
Theses and Dissertations

2014

Automated spatiotemporal and semantic information extraction for hazards

Wei Wang
University of Iowa

Copyright 2014 Wei Wang

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/1415>

Recommended Citation

Wang, Wei. "Automated spatiotemporal and semantic information extraction for hazards." PhD (Doctor of Philosophy) thesis, University of Iowa, 2014.
<http://ir.uiowa.edu/etd/1415>.

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Geography Commons](#)

AUTOMATED SPATIOTEMPORAL AND SEMANTIC INFORMATION
EXTRACTION FOR HAZARDS

by

Wei Wang

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Geography
in the Graduate College of
The University of Iowa

August 2014

Thesis Supervisor: Associate Professor Kathleen Stewart

Copyright by
WEI WANG
2014
All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Wei Wang

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Geography
at the August 2014 graduation

Thesis Committee:

Kathleen Stewart, Thesis Supervisor

Marc Armstrong

David Bennett

George Malanson

Padmini Srinivasan

To everyone who has supported and helped me over the years.

ACKNOWLEDGMENTS

It has been a long journey since I started my Ph.D. study at the University of Iowa. During the six years, I received many supports from my adviser, professors, my family and friends. First of all, I would like to give my sincerest thanks to my adviser Dr. Kathleen Stewart for her enormous contributions of advices and time. She brought me to this research field to start the journey. She gave me a lot of inspirations to explore the new path, and kept me on the right track. I also want to thank my committee members Dr. David Bennett, Dr. Marc Armstrong, Dr. George Malanson, and Dr. Padmini Srinivasan for their time and supports. Thanks to all professors and colleagues in the Department of Geographical and Sustainability Science at University of Iowa for their help during my study period. I would like to thank my family and friends for all their supports and love during the 6 years for my Ph.D. study. Last but not least, I want to give my best thanks to my husband. He gives me a lot of supports and encouragement throughout my study. Without his supports, I could not achieve to the destination of the journey. Thank you to all of you.

ABSTRACT

This dissertation explores three research topics related to automated spatiotemporal and semantic information extraction about hazard events from Web news reports and other social media. The dissertation makes a unique contribution of bridging geographic information science, geographic information retrieval, and natural language processing. Geographic information retrieval and natural language processing techniques are applied to extract spatiotemporal and semantic information automatically from Web documents, to retrieve information about patterns of hazard events that are not explicitly described in the texts. Chapters 2, 3 and 4 can be regarded as three standalone journal papers. The research topics covered by the three chapters are related to each other, and are presented in a sequential way. Chapter 2 begins with an investigation of methods for automatically extracting spatial and temporal information about hazards from Web news reports. A set of rules is developed to combine the spatial and temporal information contained in the reports based on how this information is presented in text in order to capture the dynamics of hazard events (e.g., changes in event locations, new events occurring) as they occur over space and time. Chapter 3 presents an approach for retrieving semantic information about hazard events using ontologies and semantic gazetteers. With this work, information on the different kinds of events (e.g., impact, response, or recovery events) can be extracted as well as information about hazard events at different levels of detail. Using the methods presented in Chapter 2 and 3, an approach for automatically extracting spatial, temporal, and semantic information from tweets is discussed in Chapter 4. Four different elements of tweets are used for assigning appropriate spatial and temporal information to hazard events in tweets. Since tweets represent shorter, but more current information about hazards and how they are impacting a local area, key information about hazards can be retrieved through extracted spatiotemporal and semantic information from tweets.

TABLE OF CONTENTS

LIST OF TABLES.....	vii
LIST OF FIGURES.....	x
CHAPTER 1 INTRODUCTION	1
1.1 Motivation	1
1.2 Research Problems	4
1.3 Major Results and Contributions.....	5
1.4 Dissertation Outline.....	8
CHAPTER 2 AUTOMATICALLY EXTRACTING SPATIOTEMPORAL INFORMATION FROM WEB TEXT DOCUMENTS.....	10
2.1 Introduction.....	10
2.2 Related Work in GIR and GIScience	12
2.3 Extracting Spatiotemporal Information from Text Documents.....	15
2.3.1 Creating spatial and temporal gazetteers.....	15
2.3.2 Parsing and combining spatiotemporal references.....	18
2.3.3 Exporting results for geocoding and geovisualization.....	24
2.4 Case Study- Tornadoes in Oklahoma, US, April 14-16, 2012.....	25
2.5 Evaluation	29
2.6 Conclusions.....	32
CHAPTER 3 USING A HAZARD ONTOLOGY FOR SEMANTIC INFORMATION RETRIEVAL.....	34
3.1 Introduction	34
3.2 Related Work	36
3.3 Constructing a Hazard Ontology.....	38
3.4 Integrating the Semantic Gazetteers with the Hazard Ontology.....	40
3.5 Spatiotemporal and Semantic Information Retrieval for Natural Hazard	42
3.6 Case Studies.....	44
3.6.1 The first case study--Tornadoes in the Midwest in April, 2012.....	46
3.6.2 The second case study-- Hurricane Sandy from October 24 to November 4 2012.....	47
3.7 Evaluation.....	55
3.8 Conclusions	58
CHAPTER 4 EXTRACTING SPATIOTEMPORAL EVENTS FROM TWEETS....	61
4.1 Introduction	61
4.2 Related Work.....	64
4.3 Processing Tweets using NLP and GIR	67

4.3.1	Crawling tweets and parsing features of the tweets.....	68
4.3.2	Extracting spatial, temporal, and semantic information from the text content of tweets.....	70
4.3.3	Rules of extracting spatial information	71
4.3.4	Rules of extracting temporal information	78
4.4	Case Study-Snowstorms in January 2014.....	80
4.4.1	Tweet data.....	81
4.4.2	Spatiotemporal clustering of tweets.....	82
4.4.3	Interpreting spatiotemporal hazard patterns based on semantics from tweets.....	84
4.4.4	Integrating extracted information from news reports with tweeted information	84
4.5	Evaluation.....	88
4.6	Discussions and Conclusions.....	91
CHAPTER 5 CONCLUSIONS AND FUTURE WORK.....		93
5.1	Discussions and conclusions.....	93
5.2	Future Work.....	96
REFERENCES		98
APPENDIX		109

