



Interdisciplinary Journal of Information, Knowledge, and Management

An Official Publication
of the Informing Science Institute
InformingScience.org

IJIKM.org

Volume 14, 2019

CRISIS AND DISASTER SITUATIONS ON SOCIAL MEDIA STREAMS: AN ONTOLOGY-BASED KNOWLEDGE HARVESTING APPROACH

SenthilKumar Narayanasamy*	Vellore Institute of Technology, Vellore, Tamil Nadu, India	senthilkumar.n@vit.ac.in
Dinakaran Muruganantham	Vellore Institute of Technology, Vellore, Tamil Nadu, India	dinakaran.m@vit.ac.in
Atilla Elçi	Aksaray University, Merkez, Aksaray, Turkey	atilla.elci@gmail.com

* Corresponding author

ABSTRACT

Aim/Purpose	Vis-à-vis management of crisis and disaster situations, this paper focuses on important use cases of social media functions, such as information collection & dissemination, disaster event identification & monitoring, collaborative problem-solving mechanism, and decision-making process. With the prolific utilization of disaster-based ontological framework, a strong disambiguation system is realized, which further enhances the searching capabilities of the user request and provides a solution of unambiguous in nature.
Background	Even though social media is information-rich, it has created a challenge for deriving a decision in critical crisis-related cases. In order to make the whole process effective and avail quality decision making, sufficiently clear semantics of such information is necessary, which can be supplemented through employing semantic web technologies.
Methodology	This paper evolves a disaster ontology-based system availing a framework model for monitoring uses of social media during risk and crisis-related events. The proposed system monitors a discussion thread discovering whether it has reached its peak or decline after its root in the social forum like Twitter. The content in social media can be accessed through two typical ways: Search Application Program Interfaces (APIs) and Streaming APIs. These two kinds of API processes can be used interchangeably. News content may be filtered by time, geographical region, keyword occurrence and

Accepting Editor Iris A Humala | Received: June 23, 2019 | Revised: August 19, August 21, 2019 |
Accepted: August 22, 2019.

Cite as: Narayanasamy, S., Muruganantham. D. & Elçi, A. (2019). Crisis and disaster situations on social media streams: An ontology-based knowledge harvesting approach. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, 343-366. <https://doi.org/10.28945/4420>

(CC BY-NC 4.0) This article is licensed to you under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

availability ratio. With the support of disaster ontology, domain knowledge extraction and comparison against all possible concepts are availed. Besides, the proposed method makes use of SPARQL to disambiguate the query and yield the results which produce high precision.

Contribution	The model provides for the collection of crisis-related temporal data and decision making through semantic mapping of entities over concepts in a disaster ontology we developed, thereby disambiguating potential named entities. Results of empirical testing and analysis indicate that the proposed model outperforms similar other models.
Findings	Crucial findings of this research lie in three aspects: (1) Twitter streams and conventional news media tend to offer almost similar types of news coverage for a specified event, but the rate of distribution among topics/categories differs. (2) On specific events such as disaster, crisis or any emergency situations, the volume of information that has been accumulated between the two news media stands divergent and filtering the most potential information poses a challenging task. (3) Relational mapping/co-occurrence of terms has been well designed for conventional news media, but due to shortness and sparseness of tweets, there remains a bottleneck for researchers.
Recommendations for Practitioners	Though metadata avails collaborative details of news content and it has been conventionally used in many areas like information retrieval, natural language processing, and pattern recognition, there is still a lack of fulfillment in semantic aspects of data. Hence, the pervasive use of ontology is highly suggested that build semantic-oriented metadata for concept-based modeling, information flow searching and knowledge exchange.
Recommendation for Researchers	The strong recommendation for researchers is that instead of heavily relying on conventional Information Retrieval (IR) systems, one can focus more on ontology for improving the accuracy rate and thereby reducing ambiguous terms persisting in the result sets. In order to harness the potential information to derive the hidden facts, this research recommends clustering the information from diverse sources rather than pruning a single news source. It is advisable to use a domain ontology to segregate the entities which pose ambiguity over other candidate sets thus strengthening the outcome.
Impact on Society	The objective of this research is to provide informative summarization of happenings such as crisis, disaster, emergency and havoc-based situations in the real world. A system is proposed which provides the summarized views of such happenings and corroborates the news by interrelating with one another. Its major task is to monitor the events which are very booming and deemed important from a crowd's perspective.
Future Research	In the future, one shall strive to help to summarize and to visualize the potential information which is ranked high by the model.
Keywords	disaster management, social media, ontological support, semantic search, SPARQL, RDF

INTRODUCTION

For long, there has been a huge demand to develop an efficient mechanism to effectively search and extract much-needed information from the social web. Manual annotation is effectively possible in information retrieval for a limited number of documents, but impractical for a large accumulation of content, particularly in social media. And moreover, automatic annotation processes are in an infant

stage. As Ritter, Etzioni and Clark (2012) indicated, automatic annotation has not reached its complete stage. There would be deemed requirements to properly utilize ontologies to precisely govern the types of knowledge to harvest. With the support of ontology (Sakaki, Okazaki, & Matsuo, 2010), domain knowledge extraction is very relevant and relates all possible concepts. Ontology-based knowledge extraction is expected to provide a boost to the domain of this research, disaster management.

Hence, pervasive use of ontology has been highly suggested (Ha-Thuc, Mejova, Harris, & Srinivasan, 2010; Luo, Osborne, & Wang, 2012) that builds semantic oriented metadata for concept-based modeling, information flow searching and knowledge exchanging. Once the semantic aspect of metadata is built for the content available over the Web on a domain of interest, then it will provide common grounds for understanding and sharing the information, as well as increasing relevancy and reducing, perhaps even minimizing inherent ambiguities. In the past, there were projects aimed at scavenging Web content with exemplary results for semantically annotating the metadata for domain-specific semantic searches. Several noteworthy examples may be mentioned; for instance, systems like PlanetOnto, ArtEquAKT, sparse kernel learning continuous relevance model for image annotation (Moran & Lavrenko, 2014), and integration of linked data in Knowledge-Based Systems for process planning (Rehage, Joppen, & Gausemeier, 2016). However, when it comes to working with the fast transient contents of social networks, a different paradigm is needed where this research comes in to contribute.

The next potential task of the operation is accessing the news contents from social media platforms. Though the task seems simple, it is inherently complicated due to interoperability conundrum and accessibility disabilities. Most of the social media platforms have accorded privileges to the programmers accessing the content through its appropriate APIs, but they differ from one another in what they provide and are mostly resource-limited (i.e., in the sense of the number of allowed request for a unit time). The content in the social media can be accessed through two typical ways: permitting users to access and archive the past messages, it is called Search APIs; and allowing users to subscribe for the real-time data feeds, it is known as Streaming APIs. These two kinds of API processes can be used interchangeably and allow expressing information needs, such as filtering the news content by time, geographical region, keyword occurrence and availability ratio (Celik, Abel, & Houben, 2011; Raman, Kuppusamy, Dorasamy, & Nair, 2014). The harvested data however still requires further pre-processing.

DATA PRE-PROCESSING AND NORMALIZATION

Though we have effective information extraction processes and well-established APIs to gather the news source content, there still is a pertinent need for preprocessing the extracted data. The news content extracted from social media can be pre-processed by natural language processing (NLP) toolkit. Common pre-processing operations are tokenization and labeling, part-of-speech tagging, semantic role labeling, dependency parsing and named entity linking (Kumar & Muruganantham, 2016; Lima, Espinasse, & Freitas, 2017). The next challenge on data pre-processing is to reduce the amount of data identifying and eliminating duplicate messages. It is not an easy task since every message posted in the social media can be valuable; de-duplicating the messages requires thorough clustering and then prioritizing them based on the event context. In order to prune this whole process, semantic-based technologies are used.

Apart from that, there are various issues associated with handling social media messages. The most prominent issues are scalability and content. The scalability issue concerns Twitter stream size, volume, and velocity. Particularly during any large crisis or severe havoc (Otegi, Arregi, Ansa, & Agirre, 2015; Sakaki et al., 2010), a huge volume of tweets and millions of messages pertaining to that event may be posted. In these critical situations, the tweet velocity would never be at a constant rate. Instead, it grows drastically and records a huge response from people over the event. If it is observed at various times, it would be discovered that same/similar tweets were repeated and reposted again

and again for the same event. This is the foremost challenge for the scalability issue; redundancy avoidance is the core factor for decision making and enhances the level of understanding over the specified event. Next, the content issue deals with tweets that are very brief and canonical (Abel, Gao, Houben, & Tao, 2011); most of the tweets posted in the social media are akin to normal speech and they pose a seminal challenge for the computational methods to deliver the correct form.

