology approach to scalable interoperability*. MIT Sloan Working Paper 4621-06. Working Paper CISL – Composite Information Systems Laboratory June 2006

Ziemke, J. (2012) Crisis Mapping: The Construction of a New Interdisciplinary Field? *Journal of Map & Geography Libraries*, 8(2), 101–117

Zins, C., (2007). Conceptual Approaches for Defining Data, Information, and Knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), pp.479–493.

Resources

K. Heppler. *Helping Haitians via mobile, crowdsourcing & social media*. Social Brite. [Online]. Available: http://www.socialbrite.org/2010/02/16/helping-haitians-via-mobile-crowdsourcing-social-media/ in date 21/04/2016

PASQUALE BALENA
CURRICULUM VITAE – SHORT

RECENT WORKPLACE INFO

Nome	BALENA, Pasquale
Workplace	DICATECh - Technical University of Bari, Italy, via Orabona, 4 – 70125 Bari (Italy)
Telephone	(+ 39) 3204316342
E-mail	pasquale.balena@poliba.it
Nationality	Italiana

QUALIFICATIONS

2014
DICATECh - Technical University of Bari
Ph.D. Candidate in Risk and environmental, territorial and building development - XXIX Cycle

2012
IUAV UNIVERSITY OF VENICE – Department of Design and Planning In Complex Environments
Master's Degree in Geographic Information Systems and Remote Sensing
Final Grades 110/110

2010
IUAV UNIVERSITY OF VENICE – Department of Design and Planning in Complex Environments
Post Graduate Diploma in Geographic Information Systems and Remote Sensing

2008
IUAV UNIVERSITY OF VENICE – Department of Design and Planning In Complex Environments
Bachelor's Degree in Geographic Information Systems
Final Grades 108/110

WORK EXPERIENCE

1988 - 2000

CONSULTANCY FIRMS in URBAN PLANNING AND GEOGRAPHIC INFORMATION SYSTEM
Analyst Programmer and Project Manager in Geographic information systems
Lead Auditor Quality (UN IEN ISO 9000)

2001 – Present

Director of the Laboratory Urban Planning Laboratory – Technical University of Bari

Main Publications

De Lucia, C., Balena, P., Stufano Melone, M. R., Borri, D. (2016). Policy, entrepreneurship, creativity and sustainability: The case of ‘Principi Attivi’ (‘Active Ingredients’) in Apulia Region (southern Italy). Journal of Cleaner Production, 135, 1461-1473. Special issue: ‘Paths for integrating creativity and sustainability’.

Bonifazi, P. Balena, (2016). I suoli di Puglia fra consumo e politiche per il risparmio. In Nuovo sfide per il suolo – Centro di Ricerca sui Consumi di Suolo, Rapporto 2016, Chapter: 2.10, Publisher: INU Edizioni, Editors: Arcidiacono, Andrea, Damiano Di Simine, Federico Oliva, Silvia Ronchi, Stefano Salata, pp.99-104.

A. Bonifazi, P, Balena (2015). Risparmio di suolo nei paesaggi industriali: quali direzioni? in Michele Munafò , Marco Marchetti - Recuperiamo terreno. Analisi e prospettive per la gestione sostenibile della risorsa suolo, Franco Angeli Editore.

P. Balena, V. Sannicandro , C. Torre (2014) Spatial Multicrierial Evaluation of Soil Consumption as a Tool for SEA In ICCSA 2014, Part III, LNCS 8581, pp. 446–458, 2014. Springer, Heidelber.

Torre , Balena P., Ceppi C. (2014). The devaluation of property due to the perception of health risk in polluted industrial cities In Int. Journal Business Intelligence and Data Mining, Vol. 9, No. 1.

C. Torre, P. Balena, C. Ceppi, (2013). The Estimation of the Depreciation of Property Due to Environmental Risk: The Case of Taranto in ICCSA 2013, Part IV, LNCS 7974, pp. 528-540, Springer, Heidelberg.

C. Torre, P. Balena, V. Sannicandro (2013) Spatial Analysis of Soil Consumption and as Support to Transfer Development Rights Mechanisms in ICCSA, Part IV, LNCS 7974, pp. 587-599, Springer, Heidelberg.

P. Balena, A.Bonifazi, G. Mangialardi (2013) Smart Communities Meet Urban Management:Harnessing the Potential of Open Data and Public/Private Partnerships through Innovative E-Governance Applications in ICCSA 2013, Part IV, LNCS 7974, pp. 587-599, Springer, Heidelber.

P.Balena, G. Mangialardi (2012)
Politiche innovative per la rigenerazione urbana in chiave sostenibile (in F. Rotondo, F.Selicato, C. Torre – Percorsi di Rigenerazione Urbana e Territoriale, Mario Adda Editore, Bari).

C. Torre, Pasquale Balena, V. Sannicandro (2012). Il Contenimento del Consumo di Suolo come elemento della rigenerazione territoriale e urbana (in F. Rotondo, F.Selicato, C. Torre – Percorsi di Rigenerazione Urbana e Territoriale, Mario Adda Editore, Bari).

P. Balena, G. Mangialardi and C. Maria Torre. (2012). A BEP Analysis of Energy Supply For Sustainable Urban Microgrids, Springer, Heidelberg.

C. Torre, P. Balena, R. Zito (2012). An automatic procedure to select areas for transfer development rights in the urban market, Springer, Heidelberg.

C. Torre, P. Balena, G., Mangialardi (2012). Strumenti interdisciplinari per la valutazione, la programmazione e la gestione delle risorse ambientali Break Even Point per l’autoapprovvigionamento energetico di aree urbane omogenee. Rivista Territori 9 Anno III – Giugno.