LIST OF TABLES

Table

2.1 Manually annotated and automatically processed precision and recall results based on numbers of solved spatiotemporal references.....	32
2.2 Precision and recall results based on the assigning rules	32
3.1 Manually annotated and automatically processed precision and recall results based on numbers of solved spatiotemporal semantic references.....	57
4.1 Manually annotated and automatically processed precision and recall results based on numbers of solved spatiotemporal semantic references from Tweets.....	90

LIST OF FIGURES

Figure

1.2 Dissertation structure.....	8
2.1 A framework for automatically extracting spatiotemporal information from text documents to a geodatabase and visualizing as a map using GIS	16
2.2 Annotated spatial and temporal terms in GATE.....	19
2.3 Algorithm for combining extracted spatiotemporal information from text documents.....	20
2.4 Locations automatically extracted from news documents about tornadoes that struck Oklahoma, US in April 2012.....	27
2.5 Locations extracted from news reports about tornadoes in April 2012. temporal information refines the spatial pattern of locations affected by the tornadoes associating events with more days as compared to the map in Figure 2.4.....	28
3.1 A partial view of classes in the hazard ontology created using NeOn.....	39
3.2 A view of the GATE interface for Importing Ontology.....	42
3.3 The process for incorporating ontologies into spatiotemporal and semantic information retrieval for hazard events.....	43
3.4 Mapping hazard-related events from the news reports about tornadoes that struck Oklahoma, US.....	45
3.5 Semantics extracted from a news report about tornadoes that struck Oklahoma, US on April 15, 2012	47
3.6 Extracted events relating to Hurricane Sandy from Web news reports early in the storm time period	48
3.7 Semantic information retrieval for New York and New Jersey over space and time. The news reports are from Oct 30th, 2012, although events from Oct 29-Oct 31 were reported in the texts.....	51
3.8 Hazard impact.....	54

3.9 Drought visualization	58
4.1 A tweet about an event	62
4.2 Spatial information, temporal information, and semantic information from Twitter utilized in this research.....	63
4.3 A framework for processing four features of Twitter.....	67
4.4 Crawling Tweets with Twitter Streaming API.....	69
4.5 Parse the spatial, temporal and semantic content of tweets.....	72
4.6 Algorithm of assigning spatial information.....	79
4.7 An overview of the spatial distribution of the snow storm-related tweets from Jan 27-29, 2014.....	81
4.8 Kernel density clustering based on snowstorm related events extracted from tweets	83
4.9 Text terms from tweets about snow storms impact events that struck the eastern US from Jan 27-29, 2014.....	85
4.10 Clustering of hazard-events from CNN news reports and Twitter.....	87
4.11 Spatiotemporal clustering of explosion events	91

CHAPTER 1

INTRODUCTION

1.1 Motivation

Spatiotemporal information is regularly contributed to the public through online articles, daily news reports, blogs, and Twitter feeds. Because of the massive amount of spatiotemporal information provided in the form of unstructured texts (e.g., news reports, tweets), it is necessary to automatically acquire relevant information from online text sources. Information retrieval (IR) provides valuable opportunities for users to automatically obtain information from Internet search engines or digital library applications. For example, when a user types “Hurricane Sandy” into a search engine (e.g., Google), a set of web files such as documents, images or videos with URLs that match the query will be retrieved.

As IR techniques have improved over the past decade, they accommodate and exploit a broader range of information, e.g., geographic information. Geographic information retrieval (GIR) stems from the discipline of IR, and involves not only IR methods, such as indexing, searching, browsing and querying web documents, but also includes methods that exploit the geographic content of documents (Kemp et al. 2007; Jones and Purves 2008; Teitler et al. 2008; Strotgen et al. 2010; Purves and Jones 2011; Karimzadeh et al. 2013; Andrienko et al. 2013). For example, if a document refers to “a winter storm in Chicago”, the location (Chicago) described in the document help users understand where the winter storm has occurred. Natural language processing (NLP) is used to analyze the content and context of documents, and manipulate texts to perform useful tasks with a set of computational algorithms and statistical approaches (Chowdhury 2003). Combining GIR and NLP techniques, salient geographic information can be automatically extracted from large volumes of unstructured text. For example, it’s possible to extract ZIP codes, addresses, and well-known landmarks automatically from documents. While the

fields of GIR and NLP have contributed solutions that help users find information based on their interests, the possibility of automatically tracking spatiotemporal and semantic changes relating to events in text documents, and visualizing the extracted results using GIS is a new challenge (Janowicz et al. 2012; Crooks et al. 2013; Croitoru et al. 2013; Li et al. 2013; Stefanidis et al. 2013; Wang and Stewart 2013; Stock et al. 2013; Tsou and Leitner 2013; Tsou et al. 2014; Wang and Stewart 2014). Extracting spatiotemporal and semantic information from a set of Web documents enables us to build a rich representation of the geographic knowledge described in text, capturing where, when, or what events occurred (Egenhofer 2002; Jones and Purves 2008; Sankaranarayanan 2009; Larson and Shaw 2009; Joliveau et al. 2011; Chasin et al. 2014). Twitter messages are another source for spatiotemporal event information and it is designed to work as a “micro” version of blogs or news reports. One of the most important advantages of Twitter is the rapid information transmission via the Internet (Signorini et al. 2011; Lau et al. 2014). As the major social networking platform nowadays, Twitter becomes a valuable and rich revenue for mining the “real-time Web” (MacEachren et al. 2011; Pak and Paroubek 2012; Schuurman 2013; Leetaru et al. 2013; Wang 2013; Tsou et al. 2013; Tsou and Leitner 2013).

In geographic information science (GIScience), geographic dynamics refers to change or movement events with spatial and temporal characteristics, and involves an understanding of the principle functions of relevant forces and their relationships over space and time (Yuan and Stewart Hornsby 2008). Representing dynamics of geographic domains includes determining the characteristic patterns of movement of individuals or groups (Laube et al. 2007; Dodge et al. 2008); time geography analyses, that investigate patterns of people’s activities from spatial and temporal perspectives (Miller 1991; Kwan 2000; Raubal et al. 2004; Yu 2006; Shaw and Yu 2009); and modeling the movement paths or trajectories of moving objects, such as people, vehicles or natural phenomena over space and time (Dodge et al. 2008; Stewart Hornsby and Li 2009; Theruaykt and Claramunt 2013). GIScience research is a contributor to the field of GIR, and there is a growing interest in the

intersection of text analysis and mapping of geographic information (Janowicz et al. 2012; Gelernter and Mushegian 2011; Wang and Stewart 2013; Stefanidis et al. 2013; Tsou and Leitner 2013; Leetaru et al. 2013). However, modeling geographic dynamics based on information extraction from text documents needs further study, for example, additional research is needed regarding the automated extraction of spatiotemporal information from Web text documents, geographic semantic information retrieval, and the limitations of representing extracted geographic information from Twitter.