OBJECTIVES OF THIS RESEARCH

The objective of this research is to provide informative summarization of social network content concerning happenings such as crisis, disaster, emergency and havoc based situations in the real world. A system is proposed that provides the summarized views of such happenings and corroborates the news by interrelating with one another. Its major task is to monitor the events which are very booming and deemed important from a crowd's perspective (Samuel & Sharma 2018; Sheth, Thomas, & Mehra, 2010). Important events cannot be adjudged as such. Instead, the suggested events must have the root on the social media such as Twitter after the specified news inception, and it must take the serious impact on the social media through series of takes by the social media users (Lei, Rao, Li, Quan, & Wenyin, 2014). The system proposed and evolved in this research monitors a discussion thread whether it has reached its peak or decline after its root in the social forum like Twitter. Eventually, it would give us insight into the evolutionary trends of the specified events over time.

Thus this research aims to address the problems of entity ambiguity and its associated entity types for purposes of disaster management. We categorize the disaster based entity domains using ontology and enhance searching capability by incrementing the explicit connection which mutually exists between entity and ontology class. In order to achieve this task, we identify major issues to deal with and study them thoroughly for efficient processing of the following results: (1) Twitter streams and conventional news media tend to offer almost similar types of news coverage for a specified event but the rate of distribution among topics/categories differs. (2) On specific events such as disaster, crisis or any emergency situations, the volume of information accumulated between the two news media stands divergent and filtering the most potential information poses a challenging task. (3) Relational mapping/co-occurrence of terms suits well conventional news media but due to shortness and sparseness of tweets there remains a bottleneck for researchers. Therefore, we cover the details of the above-stated problems at length and propose algorithms and methods to exercise the cases conservatively.

In the following sections, we present semantic filtering of entities from Twitter streams and identifying the potential meaning of tweets using domain ontology. Besides, we highlight a semantic model for disaster situations and detail at length different ontological tools available for effective filtering of semantic content. In that connection, we propose the model by which to analyze the semantic mapping of entities/terms over the concepts and disambiguate the potential named entities. Eventually, we detail the analysis and present the empirical results obtained from our proposed model that lead us to the conclusions.

RELATED WORKS

In recent years, social media has gained momentum over collecting real-time events, and it has been proven that it has given the appropriate responses over the time of crisis when compared to other sources of decision-making systems. Celik et al. (2011) and Lei et al. (2014) did carry out timely searches over the crisis-related events even when the access to the online events was dropped consistently due to network latency and data traffic on the social media sites. Also, Abel et al. (2011) stated that extracting the relevant content over the crisis-related situations turned mostly ambiguous and many times redundant information. The key challenges that they addressed in the paper are how to overcome the difficulties in avoiding ambiguous results and removing redundancy. Sakaki et al. (2010) empirically discovered that during the course of extracting crisis-related information, compre-

hending the phrases entered by social media users into the site is directly affecting the search results. Hence they used Wikipedia and other organized knowledge sources to map the canonical terms thus yielding better results through employing term dissimilarity.

Grolinger, Capretz, Shypanski and Gill (2011) empirically showed event identification using statistical methods to observe an event's history and concluded on some proximity-based estimation of preciseness. They proposed a robust system that employs an algorithmic method based on 'Latent Dirichlet Allocation', which monitors the events and omits the content that is not relevant. They computed the average Euclidean distance between events, segregated abnormal changes present in the streams so that unknown distribution of data would be neglected, and got the Bag-of-Words. Another approach that they proposed in the algorithm is the log-likelihood rate by which the statistical ratio of data and TF-IDF difference present in the user-generated streams can be normalized based on term weighting and similarity score. In Wirtz, Kron, Löw and Steuer (2014), classification algorithms were used largely for event identification and classification. Many supervised learning algorithms were applied to divide the user-generated streams into the already-defined topic categories and carry out the detection process much easier than before. Approaches such as text classification and named entity recognition were extensively employed for exploring hidden events and disambiguating the selection process (using a vector space model to increase the viability of event detection).

Liu, Brewster and Shaw (2013) explored the facts that covered the most common patterns on the disaster-related content and stored the details in a separate database. Weidong, Jidong, Jia, and Danni (2012) and Liu, Shaw, and Brewster, (2013) developed a system that detects online news events and searches related events on social media for the widespread collection of related event information. In contrast to our approach of ontology-based discrimination, all of the references mentioned above are statistical in nature.

Lei Li and Tao Li (2013) developed a domain ontology for analyzing multi-document summarization pertaining to disaster or any crisis management. They provided many experimental models for developing an efficient ontological framework specifically on Hurricane Wilma in 2005. With that ontology, they precisely demonstrated the ontology-based multi-document summarization that performed well compared to other existing models. Jerman-Blažič, Matskanis and Bojanc (2017) discussed differences between man-made and natural calamities in detail and delineated the strategic measures, such as preparation, response, and recovery of any crisis or disaster-related situations. Their model was approved by the European Union and funded for the project REDIRNET. Later Selvam, Balakrishnan and Ramakrishnan (2018) constructed an ontology for Social Event Detection (SED) and applied it for Flickr (online photo management and sharing application) website by extracting the metadata features, such as geolocation, photoID, tags, description, title, timestamp, etc. On the other hand, they extensively utilized the Linked Open Data (LOD), such as Last.fm, Eventful, GeoNames, FourSquare, and so on, for productive discrimination of ontological properties. Although ontology/metadata-based, these works considered full documents or photo metadata in contrast to our study of short bursty messages, for example of Twitter.

SEMANTIC TECHNOLOGIES IN DISASTER MANAGEMENT

The primary objective of Semantic Web technologies is to pave the way for users to easily find relevant information, navigate among diverse data sources and integrate heterogeneous information. For example, usage of semantic technologies will be highly important in getting relevant content and linking data elements to search concepts in the case of Twitter content during any heavy crisis or mass convergence event (Schulz, Ristoski, & Paulheim, 2013; Yates & Paquette, 2011). Nevertheless, the task is complicated because all such processes should be machine-readable and automated; this is where the semantic Web technologies come handy in affecting ontological enrichments (see Table 1).

Table 1. Semantic technologies for unambiguous data handling

Component Name	Cached	Description
RDF, RDFS	YES	Vocabulary & Markup supported
MicroFormats	YES	Coupling Vocabulary & Markup Languages
Data RSS Feeds	YES	Fetching Atom and Metadata
XSLT	NO	Define data prototyping
Web Services	NO	Interoperate with remote data objects

In the rest of this section, we expand on the need to semantically enrich social media content, survey semantic Web technologies to that end, and how to integrate data from various sources towards facilitating crisis management.

SEMANTIC ENRICHMENT OF SOCIAL MEDIA CONTENT

As the semantic Web technologies play a seminal role in extracting meaningful information from social media and in the context of any crisis management (Heath & Bizer, 2011; Liu, Shaw, & Brewster, 2013), robust methods are required in dealing with different expressions of text that all in turn point to the central concepts. To get out this discrimination existing in the social media messages, the sole ways of solving this complexity is through making the Web machine-readable. To deal with the above-stated problems, semantic technologies provide an efficient technique called “named entity linking.” Named entity linking is the process of detecting the entities prevailing in social media messages and associating the entities which are closely matched to the specified event (see Figure 1).

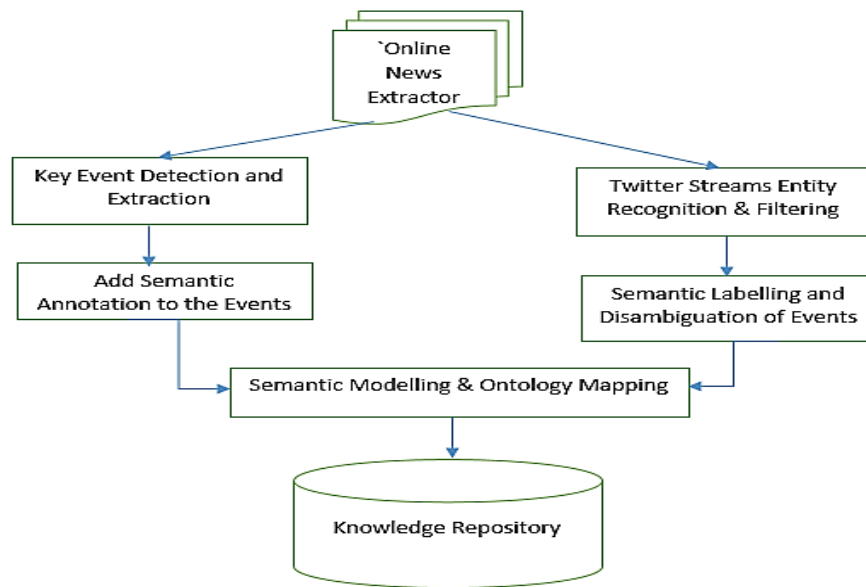


Figure 1. Semantic enrichment of news contents

The named entity linking does two things. Firstly, it crawls the social media messages related to the crisis event detecting the potentially emerging entities like names, places, locations, organizations, category, etc. Table 2 displays ontological relations instrumental in extracting potential named entities of various genres using semantic technologies.