This dissertation aims to bridge the gap in GIR, NLP and GIScience, and investigates the automatic extraction of spatiotemporal and semantic information from Web texts, and the representation of the underlying relationships using GIS. In this dissertation, *spatial information* is defined as geographic locations, such as countries, states, counties, cities, coordinates, zipcodes, street names, residential addresses, schools, or airports. *Temporal information* refers to the time, such as years, months, days, or hours. *Semantic information* is associated with the meanings of different domain-related events associated with hazards (e.g., *airport closed* or *electricity shortage*) and the higher-level classes to which events belong (e.g., *hazard impact* is the super class of *airport closed* or *electricity shortage*). This research not only transfers text content to a visual representation that preserves informational characteristics from the text documents, but also offers individuals an understanding of spatial and temporal characteristics that are otherwise buried within the text. Extracted results can be visualized in a GIS environment, transferring a “textual surrogate” of text documents to a “visual surrogate”. These new map-based representations visualized using GIS provide details about dynamic change patterns and trends of world events over space-time. The research questions to be addressed through this dissertation focus broadly on:

- Improving spatiotemporal information extraction through the development of methods to automatically extract spatial and temporal details about hazard events from Web news reports;

- Using automated semantic information extraction of hazard events for supporting the understanding of hazard phenomena from different perspectives (e.g., natural hazard, hazard impact, hazard response, hazard recovery, etc.);
- Developing methods for automatically extracting spatiotemporal semantic information from Twitter messages affording the extraction of hazard event information at multiple granularities.

1.2 Research Questions

Automatically extracting spatiotemporal and semantic information from news reports and social media is challenging due to the complex contents and context of these data. To support effective spatiotemporal and semantic information extraction, three main research topics and their associated research questions are addressed in this dissertation:

- Can spatial and temporal information presented in Web news reports be extracted to retrieve a temporal ordering of extracted hazard events and correctly assign locations and times to these events? In addition to using software tools such as GATE 8.0, what additional approaches are needed so that spatial and temporal information about hazard events can be extracted from text documents and these events can be mapped using GIS?
- Can gazetteers and ontologies be integrated in order to contribute to semantic information retrieval over multiple granularities of hazard information? Can ontologies be used to provide semantic information that supports a GIR process? How does mapping semantic information associated with hazard events contribute to an understanding of event dynamics?
- Can the same methods for spatial, temporal and semantic information extraction developed for Web news reports be applied to tweets? Does it need any additional processing? What are the benefits of augmenting the results from

processing Web news reports with data extracted through processing Twitter feeds?

1.3 Major Results and Contributions

This dissertation follows a three papers format. Figure 1.1 diagrams the three main research topics in this dissertation. The topics investigate methods for automatically extracting spatiotemporal and semantic information from hazard-related Web news documents and Twitter data. For this dissertation, more than 300 web news reports related to different types of natural hazards (tornadoes, hurricanes, and blizzards) and 270,000 tweets related to winter storms, are studied as a basis for developing approaches for automatically extracting spatiotemporal and semantic information about hazards from text documents, and to represent the geographic dynamics of the extracted information about hazards using a GIS. For this dissertation, an open source software tool, General Architecture for Text Engineering (GATE 8.0) (<http://gate.ac.uk/>) is used to implement the extraction tasks and GIR. GATE provides a Java-based environment for developers to implement multilingual text processing tasks. ESRI ArcGIS 10.1 is used to map the extracted results and perform spatiotemporal analyses.

The first paper describes a framework for automatically extracting and combining spatial and temporal information from text documents in order to capture and model geographic dynamics about hazard events. Spatial and temporal references in new reports are extracted using GATE and its supported NLP processing techniques. As part of this work, spatial and temporal gazetteers are created for contributing to the text matching steps. In this work, *gazetteer* refers to a dictionary with lists of specific terms or phrases that are used to match the corresponding information from text documents. Algorithms are developed and implemented in Java to assign proper spatial and temporal references to relevant hazard event information according to the content of the Web documents. Results are exported to a local geodatabase for geocoding. The exported data is then mapped in a

GIS environment. A validation of the developed techniques is conducted and presented for discussion.

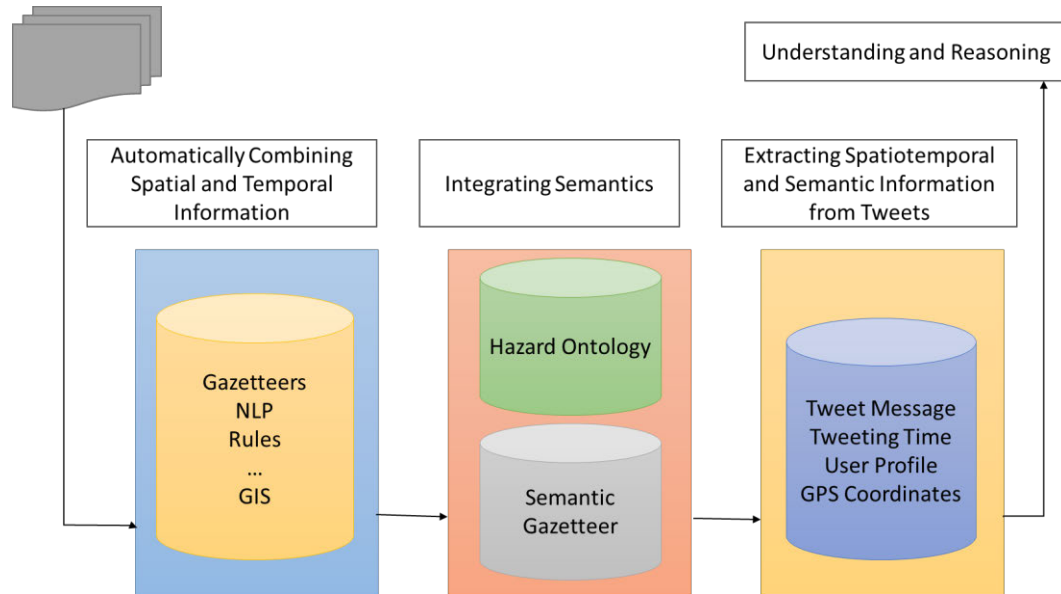


Figure 1.1 The three main research topics

In the second paper, a new semantic gazetteer is developed and the gazetteers are combined with a hazard ontology to support semantic information extraction from Web news reports about hazards. In this study, natural and human-related semantics about hazards are sourced from web texts that describe weather-related topics as well as the human side of hazards (e.g., response, impact and recovery aspects). The semantic gazetteer and the hazard ontology address the extraction of semantics associated with spatiotemporal events. In this chapter, dynamics *and* semantic information about hazard events are automatically extracted from news reports. Combining semantic information with the spatiotemporal information in a mapped representation reveals the different kinds of events and how they unfold over space and time. Making sense of the extracted semantic

information through the addition of spatiotemporal information contributes further understanding of the hazard (e.g., environmental or natural aspects vs. human hazard-related activities) to the dynamic pattern of hazard outbreaks for Web users. Maps portray the spatial and temporal characteristics associated with different events described in text documents without users having to read through each document.

In the third paper, an approach for extracting spatiotemporal and semantic information from hazard-related tweets is demonstrated. In this chapter, we investigate how the processing of hazard information extracted from news reports can be augmented with information gleaned from Twitter feeds. Four features (coordinates attached to tweets using GPS-enabled devices, user profile locations, tweet text content, and tweeting time) are parsed from Twitter to obtain spatial, temporal and semantic information. The rules developed to automatically assign appropriate spatial and temporal information to the relevant hazard events are extended here for tweets. Patterns of events from tweets are detected through aggregating spatiotemporal tweets into clusters using a kernel density approach, revealing the evolution of severe weather events monitored over time. The results provide information about the different kinds of events that occur, patterns of change, and spatiotemporal trends of hazard events. The research also shows how information processed from tweets can be combined with event information similarly extracted from Web news reports in order to capture hazard event information over multiple granularities and increase the richness of real-time hazard information especially with regard to local hazard event details.

This dissertation makes a unique contribution by bridging GIScience, GIR, and NLP and applying new methods for the extraction and visualization of spatiotemporal and semantic text information. Three hazard-related case studies in Chapter 2, 3 and 4 highlight the application of these methods for modeling event dynamics.

1.4 Dissertation Outline

The dissertation is organized as follows (Figure 1.2). In Chapter 2, methods are illustrated to automatically extract and combine spatial and temporal information from web news reports to reveal geographic dynamics (e.g., the evolution of storm events) about hazard events described in text. This topic represents the work for the first paper in this 3-paper dissertation. A case study and evaluation of methods is included in this chapter using a collection of web news reports on tornadoes that occurred over a large portion of the US Midwest, especially Oklahoma, during April 2012.

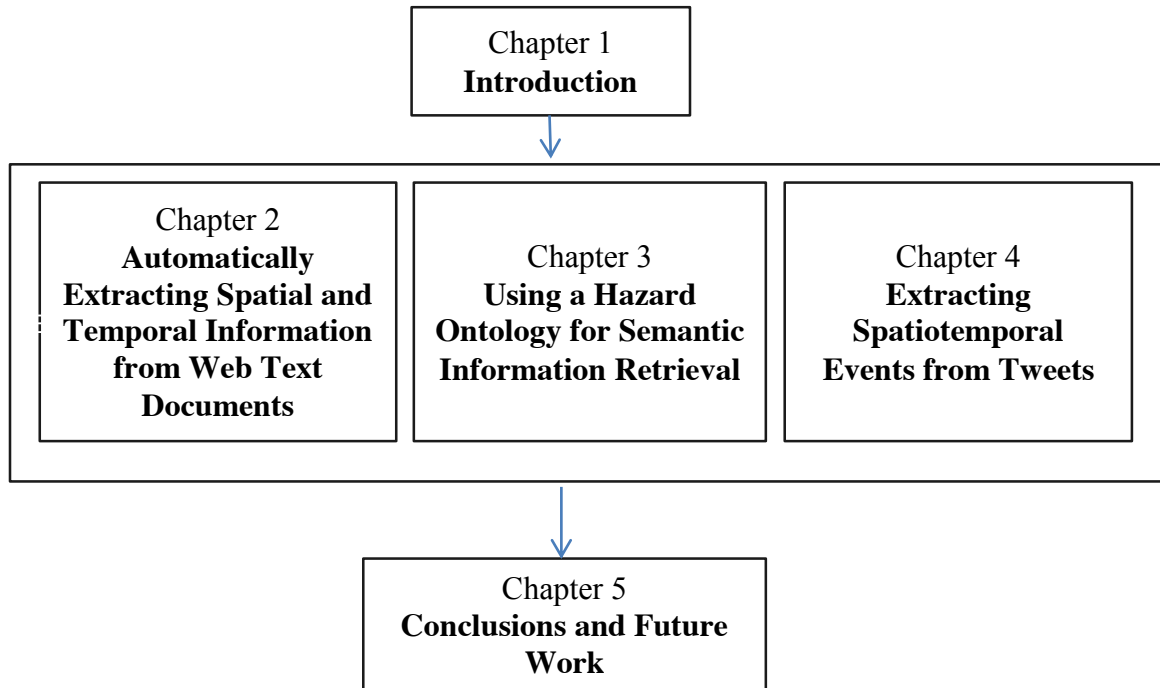


Figure 1.2 Dissertation structure

In Chapter 3, semantic information extraction and retrieval is undertaken by integrating a semantic gazetteer with a hazard ontology to represent the different kinds of hazard event information (e.g., natural or environmental and that from a human perspective), automatically and map the extracted text information at multiple granularities using the ontology. This is the subject of paper two in the dissertation. Two case studies using a collection of Web news reports about tornadoes in the Midwest in April 2012 and hurricane Sandy, from October 24-November 4, 2012 are demonstrated to examine the approach discussed in this chapter. An evaluation is conducted to assess the performance of the approach demonstrated in this chapter.