Table 2. General domain-based ontological relations

Event	General Relation
Identify Relation	is-a, isAbout, defines, occurs, exists, classifies, express, describes, isRelatedTo, sameAs.
Temporal Relation	hasTime, timeInterval, time-span, timeStamp, during, eventDate, begin, end, since, nextTo.
Spatial Relation	place, region, space, location, hasBoundary, nearTo, direction, overlap, placeName
Causal Relation	cause, result, factor, agent, actor, action, activities, impact, consequence, result, participant, role, product, instrument.
Exceptional Relation	isPartOf, hasSubEvent, hasComponent, hasMember, unifies, includes, involves, transitive, symmetric, negative, opposite

Secondly, for every candidate event found, the named entity links searches for events in proximity to the named entity in that context. Unlike conventional search engines, it never searches based on a keyword. Instead, it makes use of the ontologies to augment the search process rendering the system to automate the process through enabling techniques such as RDF, RDFS, and OWL (Celik et al., 2011; Gruhl, Nagarajan, Pieper, Robson, & Sheth, 2009). Table 3 provides a sample of ontologies relevant in crisis management domain.

Table 3. Domain specification of ontologies for crisis management

Domain Specification	Ontology Name	SW Representation
Resources	SoKNOS	OWL-DL
	MOAC	RDF
	SIADEx	Unknown
People	FOAF	RDF
	BIO	RDF
Organizations	IntelLEO	RDF
	Organization OL	RDF
	EM-DAT	Online Query
Disaster	UNEP-DTIE	Online query
	Canadian Disaster Database	Application System
Damage	HXL	RDF
Infrastructure	OTN	RDF
	EPANET	Development
Geography	GeoNames	RDF
Meteorology	NEW weather ontology	OWL
	Ordnance Survey	OWL
Hydrology	Hydrology Ontology	

Once the named entity linking process is completed, thus having semantically enriched the messages present in the social media, then users can search for the information they want. This is called “faceted search”. In faceted search, information is not crawled based on the keyword supplied but by associating the concepts for the term through proper ontological support (Grolinger et al., 2011; Malizia, Onorati, Díaz, Aedo, & Astorga-Paliza, 2010). For example, if we give the search word “virus”, it will not search the information through using the keyword “virus”; instead, it finds relevant concepts for the keyword and associated terms in the ontology hierarchy and then related media messages based on the ontological concepts are returned.

PERTAINING CHALLENGES

Tweets extracted from Twitter entails a huge amount of challenges to overcome in order to put them to use for a particular purpose. Wirtz et al. (2014) pointed out precisely that there would be no metrics stipulated about how much of information need to be monitored, extracted and evaluated. The following are major problems that present various challenges in yielding useful results:

- a) Tweets extracted cover different aspects of the event and fail to make the distinction between the choices of the events.
- b) Many tweets tend to be very noisy and sometimes irrelevant to the event thus causing unnecessary computational problems.
- c) No measure to counteract rumors germinated during the events and it has been feared that it would spread vehemently over the short course of a period.
- d) Dealing with misspelled tweets is a big task since there had been no apt use of a dictionary to makeover.

EVENT EXTRACTION

Event extraction from Twitter is carried out through the Twitter Streaming API, which is the standard application fulfilling the filtration process effectively (El-Halees, & Al-Asmar, 2017; Sakaki et al., 2010). It extracts the tweets for the event by crawling through the hashtags created for the events by various NEWS sources on Twitter and monitors the user-generated posts subsequently (Silva, Wuwongse, & Sharma, 2013). The crawler is started to fetch the tweets that have been monitored and stored in the data store. Each and every tweet consists of the original post, author, timestamp, geographic information and hashtag from which it was obtained. All these properties are very useful in deriving the patterns and identifying the purpose of the posts.

The formal definition for entity extraction of Twitter streams is expressed in the graph theory as $G(E,V)$ where E and V represent some set of edges for the given set of vertices. To determine the potential social entities prevailed in the twitter streams and build the appropriate relationships between the entities, we link the set of edges E , in which $v_i \in V, i = 1 \dots N$ denotes the extracted entities in twitter streams and $v_i v_j \in E$ denotes relationship between entities $v_i, v_j \in V$. In this connection, to estimate the candidate entity for the query q , the search engine would normally generate most ambiguous entity sets about the given candidate entity, and it is termed as

$$a_i = q \leftarrow a_i \quad (1)$$

However, in our proposed approach, we have introduced a novel method to tackle this ambiguity prevailed over the search results by incurring the semantic web ontology for the domain at which the entity is dealt with through the appropriate level of ontological weight, and it can drastically reduce with the addition of ontology candidate keyword KW , i.e. consequent of

$$aw_i = q \leftarrow a_i, KW \quad (2)$$

which reduced the entity ambiguity with which $|aw_i| \leq |a_i|$, $|a_i| \in a_i$ is a cardinality of a_i and $|aw_i| \in aw_i$ is a cardinality of aw_i , KW. By utilizing a well-formed query for the candidate entity in the query, the named entity information would come as

$$a_{i''} = q \leftarrow "a_i" \quad (3)$$

In some cases, $|a_{i''}| \leq |a_j|$ and $|a_{i''}| \in a_{i''}$ is a cardinality of " a_i ". As well as with

$$aw_{i''} = q \leftarrow "a_i", KW \quad (4)$$

is about one of the information concentrations of a named entity. Then after pruning the entity cardinality, in the next process, the relationship between two named entities is based on the concept of co-occurrence. Thus,

$$a_i a_j = q \leftarrow a_i, a_j \quad (5)$$

which is a process to augment the semantic similarity between the two named entities and build the relationships between them, with which $|a_i \cap a_j| \leq |a_i|$ and $|a_i \cap a_j| \leq |a_j|$ and $|a_i \cap a_j| \in a_i a_j$ is a cardinality of $a_i a_j$. Besides, with the supplement of a keyword towards the co-occurrence will usually subside the number of entities given, and that is

$$aw_i aw_j = q \leftarrow q a_i, KW \quad (6)$$

But it should satisfy that $|aw_i \cap aw_j| \leq |a_i \cap a_j|$, $|aw_i \cap aw_j| \in aw_i aw_j$ is a cardinality of $a_i a_j$, KW. Similarly, effective utilization of the well-defined entity set for the query will yield the appropriate relationships between the two named entities.

ONTOLOGICAL INCLUSION FOR DISASTER MANAGEMENT

Semantic technologies that support crisis events identification are often required for interacting between different developers and software applications operated by various agencies. In this context, affecting the semantically enabled system to communicate the information in a unified format is the critical challenge, and social media platforms behave differently to address the dimension of the problem to interrelate with one another. Interoperability shortcomings at the semantic level of concepts can be alleviated using common vocabularies as well as shared concepts in linking the whole processes (Liu, Brewster, & Shaw, 2013; Rajpathak & De, 2016). The best way of accomplishing this critical task is through the use of machine-understandable ontologies that can precisely define the concepts, categorize the events based on clustering approach and build an appropriate path between concepts and unified communication.

The next aspect of the challenge in retrieving the crisis-related information is in extracting related events or messages from blogs, forums, and referring wikis. These are the places where vibrant information is present at a high rate and shared opinions and suggestion made by several authors. Now the critical challenge is interlinking not only social media platforms but also these blogs and forums. It is very difficult in repurposing the content and tough to identify the common events among these sites. For instance, take Wikipedia – it is a huge repository of publicly accessible knowledge source but reusing the same knowledge for other applications presents new challenges and difficulties (Kumar & Muruganatham, 2016; Yates & Paquette, 2011). Furthermore, a user can create accounts on many sites like in blogs, forums, wikis, and other social media platforms, but it is very complicated to inter-relate the candidate entities among these different social sites. The major problem pertaining to these media sources is that the information items in such sites are entirely disconnected and completely separated from one another. There is an absolute lack of exchanging the semantics of entities and unable to derive the facts from such information silos. Every site holds the information posted by their registered users independently and, at times, it has turned into stagnant information silos which are untapped by others.

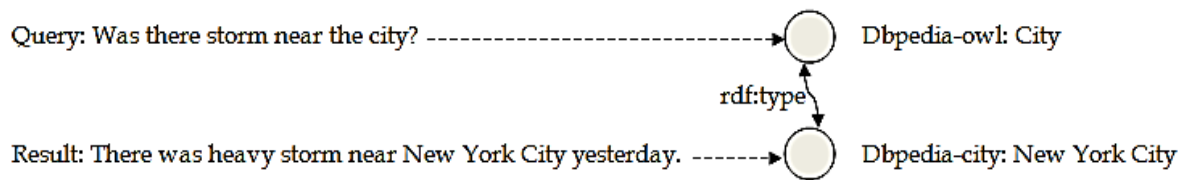
To meet the challenges rising from crisis or havoc situations, there would be a huge demand to de-centralize the process and enable interactive processes to fetch hidden relevant facts from the content. Table 4 gives the details of handling the events from the time news originates to planning for future action taking against the crisis events (i.e., past to present).

Table 4. Information handling for crisis situations

	News Inception	Present State	Future Plan
Goal	Inform & Publish the news	Share & Collect the news	Engage & Prune
Main Activity	Gather relevant event-based information	Track & Monitor the events	Prepare the action
Content	Initially, discrete data	Clustered Data	Find Relationships
Information Handling	Confidential	Privileged Access	Absolute transparency
Software Tools	In-house Software	Commercial Software	Open Source Software

Semantic Web recently provided the necessary tools for effective information linking and interoperability. Moreover, many semantic Web vocabularies have successfully been deployed at various social platforms facilitating machine-understandable message processing (Benali, & Rahal, 2017; Ritter et al., 2012). Some of the semantic Web vocabularies are RSS, FOAF (Friend of a Friend), and SIOC (Semantically Interlinked Online Communities). With the help of these and other more refined semantic vocabularies, interlinking communities and social sites became effective and helped curtail down information redundancy.

For example, let’s consider the query for crisis-related content on Twitter such as “Was there a storm near the city?” In this query, the name of the city is not mentioned, but the tagging engine like DBPedia Spotlight (<http://wiki.dbpedia.org/projects/dbpedia-spotlight>) and OpenCalais (<http://www.opencalais.com/>) would annotate the given query and fix the appropriate entity for the annotated tokens based on the Agglomerative Clustering techniques applied on the collected tweets. Since there is a semantic link (i.e., `rdf:type`) between the query and DBPedia, it can be computed based on the similarity score and higher relevance of content. It is illustrated in the following:



ONTOLOGY LANGUAGES

The so-called ‘semantic Web stack’ comprising a number of semantic Web languages was suggested for effective utilization in information processing and that, by the way, leads to efficient semantic implementations on the retrieval systems. The first language that was suggested to use was the Resource Description Framework (RDF), by which the basic framework to represent the potential information on the Web content was availed. The basic structure of any RDF statements is just a triple (subject, predicate, and object), which further yields the hierarchical RDF graph to prune the data in a much effective way. In simpler terms, RDF statement is just denoting the relationship existing between the existing things called nodes and that node is interconnected with other nodes (i.e., semantically related nodes). RDF serialization uses XML syntax and terms such as element name, attributes, values, etc. The scope of using RDF is to make the system machine-readable and process the infor-

mation semantically (see Table 5). Besides, it integrates the data without any serious glitches as it follows the well-formed logic which is universally acknowledged.

Table 5. Semantic Web languages and their structural contents

SW Languages	Structural Contents
OWL-S	Services in a machine-understandable format
SWSL	
WSML	
OWL	Equality Relation between Classes Cardinality Restrictions for Properties Relation of Properties
RDF Schema	Classes Objects Properties
RDF	Prepositions as Triplets
WSDL	Services in an exact human understandable format

A set of RDF statements form an RDF graph through interconnected nodes. As the RDF graph is conventionally followed in expressing the logical facts about the potential named entities, ontologies are used to give the domain and category of each thing (i.e., entity) and yield the appropriate relationship between them. Ontologies contain features to wholesomely express the rich relationship between entities and also set appropriate constraints on them (Grolinger et al., 2011; Wang et al., 2018). The language followed for effective creation of ontology is RDF Schema and OWL (see Tables 6 and 7). The RDF Schema (RDFS for short) employs the set of classes related to the entity and chooses the properties according to its domain.

The basic objective of RDFS is to provide a well-structured description of entities properties. Some ontologies used to set classes and properties are called RDF vocabularies, and examples include FOAF, SKOS, MOAC, Dublin Core, etc., whereas OWL is facilitating the information interoperability and providing additional vocabularies to enhance the formal semantics of RDF Schema.

Table 6. RDF and OWL Ontological Constructs

RDF/OWL	Category	Functions
Class	Definition	Class
		Enumerated Class
		Restriction
	Axiom	IntersectionOf
		UnionOf, ComplementOf
		subClassOf
Relation	Definition	Equality
		disjointWith
		Property
	Axiom	Domain, range
		subPropertyOf (Inverse)Functional

RDF/OWL	Category	Functions
		Equality, inverseOf
		Transitive, symmetric
Instance	Definition	Type
	Axiom	(In)Equality

ONTOLOGY EDITORS

In order to derive the meaning out of the collected information from Twitter streams and as we process the tweets into the respective semantic representations, there is a need to design an ontology that is well mapped to the information and facilitates fetching the content in the hierarchical format (Liu, Shaw, & Brewster, 2013; Malizia et al., 2010). In this connection, we are required to build ontology using standard text editors that are very simple in design and development. Besides, it can be formatted with the semantic Web languages called RDF/XML format. One such editor is called Protégé (<https://protege.stanford.edu/>) which is a tool that permits designing OWL ontologies and further helps to connect to the data interrelated with one another in the overall ontological framework. With Protégé, querying and reasoning help to disambiguate the information and assist in making the filtration absolutely error-free. Other popular ontology editors are SWOOP, OntoStudio, NeOn, Altova, WebODE, and so on. Among all the editors, Protégé was the first, freely available, and open-source ontology editor and framework for building intelligent systems thus tops the list and widely deployed in much recent research.

Table 7. A notional mapping between RDF/OWL and relational concepts

RDF/OWL Terms	Relational Concepts
rdf:class	Table
rdf:property	Column
rdfs:domain	Table that the rdf:property is a column of
rdfs:range	Data Type of the column
rdf:type	Values of the Primary Key column

SEMANTIC MATCHING AND TRANSLATION

As ontologies play a seminal role in semantic processing of information (Celik et al., 2011; Sheth et al., 2010), we should, therefore, try harnessing the potential meaning hidden in the collected information streams. Ontologies help process the meaning of different terms represented in the information and avail the system to understand the very basic structure of information in a precise way. The system processes the collected data automatically and does matching and translation with its rich vocabulary sets, which is fed as the dataset to the ontology while designing the complete framework of the domain ontology (Gruhl et al., 2009; Madani, Boussaid, & Zegour, 2015). Each term (i.e. a concept) in a tweet is mapped with corresponding vocabulary sets in the specific ontology domain, and sometimes it arches to other domains of vocabulary set to find the exact meaning of the concepts represented in the tweet. Precisely, the inclusion of mapping the terms over multiple ontologies is the biggest challenge in designing an application that is used to integrate the ontologies to disambiguate in parallel; that is a challenging research area to deal with. Further, to make the automation of semantic matching and translation effective, appropriate use of mapping rules over the information is necessary and should be defined using ontology matching tools. Ontology matching is variously called also as ontology aligning, mapping, and translation (for example, for Web services discovery: Fellah, Malki, & Elçi, 2016). Conventionally, ontology mapping tools come in two categories: element-based approach such as name similarity, entity similarity, concept similarity, etc., and structure-based approach such as sub/super-categories, -domains, -levels, etc. Besides, mapping requires infus-

ing external knowledge such as Thesauri, WordNet, etc., to yield precision and high recall. Some of the ontology mapping tools available on the market are RiMOM, ASMOV, and AgreementMaker.