Chapter 4 addresses the third paper that focuses on automatically extracting spatiotemporal and semantic information from hazard-related tweets. A case study of winter storms that occurred in the southeastern United States during January 2014 and 270,000 tweets, is used to illustrate the steps necessary for spatiotemporal extraction from tweets. The extracted results from Twitter support the interpretation of spatiotemporal patterns of hazards (e.g., snowstorms in this case) as well as the impact of this hazard over space and time and we show how information derived from tweets can be used to augment that collected from processing Web news reports.

The final chapter in the dissertation presents a summary of the major results arising from the investigation of research questions, a discussion of these results, and topics open for future research, for example, improving and extending the gazetteers, improving the results of geocoding, and improve the quality of spatiotemporal and semantic information retrieval from tweets.

CHAPTER 2

AUTOMACTICALLY EXTRACTING SPATIOTEMPORAL INFORMATION FROM WEB TEXT DOCUMENTS

2.1 Introduction

Digital text information is widely available in the form of Web articles, news reports, blogs, Twitter feeds, and other formats. Spatial and temporal information is commonly referenced in these Web documents, especially for articles about natural hazards, as this information is related to dynamic occurrences (e.g., the track of a storm) for severe weather events and their related human activities (e.g., the movement of relief supplies in the wake of a natural disaster). Automatically extracting spatial and temporal information from a group of Web text documents and representing the extracted results using GIS enables us to build a rich representation of the geographic knowledge described in the texts. Such representations provide details about dynamic change patterns and trends of world events over space-time (Sankaranarayanan 2009; Janowicz et al. 2012; Li et al. 2013). The main objective of this chapter is to present an approach that automatically represents the spatiotemporal characterizations of hazard-related events described in Web news articles in a dynamic mapping environment. In this way, it is possible to map, for example, event sequences from documents, as they are described in text documents.

The research described in this chapter of the dissertation focuses on developing methods to automatically extract and represent geographic dynamics from hazard-related Web news reports through spatiotemporal information extraction. *Geographic dynamics*, refers to the change or movement of an event with spatial and temporal characteristics (Yuan and Stewart Hornsby 2008). Geographic dynamics is an important topic in the field of GIScience, and involves understanding the fundamental characteristics of relevant forces and their underlying relationships in space and time (Yuan and Stewart Hornsby 2008; Dodge et al. 2008; Stewart Hornsby and Li 2009; Stewart and Wang 2013; Yuan

2013; Kwan et al. 2014). In GIScience, modeling geographic information from web documents is a growing topic of research to which this dissertation makes an important contribution by focusing on handling the dynamics associated with spatiotemporal information in text (Egenhofer 2002; Jones and Purves 2008; Larson and Shaw 2009; Joliveau et al. 2011; Crooks et al. 2013; Croitoru et al. 2013; Stefanidis et al. 2013; Wang and Stewart 2013; Stock et al. 2013; Chasin et al. 2014). The novel feature of this research is to extract the spatial and temporal information related to events based on the context of the text documents, and dynamically represent the results in a GIS environment, instead of exploring these two elements in isolation.

This chapter presents methods for automatically extracting spatial and temporal information from hazard-related web news reports. The research questions addressed in this chapter are:

- Can spatial and temporal information presented in Web news reports be extracted to retrieve a temporal ordering of extracted hazard events and correctly assign locations and times to these events?
- In addition to using software tools such as GATE 8.0, what additional approaches are needed so that spatial and temporal information about hazard events can be extracted from text documents and these events can be mapped using GIS?

In this research, a framework is presented for automatically extracting spatiotemporal information from unstructured Web documents and detecting the hidden spatiotemporal patterns of hazard events using GIS. For this work, GATE 8.0 is used as the primary tool to support spatiotemporal information extraction from texts. GATE provides users a Java-based environment to extract salient information from a wide range of text documents, including web pages, RSS news feeds, and Facebook context pages. This research extends the capabilities of GATE through additional processing to show how spatial and temporal references available in text can be extracted and combined together to

inform users about the spatiotemporal pattern of hazard-related events. In this research, only text documents in English are considered and the text corpus is primarily concerned with locations in the United States. As part of this research, we employ GIS (ArcGIS 10.1) with respect to mapping extracted results from text and representing geographic dynamics for users.

2.2 Related work in GIR and GIScience fields

Geographic information such as locations, street addresses, zip codes, and x,y coordinates, are extracted from text documents and used in different applications (Mani et al. 2006; Jones and Purves 2008; Goodchild and Glennon 2010; Sakaki et al. 2010; Janowicz et al. 2012; Crooks et al. 2013). For example, geographic locations extracted from authors' affiliations can be assigned to their publications to examine a geographic flow of citations (Pan et al. 2011).

A key objective of GIR is to detect and capture location-based information from natural language text. Most GIR systems are based on detecting spatial references in text. To extract geographic information from text documents, a spatial gazetteer is a key element for data processing that affects the accuracy of extraction results. Specifically, the references in the documents are compared with the terms in the gazetteers and, if a match is found, those words or phrases from text documents are annotated by the NLP system. Numerous systems have been developed based on GIR techniques. GIPSY, a geo-referenced information processing system, supports automatic geographic indexing of text documents (Woodruff and Plaunt 1994). Using GIPSY, geographic words and phrases are identified from documents by matching terms in a document to terms in a thesaurus. The thesaurus contains place names and the names of other geographically significant objects (e.g., rivers, lakes, bioregions, animal and plant habitats, and land use types). The thesaurus plays an important role as it directly affects the accuracy of identifying locations. Two datasets serve as key components for the thesaurus in GIPSY: the geographic names

information system (GNIS) and the geographic information retrieval and analysis system (GIRAS) (Price et al. 2003). Based on these datasets, geographic references are extracted and then located on a map by using basic spatial operations including overlay and statistical methods. For example, probability weights calculations are used for determining the frequency of locations correctly displayed on the map. Since GYPSY, the idea of designing and applying gazetteers has become a standard component in most GIR systems.