SEMANTIC SEARCH

The next level in semantic utilization of social media harvested information is through semantic search, which should return results without any ambiguity and sparseness. As semantic mapping directly links information repositories, domain ontologies should facilitate the operation of effective semantic search (Kumar & Muruganatham, 2016) retrieving facts that are interconnected with one another in the ontological framework. In order to render the search process easier, appropriate use of indexing methods in the ontological inclusion over the concepts is deemed important. Ontologies follow the semantic indexing approach using its standard principle of “indexing the RDF triplets”, thus smoothen the way of semantic search over the collected information. Several pieces of research have been carried out in this regard to fetch the precise and unambiguous results by semantically integrating information from diverse ontological frameworks in retrieving the results from multiple repositories.

INTEGRATION OF DATA

Another critical issue faced in utilizing social media content is the integration of data, which may be considered from two different perspectives: data source (database/stream) reconciliation, and information integration. The basic objective of overcoming the problems of semantic heterogeneity between these two categories using the appropriate ontology framework is a challenging task to any research (Liu, Shaw, & Brewster, 2013). In considering database integration, the role of ontologies lies on the upper layer of the schema (i.e, semantic matching of information and table schema should be shared with the domain ontology and additionally, make use of the ontology to integrate the database schema rightly towards its order). While integrating information the core problem lies at integrating the terms from various sources and mapping the potential candidate terms relevant to its vocabulary sets and bring into the consolidated view called the new collection of a derived set. The challenge that lies here is to not change the original sense of the terms while mapping to the appropriate sense of terms in the vocabulary sets. Besides, in database integration, ontologies convert tables into respective classes in RDF triplet and columns in the table into data relation in the RDF Schema. The data models followed in the semantic conversion would always be 1:1 mapping cardinalities. In recent works (Kumar & Muruganatham, 2016; Kwak, Lee, Park, & Moon, 2010), the authors propose rules to dynamically map the data models into ontologies and consider mapping instances to class levels. Also, some language constructs are given to fetch data objects and, using the queries, they can be annotated dynamically representing in RDF.

Whereas in information integration, accumulating the terms from various sources of datasets to bring them into a unified collection, several efforts were carried out in the recent past but failed to resolve it. Some early researchers have tried to apply the Description Logic (DL) as ontology language and observed few changes in the outcome. Later, the Prolog programming language is employed for expressing the information formally and integrating the terms using appropriate domain ontology.

In the next section, we introduce our proposed information-centric model for management of crisis and disaster based situations through integration of many of the technologies mentioned above combined by our innovative approach. Our proposed model is introduced, followed by the empirical tests and discussion of findings.

THE PROPOSED MODEL

The objective of this research is to potentially harness the information gathered from various social media platforms and render it relevantly interconnected with the selected news articles. In doing so,

here we introduce some of the notation and problems that we define formally before presenting the Semantic Search for Events Algorithm.

Problem 1 (News stream): For every news article related to disaster or crisis situation, content must be analyzed and scrutinized for a further level of comprehension. Let $N = \{n_0, n_1, \dots, n_i\}$ be the news posted on various sites and gathered from various news agencies. For every news article posted say n_i , we find the actually published time $t(n_i)$. Since the origin of news story gives the real arrival of news, it brings in the proximity among related news articles.

Problem 2 (Tweet stream): Upon the arrival of every news article related to disaster or crisis, the next task of the system is to identify the equivalent social media content such as Twitter where the relevant news item is discussed and promulgated. Let $S = \{s_0, s_1, \dots, s_j\}$ be the Twitter Streams for the taken news articles and load the inter-related Twitter messages posted by various potential social users. For every tweet s_j , we find the actually posted time $t(s_j)$ and responses for the message.

Problem 3 (News recommendation problem): Once the news items N and its associated social media contents (Twitter Streams) S are mapped, then the real task is to find the top-k most relevant news for the topic. Let's take the set of users interacted on the particular news topic $U = \{u_0, u_1, u_2, \dots, u_n\}$ in the social media platforms and explicitly categorize the social messages and news streams of general interest (i.e., for any social user $u \in U$ at any point of time T , we recursively adopt the functional ranking which links the users interest among its neighbors).

Problem 4 (Social influence): In order to find whether the news item $n_i \in N$ has influenced the social media users $U = \{u_0, u_1, u_2, \dots, u_n\}$ effectively, we give the social influence model $S = |U| \times |U|$ matrix where $S(i, j)$ calculates the cumulative interest of the selected users u_i to the user-generated content by u_j . This process states that each user in the context would pose an absolute interest to the user-generated content posted by the other user.

Problem 5 (Tweets-to-news model): To merge the process, let N be the order of news collected and S be the streams of social media messages, we model the relationships between user-generated content and news items as $M = |S| \times |N|$ matrix Z where $S(i, j)$ is the closest proximity of user-generated content s_j to news item n_i .

Algorithm (Semantic Search for Events)

Input: Seed words for each crisis event

Output: Generation of semantic classes

BaseTerms \leftarrow set of seed words given;

for i : 1 to N (Number of Iterations) do

BaseTerms \leftarrow ExpansionOf (Seed words, Corpus);

BaseTerms \leftarrow Cluster (BaseTerms, Seed words, Corpus);

end

return BaseTerms

The algorithm carries two significant operations: (1) expand the seed words with the assistance of ontologies; and (2) cluster the events based on the similarities existing in the classification. For the given seed words for the event, it crawls for new terms that possess the similar distributional features to the seed word and assigned to the set of seed words (also called as candidate words). In fact, extracting the new terms for the seed word can be done based on the contextual features and top score similarity measure. For the clustering, the selection procedure would process the learned terms and

seed terms based on the distributed similarity and set the minimum threshold value for estimating the exact precision of the terms.

The sole plan of this algorithm is to get the patterns which are interlinked with a semantic relation and bring in the semantic class for the search terms (Kumar & Muruganatham, 2016; Wang et al., 2018). As listed in the algorithm, it has three core operations for finding the patterns existing in the disaster based situations:

1. using the semantic class expansion algorithm, extract the candidate terms for the disaster event tags;
2. find the patterns for the candidate terms selection and fix the semantic category of the events; and
3. choose the cluster events which hold similar action terms and evaluate the patterns for further classification.

It has been noted at several instances (Abel et al., 2011; Sheth et al., 2010; Wirtz et al., 2014) that news items and user-generated content at social media platforms (say, Twitter Streams) co-exist with one another with same news topic (see Figure 2). Sometimes, a published news story is pushed into social media platforms for further discussion and circulation. And, at many times, a news item first discussed vehemently on social networks then becomes a topic in news stories (Lei et al., 2014; Li, Liu, Li, Qin, 2016). In these two cases, the predominant factor is holding the current trending entities, which give the unflinching bond between social networks and news sites. There is an absolute relationship between user-generated content and news stories, which create an intermediary layer that paves the way to generalize the analysis. Hence, this would make our work equally applicable in deriving the ultimate decision for disaster management and assess the core patterns for decision making.

During the analysis of the relationship between events and results, there would emerge a need to attain a similarity score for the ultimate decision process. The similarity score for the crisis/disaster management (Liu, Shaw, & Brewster, 2013; Malizia et al., 2010; Schulz et al., 2013) can be accepted and formulated based on the following assumptions:

1. If there would be a high or low hazard during the disaster situations, then it requires the scientific or technical measure to be assured, and precautionary steps should be taken based on scientific or technical grounds.
2. If there would be high or low outrage, then it is an emotive issue and should be tackled through proper negotiations or political balance.

In these two cases, the analysis of the events played a crucial role in disseminating the user-generated content posted on social networks and determining the effective decision-making process (Heath & Bizer, 2011).

DISASTER ONTOLOGY

To substantiate our proposed model, we constructed an ontology for disaster datasets with a glossary consisting of more than 150 definitions (i.e. mostly of recurring terms) and further accumulated terms related to disaster from books, papers, survey on seismic risk and other relevant disaster web sites. We constructed the ontology with associated concepts, its attributes and proper relationships between concepts. In constructing this ontology, we followed many dictionary terms (i.e. entities related to disaster) with their associated meanings (axioms) and connected the terms with taxonomic relationships. Relationship mapping of terms can be done in many ways such as Taxonomic (IS-A relationship), Meronomic (PART-OF relationship) and Telic (PURPOSE-OF relationship). The relationship mapping of terms can be achieved through inference rules to augment better reasoning and increase the credibility ratio of knowledge representations. On this proposed system, we used Protégé, an ontology tool which is more of an object-oriented paradigm and well suited for term inher-

itances. The relationship IS-A is a generalization/specialization between the candidate entities: super-class entities publically generalize the subclass entities and the sub-class entities particularly employ specialization of superclass entities. Likewise, Protégé permits to formulate the disaster ontology by considering different instances to insert and able to accommodate a huge set of information for a digital archive.

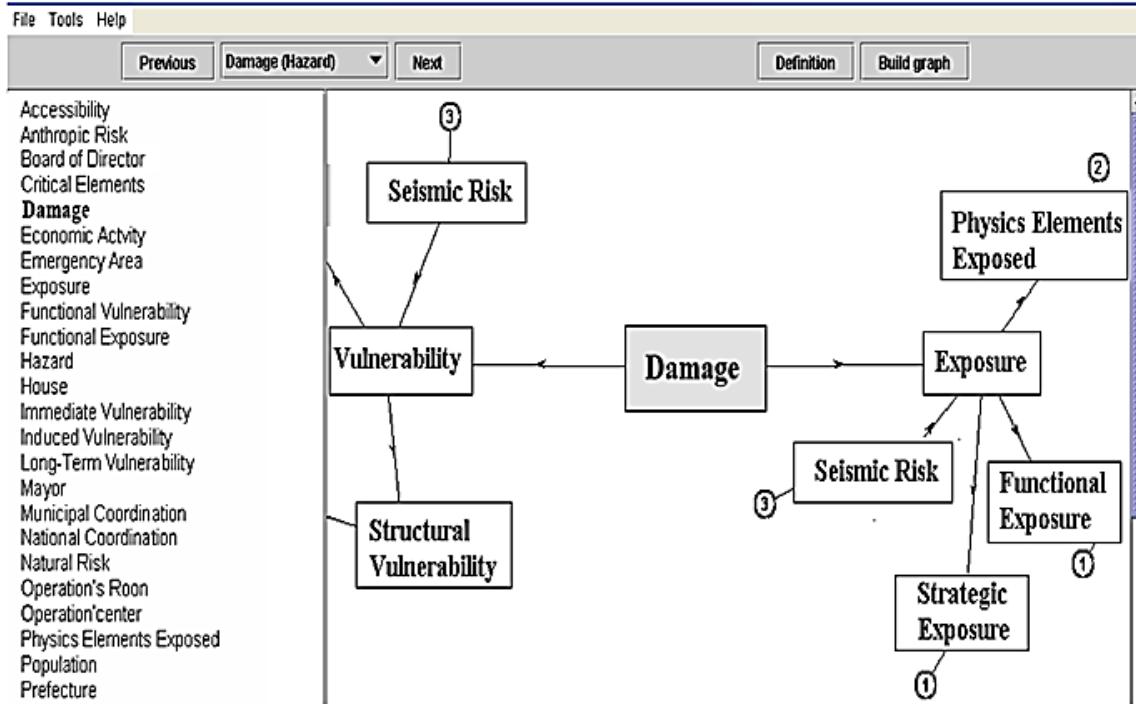


Figure 2. Disaster Ontology using Protégé

Our proposed system concerns mostly about urban risk with specific governance on seismic risk management. The effective building of this ontology paves the way for common knowledge, makes the concepts understandable, and prompts information into unambiguous semantics. This ontology construction has been performed in three steps:

1. Fetch the core concepts of the domain (Seismic Risk) and relevant terms in the glossary.
2. Extract the Super-Classes and Sub-Classes of the concepts using the IS-A relationship.
3. Find other related types of relationships using inference rules (properties, slots, and roles associated with each concept).

Relation mapping for the collected tweets can be performed and filtered using the relational properties displayed in Table 8). Entity resolution and disambiguation have been effectively dealt with in Disaster Ontology constructed above and resolve the term ambiguity persisting over the collected documents (Twitter streams).

Table 8. Relationship mapping between concepts and classes

Relation Name	Source	Target	Description
isResponsibleFor	Department	Process	Identify which sector is responsible for the event and map the relationship between department and process
workIn	Actor	Department	Map the relationship between the person and the department. Identify the actor responsible for the event.
isPartOf	Task	Process	Find the task which is responsible for the process and filter out the concepts related to the event.
isA	-	-	Relationship between super-class and sub-class
Perform	Actor	Task	Group the actor performed the task on the event.
Produce	Task	Information	Filter the information for the task on the event.

ENTITY RELATIONSHIP AND RANKING SCORE

The disaster ontology has now become a knowledge source for our disambiguation effort. When we process each and every tweet, we find the exact match of those entities against the knowledge source such as DBpedia or YAGO. If it is not present, then it sends the NIL result. Now, by means of our proposed method, we can again cross-match with our own ontology created from news articles and find the exact match of those entities. In this method, the accuracy is relatively high because the created ontology is extracted from news articles related to the tweets and context of the news articles is highly relevant and appropriated match with the tweets. If we go for the entity–mention match with DBpedia, it lists out candidate mentions for the entity, and we need to probe for the context pertaining to the tweet. But if we match the same with our own ontology, it is exact and gives an appropriate match.

Hence, in our approach, we take the link probability (Kumar & Muruganatham, 2016; Yates & Paquette, 2011) for the entity with DBpedia mention, and it can be defined as follows:

$$F_{(e,m)} = \frac{\text{Count}(m,e)}{\text{Count}(m)} \quad (7)$$

Here, we utilized an outlined ontology to arrange the mentions for the given named entities and appropriately estimate the similarity distance between them. Now the task is to estimate the distance between the entity and the suggested set of mentions from DBpedia. In this connection, we have taken the Cosine Similarity measure to access the similarity difference existing between the entity and candidate mentions as follows:

$$\text{CosSim}(e, m) = \frac{\text{Product}(e,m)}{\|e\| * \|m\|} \quad (8)$$

By this method, we categorically filter the exact match of mention for the given entity and appropriately reference with DBpedia URI as stated in (Liu, Brewster, & Shaw, 2013; Malizia et al., 2010; Schulz et al., 2013). We utilized the DBpedia Spotlight to get the URI match of each entity and return the JSON results for our implementation.

```
def filter(entity):
    return JSON (DBpediaSpotlight.annotate(entity));
```

The result of the proposed approach would create a binary mapping of the entity and mentions, as seen in Table 9.

Table 9. Identifying the relation between named entity and candidate mention

<i>Mention</i>	<i>NE Class</i>	<i>NE Link</i>	<i>DBpedia Ontology Class</i>	<i>Score</i>
<i>Barack Obama</i>	<i>Person</i>	<i>Dbpedia: Obama, USA</i>	<i>Dbpedia-owl: Person</i>	<i>3</i>
<i>Chennai</i>	<i>Location</i>	<i>Dbpedia: Chennai, India</i>	<i>Dbpedia-owl: Place</i>	<i>1</i>
<i>Cricket</i>	<i>Sports</i>	<i>Dbpedia: Cricket</i>	<i>Dbpedia-owl: Sports</i>	<i>2</i>

Generally, entities in DBpedia have its name, label, type, etc. and, to fetch the entity name given in the DBpedia for the specified URI, it can be queried through the SPARQL query as follows. For example, searching for ‘Sachin Tendulkar’:

```
Select distinct *
where {
  ?URI rdf:label ?name
  ?URI dbpprop:iupacname ?name
  filter(str(?name) = "Sachin Tendulkar")
}
```

In order to get the category of a given entity from the DBpedia, we issue the following SPARQL query. For example, for ‘Vehicle’:

```
Select *
where {
  <http://dbpedia.org/resource/Vehicle>
  <http://purl.org/dc/terms/subject>
  ?categories.
}
```

EMPIRICAL TEST AND ANALYSIS

We used Twitter4J API to gather disaster-related tweets from Twitter and utilized TextRazor API to effectively recognize the potential named entities present over the tweets and link them accordingly to its respective DBpedia URI. Additionally, we used the rich natural language processing tools of Stanford Core NLP Library to segregate tweet patterns and performed sentiment analysis for grasping the sense of the tweets. Tweets were collected on the month of August 2017 and, to witness the trust, we followed the leading news agencies on Twitter such as BBC World, CNN, New York Times, NDTV, and Breaking News. Tweets were crawled and stored only if they had at least one named entity that has its link on DBpedia URI. In our datasets, we were able to filter out 20 different topics and classified the tweets successively based on seismic risk by applying the classification rules. The algorithm proposed above is able to detect the factual information containing about 3 out of 5 tweets.