Similar to GIPSY, the *web-a-where* system employs a thesaurus-based approach to identify geographic locations from web pages (Amitay et al. 2004). This system uses *geotagging*, a process of adding geographic references (e.g., place names) to different medias (e.g., web pages), to tag each place name with geographic coordinates (Scharl et al. 2008). The *web-a-where* system determines a geographical focus for each document (Amitay et al. 2004). Instead of detecting the focus of a single document, *MetaCarta* displays all spatial references in a set of documents on a map in order to produce a visualization of the locations for each document in the set (Kornai 2005). Another system, STEWARD (Spatial-Textual Extraction on the Web Aiding the Retrieval of Document), combines searching and mapping functions together, creating a system for extracting, querying, and visualizing textual references to geographic locations in unstructured text documents (Lieberman et al. 2007; Lieberman et al. 2010). STEWARD uses two NLP techniques, Part-of-Speech (POS) tagging and Name Entity Recognition (NER), to help select location-based words. This system also enables users to visualize all locations extracted from the text documents through a map interface. *NewsStand*, another GIR system, detects geographic-related information from RSS feeds using a custom-built geotagger (Teitler et al. 2008; Lieberman and Samet 2012). The extracted locations are displayed via a map viewer that dynamically displays the locations associated with news articles (Lieberman et al. 2010; Teitler et al. 2008).

In the field of GIScience, researchers are growing increasingly interested in incorporating GIR techniques into their studies (Egenhofer 2002; Purves and Clough

2006; Purves et al. 2007; Jones and Purves 2008; Derungs et al. 2012; Janowicz et al. 2012; Crooks et al. 2013; Croitoru et al. 2013; Li et al. 2013; Stefanidis et al. 2013; Stock et al. 2013; Chasin et al. 2014). In GIScience, researchers are using GIR techniques to extract spatial information from a group of text documents related to routes, and represent paths on a map (Klippel et al. 2008; Zhang et al. 2012). In this work, direction-related documents containing route descriptions from different websites are tagged for some route direction elements (*origin*, *destination*, and *route parts*). Specifically, an HTML document is converted into a document object model tree and traversed in a depth-first order. The plain text part is separated from the document and stored in a text list. Using NLP techniques, the text is parsed into sentences, in which a variety of features are tagged: basic features, surficial features, visual features, domain-specific features, and window features. These different features serve as a basis for extracting route-related sentences. Domain-specific features could be a list of nouns and noun phrases that relate to a place, such as a school or a hotel. With these features, the sentences are classified into one of four classes: *origin*, *destination*, *instruction*, and *other*. The route can then be specified based on these elements and is visualized in a map viewer (Zhang et al. 2012). One application on narrative materials traced residential life histories based on narrative materials, such as oral histories and biographies (Kwan and Ding 2008). This work combined qualitative GIS, narrative analysis, and 3D GIS-based time-geographic frameworks to provide a multimedia environment for the interpretation, analysis, and visualization of the life path for each individual. GIR techniques were used to extract spatial, temporal, action, and emotional terms from narrative materials. These extracted terms were incorporated in a geodatabase to facilitate the exploration of the relationships among individual's feelings and the locations being visited at specific times. For example, in studying the movements of Muslim women in Columbus, Ohio after September 11, 2001, movement paths modeling the women's daily activities and qualitative data about locations that were visited over time by each woman,

are represented to show changes in Muslim women's sense of safety or danger in relation to specific locations and times after 9/11. In other recent work, a context discovery application is designed for the production of geo-historical context from *RSS* feeds (Tomaszewski 2008; Tomaszewski and MacEachren 2010). In this work, based on identifying and extracting context for humanitarian crisis situations from the ReliefWeb Sudan *RSS* feed, humanitarian terms are extracted from articles describing the long-term humanitarian crisis situation in the Sudan and geo-located using GoogleEarth.

In this chapter, we show how a composition of GIR and NLP techniques can be employed in GIScience research to provide news ways for representing Web news content related to natural hazards from a spatiotemporal perspective. The methods demonstrated in this research supports spatial and temporal information extraction from unstructured data, saving the extracted locations as well as temporal ordering information about the extracted events in a geodatabase for subsequent event sequence visualization. Designing an approach for assigning appropriate spatial and temporal information to the relevant event information is an important objective for this research.

2.3 Extracting Spatiotemporal Information from Text

Documents

In this research, a framework is employed for automatically extracting spatiotemporal information from text documents (Figure 2.1) that includes several key components: 1) creating spatial and temporal gazetteers, 2) parsing and combining spatiotemporal references with event information, 3) exporting annotated results to a local geodatabase for geocoding and geovisualization. Such a methodology provides a systematic way to process spatial and temporal content along with the corresponding event information from text documents.

2.3.1 Creating spatial and temporal gazetteers

Traditionally, a gazetteer is regarded as a dictionary that contains lists of geographic

references (Goodchild and Hill 2008), and is used for extracting place names in information retrieval systems. Geographic terms or phrases in text documents are compared with locations in the gazetteer, and when a word or a phrase in the text document matches a reference in the gazetteer, the word or the phrase is annotated as the spatial information. The term ‘gazetteer’ in NLP is applied more broadly than for a geographic gazetteer. In this field, *gazetteer* refers to a dictionary with lists of specific terms or phrases (e.g., organization, facility, locations, etc.) that are used to match the corresponding information in text documents.