Table 10. Event relevance and categories

Event Category	Total Events	Potential Sub-Events by Relevance		
		R3	R3+R2	R3-R1
Earthquake	75	35(46%)	51(68%)	59(78%)
Tsunami	120	46(38%)	79(65%)	88(73%)
Cyberattack	114	51(44%)	87(76%)	95(83%)
Unrest in a Country	150	77(51%)	90(60%)	97(64%)
Celebrity Death	115	43(37%)	66(57%)	81(70%)
Terror Attack	120	68(56%)	79(65%)	85(70%)

We tested the DBpedia corpus to identify potential events on seismic risk, which provided the six complex event categories listed in Table 10. The entities were extracted based on the recommendations stated above and identified their relationship types in corresponding DBpedia URI. Besides, we again queried the DBpedia Knowledge Source for the sub-events correlated with the events extracted from the tweets. We substantially ranked the sub-events on the basis of frequency of occurrence and chose the best-matched event category to a tweet. After evaluating the event categories against DBpedia, we determined whether the event is of positive instance or not. Sometimes, the retrieved events would pose a challenging task such as if it is partially relevant but not exactly appropriate to the categorized concepts. During these anomalies, we assigned the following three relevance scores in order to fit the events into their appropriate decks:

- Relevance (R1): Events with fuzzy relationship to the concept/category.
- Relevance (R2): Events with positive occurrences of sub-events or subject-object mapping.
- Relevance (R3): Events are positive instances and fit into the category for the posted query.
- Otherwise, the relevance zero indicates the events with absolutely NIL relationship.

Table 10 displays detected event categories and potential sub-events or co-occurrence of events with relevance scores. As was witnessed, the precision values varied considerably among the categories. The Stanford NLP Library was deemed fit to extract the potentially relevant tweets, and type filtering of events was absolutely effective at identifying the appropriately named entities. We obtained an accuracy of 74.13% and computed the Precision (0.641), Recall (0.716) and F-Measure (0.691) respectively for the given datasets.

DISCUSSION

The dynamic change in the amount of information gathered at the various medium of platforms indicates the need for a rapid decision-making process in crisis events. It was observed that the information gotten from these sources rapidly varied. Statistics (Wirtz et al., 2014) showed that the frequency of report variation grows ten times greater than the previous day. Besides, to better account for the report variation of the information accumulation, the report dimensions were categorized into three crucial breakpoints, i.e., $D + 1$, $D + 5$, $D + 10$. This elapsed gap fetches the detailed overview of the crisis or disaster based events and showed us the real potential of the event happenings (see Figure 3).

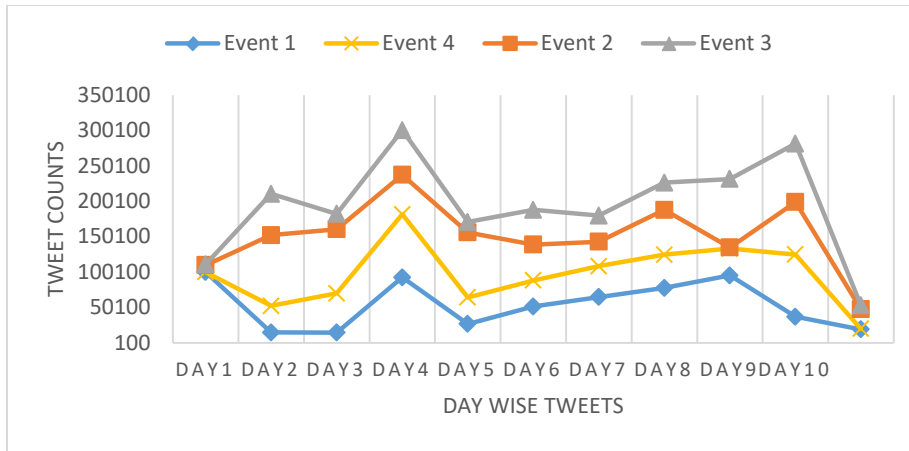


Figure 3. Daily frequency of information on social media platforms

Through the data obtained from various sources and on different days of report gathering, we can formulate deviance of patterns and get through the details of anomalies that exist in the report. By applying the pruning algorithm, we can sort the crisis events for the decision-making process and get to the core base of the events. In this research, the real task is to find the actual reason for the crisis event and get the substantiated evidence for its occurring. To augment this process, we classified the events into many chronological orders influenced by the usage of ontological background with semantic technologies. By mapping different day event reports, we scrutinize the process for discrimination (i.e., fetch the positive or negative or neutral feedback from the potential users on the social media) and allow filtering the facts based on cross-checking in tabulating the actual events of the situation.

Our approach achieved the accuracy rate of 74.13% where other existing models succeeded getting 68.42% using Support Vector Machine (SVM), 67.93% using Maximum Entropy Model (MEM), and 64.71% using Conditional Random Fields (CRF) based on the analysis successfully performed with the help of Table 9. Since our proposed model extensively uses the dedicated ontology of Crisis and Disaster, instead of employing the Bag-of-Words (BoW) method, we employed Bag-of-Concepts (BoC) and Relevance of Concepts, as well as calculating the semantic similarity score between ambiguous terms. Deep proliferation of the ontological network paved the way to yield the sub-categories of a topic and skimmed the words that are completely unambiguous. The relevance R of the concepts were derived with other three relevances R1, R2, and R3 as shown in Table 10, whereas the other existing methods mostly used only a single relevance score and restricted the research scope to Bag-of-Words model.

The major contribution of this research is in collecting crisis-related temporal data from multiple bursty short-message sources and decision making through semantic mapping of entities over concepts disambiguating potential named entities. The problems persisting over entity ambiguity and its associated entity types were addressed as well. We categorized the disaster-based entity domains using ontology and enhanced the searching capability of the system by incrementing the explicit connection mutually existing between entity and an ontology class.

CONCLUSIONS

In this paper, we proposed a novel solution to harvest and compare the content of Twitter streams and conventional news sources such as CNN, New York Times, BBC World, NDTV, and Breaking News in the cases of havoc situations. We developed a semantic filter that can map the concepts correlated between Twitter streams and traditional news sources, and can disambiguate the candidate entities based on the ontological framework particularly loaded with disaster/crisis events.

The major advantage of our work is that, instead of pruning a single news source, it paves the way for clustering the information from diverse sources and harnessing the potential information to derive the hidden facts in it. We also developed a disaster ontology for this research and used it to segregate the entities which pose ambiguity over other candidate sets.

Empirical results show that the approach based on our model outperforms other models available in the literature to solve this research gap by various other approaches. In the future, we shall strive to extend the model in order to help summarize and visualize the potential information ranked high by the model.

REFERENCES

- Abel, F., Gao, Q., Houben, G.-J., Tao, K. (2011). Analyzing user modeling on Twitter for personalized news recommendations. In J. A. Konstan, R. Conejo, J. L. Marzo, & N. Oliver (Eds.), *User modeling, adaptation and personalization. UMAP 2011. Lecture notes in Computer Science* (Vol. 6787). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-22362-4_1
- Benali, K., & Rahal, S. A. (2017). OntoDTA: Ontology-guided decision tree assistance. *Journal of Information & Knowledge Management*, 16(3), 1750031. <https://doi.org/10.1142/S0219649217500319>
- Celik, I., Abel F., & Houben, G. J. (2011). Learning semantic relationships between entities in Twitter. In S. Auer, O. Díaz, & G. A. Papadopoulos (Eds.), *Web engineering. ICWE 2011. Lecture notes in Computer Science* (Vol. 6757). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-22233-7_12
- El-Halees, A., & Al-Asmar, A. (2017). Ontology based Arabic opinion mining. *Journal of Information & Knowledge Management*, 16(3), 1750028. <https://doi.org/10.1142/S0219649217500289>
- Fellah, A., Malki, M., & Elçi, A. (2016). Web services matchmaking based on a partial ontology alignment. *International Journal of Information Technology and Computer Science*, 8(6), 9-20. <https://doi.org/10.5815/ijitcs.2016.06.02>
- Grolinger, K., Capretz, M. A., Shypanski, A., & Gill, G. S. (2011, May). Federated critical infrastructure simulators: Towards ontologies for support of collaboration. *Proceedings 24th Canadian Conference on Electrical and Computer Engineering (CCECE), Niagara Falls, ON, Canada*, 1503-1506. <https://doi.org/10.1109/CCECE.2011.6030715>
- Gruhl, D., Nagarajan, M., Pieper, J., Robson, C., & Sheth, A. (2009). Context and domain knowledge enhanced entity spotting in informal text. In A. Bernstein et al. (Eds.), *The semantic web. ISWC 2009. Lecture notes in Computer Science* (Vol. 5823, pp. 260-276). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-04930-9_17
- Ha-Thuc, V., Mejova, Y., Harris, C., & Srinivasan, P. (2010, September). News event modeling and tracking in the social web with ontological guidance. *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*, 414-419. <https://doi.org/10.1109/ICSC.2010.75>
- Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: Theory and technology* (Vol. 1, pp. 1-136). San Rafael: Morgan & Claypool. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>
- Jerman-Blažič, B., Matskanis, N., & Bojanc, R. (2017). Semantic ontology design for a multi-cooperative first responder interoperable platform. *Computing and Informatics*, 35(6), 1249-1276.
- Kumar, N. S., & Muruganantham, D. (2016). Disambiguating the Twitter stream entities and enhancing the search operation using DBpedia ontology: Named entity disambiguation for Twitter streams. *International Journal of Information Technology and Web Engineering*, 11(2), 51-62. <https://doi.org/10.4018/IJITWE.2016040104>
- Kwak, H., Lee, C., Park, H. and Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International World Wide Web Conference* (pp. 591-600). New York: ACM. <https://doi.org/10.1145/1772690.1772751>
- Lei, J., Rao, Y., Li, Q., Quan, X., & Wenyin, L. (2014). Towards building a social emotion detection system for online news. *Future Generation Computer Systems*, 37, 438-448. <https://doi.org/10.1016/j.future.2013.09.024>

- Li, L., & Li, T. (2013). An empirical study of ontology-based multi-document summarization in disaster management. *IEEE transactions on systems, man, and cybernetics: systems*, 44(2), 162-171.
- Li, Z., Liu, Y., Li, Q., & Qin, B. (2016). Relationships between knowledge bases and related results. *Knowledge and Information Systems*, 49(1), 171-195. <https://doi.org/10.1007/s10115-015-0902-z>
- Lima, R., Espinasse, B., & Freitas, F. (2017). OntoILPER: An ontology-and inductive logic programming-based system to extract entities and relations from text. *Knowledge and Information Systems*, 56(1), 223-255. <https://doi.org/10.1007/s10115-017-1108-3>
- Liu, S., Brewster, C., & Shaw, D. (2013). A semantic framework for enhancing information interoperability in emergency and disaster management. *International Conference on Social Media and Semantic Technologies in Emergency Response*, 1-20.
- Liu, S., Shaw, D., & Brewster, C. (2013). Ontologies for crisis management: a review of state of the art in ontology design and usability. *Proceedings of the Information Systems for Crisis Response and Management Conference, Baden-Baden, Germany*, 349–359.
- Luo, Z., Osborne, M., & Wang, T. (2012, May). Opinion retrieval in twitter. In *Sixth International AAAI Conference on Weblogs and Social Media*.
- Madani, A., Boussaid, O., & Zegour, D. E. (2015). New information in trending topics of tweets by labelled clusters. *Journal of Information & Knowledge Management*, 14(3), 1550019. <https://doi.org/10.1142/S0219649215500197>
- Malizia, A., Onorati, T., Díaz, P., Aedo, I., & Astorga-Paliza, F. (2010). SEMA4A: An ontology for emergency notification systems accessibility. *Expert Systems with Applications*, 37(4), 3380-3391. <https://doi.org/10.1016/j.eswa.2009.10.010>
- Moran, S., & Lavrenko, V. (2014). Sparse kernel learning for image annotation. *Proceedings of the International Conference on Multimedia Retrieval, Glasgow, UK*, 113-120. <https://doi.org/10.1145/2578726.2578734>
- Otegi, A., Arregi, X., Ansa, O., & Agirre, E. (2015). Using knowledge-based relatedness for information retrieval. *Knowledge and Information Systems*, 44(3), 689-718. <https://doi.org/10.1007/s10115-014-0785-4>
- Rajpathak, D., & De, S. (2016). A data- and ontology-driven text mining-based construction of reliability model to analyze and predict component failures. *Knowledge and Information Systems*, 46(1), 87-113. <https://doi.org/10.1007/s10115-014-0806-3>
- Raman, M., Kuppusamy, M. V., Dorasamy, M., & Nair, S. (2014). Knowledge management systems and disaster management in Malaysia: An action research approach. *Journal of Information & Knowledge Management*, 13(1), 1450003. <https://doi.org/10.1142/S0219649214500038>
- Rehage, G., Joppen, R., & Gausemeier, J. (2016). Perspective on the design of a knowledge-based system embedding linked data for process planning. *Procedia Technology*, 26, 267-276. <https://doi.org/10.1016/j.protcy.2016.08.036>
- Ritter, A., Etzioni, O., & Clark, S. (2012). Open domain event extraction from Twitter. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1104-1112. <https://doi.org/10.1145/2339530.2339704>
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010, April). Earthquake shakes Twitter users: Real-time event detection by social sensors. *Proceedings of the 19th International World Wide Web Conference*, 851-860. <https://doi.org/10.1145/1772690.1772777>
- Samuel, A., & Sharma, D. K. (2018). A novel framework for sentiment and emoticon-based clustering and indexing of tweets. *Journal of Information & Knowledge Management*, 17(2), 1850013. <https://doi.org/10.1142/S0219649218500132>
- Schulz, A., Ristoski, P., & Paulheim, H. (2013). I see a car crash: Real-time detection of small scale incidents in microblogs. In P. Cimiano, M. Fernández, V. Lopez, S. Schlobach, & J. Völker (Eds.). *The Semantic Web: ESWC 2013 Satellite Events. Lecture Notes in Computer Science* (Vol. 7955, pp. 22-33). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-41242-4_3

- Selvam, S., Balakrishnan, R., & Ramakrishnan, B. (2018). Social event detection: A systematic approach using ontology and linked open data with significance to semantic links. *The International Arab Journal of Information Technology*, 15(4), 729-738.
- Sheth, A., Thomas, C., & Mehra, P. (2010). Continuous semantics to analyze real-time data. *IEEE Internet Computing*, 14(6), 84. <https://doi.org/10.1109/MIC.2010.137>
- Silva, T., Wuwongse, V., & Sharma, H. N. (2013). Disaster mitigation and preparedness using linked open data. *Journal of Ambient Intelligence and Humanized Computing*, 4(5), 591-602. <https://doi.org/10.1007/s12652-012-0128-9>
- Wang, M., Liu, J., Wei, B., Yao, S., Zeng, H., & Shi, L. (2018). Answering why-not questions on SPARQL queries. *Knowledge and Information Systems*, 58(1), 169-208. <https://doi.org/10.1007/s10115-018-1155-4>
- Weidong, H., Jidong, Y., Jia, Z., & Danni, Z. (2012). Study on construction of emergency plan ontology model. *Information Technology Journal*, 11(4), 414. <https://doi.org/10.3923/ij.2012.414.419>
- Wirtz, A., Kron, W., Löw, P., & Steuer, M. (2014). The need for data: Natural disasters and the challenges of database management. *Natural Hazards*, 70(1), 135-157. <https://doi.org/10.1007/s11069-012-0312-4>
- Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management*, 31(1), 6-13. <https://doi.org/10.1016/j.ijinfomgt.2010.10.001>

BIOGRAPHIES



Prof. SenthilKumar Narayanasamy received his Master Degree in M.Tech IT from VIT University, Vellore and currently working as Assistant Professor in VIT University, Vellore, India. He has 10 years of teaching experience and his research areas include Semantic Web, Information Retrieval and Web Services.



Dr. Dinakaran Muruganantham received his Doctorate in Computer Science from Anna University, Chennai and Master Degree in M.Tech IT from VIT University, Vellore. He is currently working as Associate Professor in VIT University, Vellore, India. He has good teaching experience of more than 8 years. His areas of research include Information Retrieval, Networking and Web Service Management.



Atilla Elçi is full professor emeritus chairman of the Department of Electrical & Electronics Engineering at Aksaray University, Turkey. He established the Internet Technologies Research Center (2003-2009) and is founder and managing partner of IT&T Inc., Turkey (2008-2003). He was also at ITU, Switzerland, where he was chief technical advisor (1985-1997); METU, Turkey, where he was assistant chair and chair of Computer Engineering Department (1976-1985); and Purdue University, USA. His research is on web semantics, agent-based systems, robotics, machine learning, ontology, information security, and software engineering. He has published over a hundred edited books by Springer and IGI Global. He has organized IEEE Engineering Semantic Agent Systems Workshops

since 2006, Security of Information and Networks Conferences since 2007; and COMPSAC since 2005. He has B.Sc. (with Honors) in Computer/Control at METU (1970), M.Sc. and Ph.D. (5.63/6.00) in Computer Sciences at Purdue University, USA (1973, 1975).