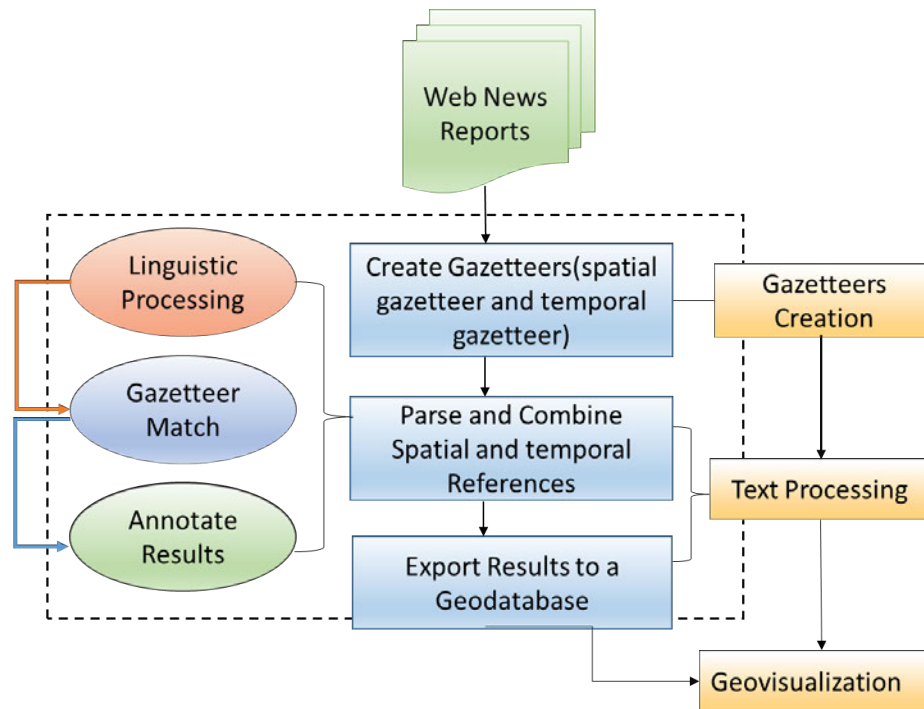


Figure 2.1 A framework for automatically extracting spatiotemporal information from text documents to a geodatabase and visualizing as a map using GIS.

In this work, different gazetteers store different kinds of vocabulary commonly found in news reports on hazards. GATE's default gazetteer supports extracting locations and dates from text documents. Most of GATE's regional references are related to general world geographical references, especially geographical locations in the UK, where GATE is developed. US geographical information is not widely covered in GATE's default gazetteer. The spatial gazetteer developed for this research extends GATE's original gazetteer by importing U.S. state abbreviations, county names for all states, and 25,150 regional places (e.g., cities, towns, villages, census designated places, airports, schools, etc.). To compare our gazetteer with GATE's, the developed spatial gazetteer detected 801 locations from 11 test documents (news reports from CNN) as compared to only 380 locations extracted using GATE's default gazetteer for the same documents. The geographic data is obtained from StreetMap Premium for ArcGIS (<http://www.esri.com/data/streetmap/>), Geonames (<http://www.geonames.org>), and the U. S. Gazetteer Files for the 2010 Census (<https://www.census.gov/geo/mapsdata/data/gazetteer2010.html>). This data is stored as a set of .lst files where spatial terms belonging to the same type are grouped into the same .lst, e.g., *city.lst*, *county.lst*, *state.lst*, *airport.lst*, etc. Currently, this gazetteer contains locations for the US and the Caribbean that supports the extraction of places relating to the hazard events in the news reports used in this research, however, the gazetteer could be expanded as necessary.

Temporal information extraction is also considered in this research where the extracted information is based on textual descriptions of time. It is common for events or happenings that appear in text to occur in a temporal order and this ordering can be exploited when extracting temporal information (Alfonseca and Manandhar 2002; Ling and Weld 2010). However, temporal expressions in text documents are often not explicit. Some temporal information is expressed as intervals, for example, ten years or two months, and some are vague, such as *last Sunday morning* and *early Sunday evening*. For this

reason, extracting only absolute temporal expressions from text documents may limit the amount of information possible relating to temporal extraction results. Similar to the spatial gazetteer, a temporal gazetteer for capturing temporal attributes, has been developed to complement GATE's built-in support (e.g., common temporal references, such as day, week, month, and year). This affords temporal processing of 150 additional references, such as *early morning* or *late Monday evening* to expand further the temporal annotation capabilities.

2.3.2 Parsing and combining spatiotemporal references

The second component of the extraction process involves parsing and combining spatial and temporal terms. Text parsing is the practice of recognizing references (e.g., spatial information and temporal information) from Web text documents with the help of NLP techniques and the gazetteers developed in the previous subsection. The general process of text parsing includes linguistic processing (in this case GATE ANNIE for the tokenizer, sentence splitter, part of speech tagger, etc.), gazetteer matching, and annotating extracted information. Specifically, as a part of linguistic processing, text documents are split into simple units (i.e., tokens), differentiating between different elements (e.g., uppercase letters, lowercase letters, mixed uppercase and lowercase letters, symbols, numbers, and punctuation). For gazetteer matching, tokens in text documents are compared with spatial references and temporal expressions in the gazetteers, and when a term or a phrase in the text document matches a reference in the gazetteers, the term or the phrase will be annotated as either spatial information or temporal information (Figure 2.2).

The process of automated spatiotemporal information extraction from Web text documents includes combining annotated spatial and temporal information according to how they are presented in text documents. Spatial and temporal data are combined to capture the salient details of spatiotemporal dynamics in articles. Designing an approach for assigning appropriate spatial information and relevant temporal information to event information is an important objective for this research. For many documents, geographical

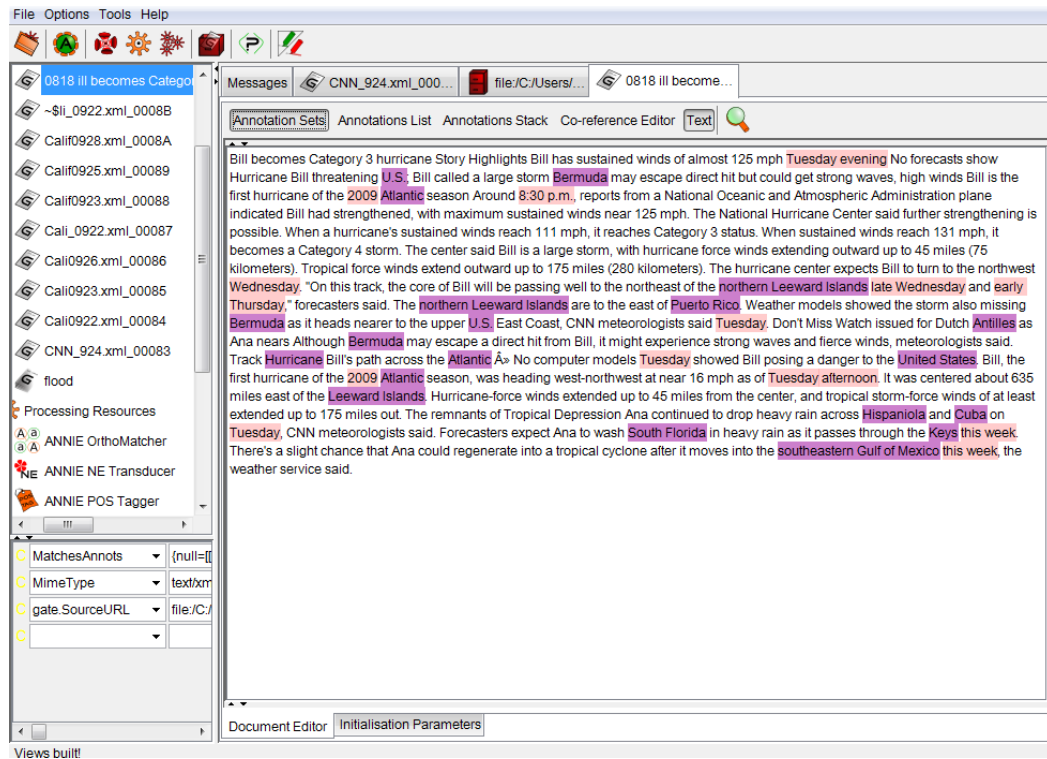


Figure 2.2 Annotated spatial and temporal terms in GATE

references are associated with temporal expressions (Strotgen et al. 2010). For example, *More than 1,000 volunteers rushed to fill sandbags Wednesday as many in Fargo tried to protect themselves from a historic flood that is expected to swamp the area.* In this example, *Fargo* is associated with the temporal expression *Wednesday*, affording both spatial and temporal details about an event of interest in a sentence. A sentence has been considered as a unit for reasoning temporal/spatial information or exploring spatiotemporal information by other researchers (King and Weld 2010; Strotgen et al. 2010; Strotgen and Gertz 2010). In this research, we follow the same pattern. Each sentence is treated as a processing unit, however, we extract the spatial and temporal information jointly with our own developed rules instead of exploring these two elements separately.

Algorithm

```

1. Input: Document  $D$ , Sentence  $E$ , Spatialterm  $S_E$ , Temporalterm  $T$ ,
   Publisheddate  $TP$ 
2. Output: Combine ( $S_E, T$ )
3. begin
4.   Parse  $D$ 
5.   for Each sentence  $E$  in  $D$  do
6.     If only  $S_E$  exist in  $D$ , then
7.       add  $TP$  to  $S_E$  combine( $S_E, TP$ )
8.     If one  $S_E$  and one  $T$  in  $E$ , then
9.       assign  $T$  to  $S_E$  combine( $S_E, T$ )
10.    If one  $S_E$  and multiple  $T$  in  $E$ , then
11.      assign each  $T$  to  $S_E$  combine( $S_E, T1$ ), combine( $S_E, T2$ )...
12.    If multiple  $S_E$  and one  $T$  in  $E$ , then
13.      assign  $T$  to each  $S_E$  combine( $S1, T$ ), combine( $S2, T$ )....
14.    If multiple  $S_E$  and multiple  $T$  in  $E$ , then
15.      Check left and right context of  $S1$ ,
16.      if a comma is found, then searching is stopped
17.      If  $T$  exists, then assign  $T$  to  $S1$ 
18.      Else jump to  $S2$ , keep checking left and right context of  $S2$ ,
19.      .....
20.    Else
21.      Jump to the next sentence
22.  Return combine( $S_E, T$ )
23. End

```

Figure 2.3 Algorithm for combining extracted spatiotemporal information from text documents

However, not every sentence will have both spatial and temporal information. Five possible cases can arise with respect to spatiotemporal information in a sentence (Stewart Hornsby and Wang 2010), and rules are implemented in Java for assigning proper temporal expressions to spatial references according to how spatial and temporal references available in each sentence (Figure 2.3).

- 1) only spatial information is present;
- 2) one spatial term and one temporal reference;
- 3) one spatial term and multiple temporal references;

- 4) multiple spatial terms and a single temporal reference;
- 5) multiple spatial and multiple temporal references.

Specifically, for each sentence, tokens on the left and right side of each spatial term are checked as a first step for assigning events to a location. The parsing of a sentence starts on the left of the spatial term (a left context), and stops when the first period in the text is reached (the end of the current sentence). Then the search process moves to the right side of the spatial term (a right context), and stops when the first period on the right side is encountered (the end of the current sentence). If a temporal expression is discovered either context, then the expression will be assigned to the spatial term according to one of the five possible cases.

1. When only spatial information is present

In some cases, only spatial information is given in a sentence with no temporal information being available, for example, *LaGuardia airport remains shut down due to flooding*. *LaGuardia airport* is extracted as a spatial location. After checking the left and right context of the spatial term (i.e., the context is the surrounding characters of the spatial term) within a sentence, it is found that no relevant temporal information is contained in the text. In this case, since no temporal expression is present here, the published date of the article can serve as the temporal information that is assigned to the spatial information, therefore, *LaGuardia airport* is assigned the document's date (e.g., 10/31/ 2012).

2. Sentences with one spatial term and one temporal reference

For documents, such as news reports that describe dynamic happenings, it is likely that sentences contain a spatial reference along with an associated unit of time. For example, *wind blows across a flooded street on October 29, 2012 in Atlantic City, New Jersey*. In this case, *Atlantic City, New Jersey* and *October 29, 2012* are parsed as spatial and temporal references using the augmented gazetteers, and a check of the left and right context finds one temporal term in the left context of *Atlantic City, New Jersey*. In this

case, *October 29, 2012* is assigned as a temporal entity to this location. In the example, *Tropical Storm Claudette is expected to make landfall in the Florida Panhandle by Monday*, the expressions *Monday* and *the Florida Panhandle* are extracted, and *Monday* is assigned to *the Florida Panhandle* following the same rationale (using the right context in this case).

3. Sentences with one spatial term and multiple temporal references

It is also possible for multiple temporal references to be present in a sentence. Temporal expressions can occur either before or after the spatial reference. All temporal expressions are assigned to the spatial reference respectively. For example, *the core of hurricane Bill will be passing well to the Leeward Islands on Late Wednesday and Early Thursday*. In this example, the spatial term *Leeward Islands* is associated with two temporal expressions *Late Wednesday* and *Early Thursday*. Both of these are assigned to *Leeward Islands* respectively. In the study, a more refined temporal modeling is provided, for example, *early Monday*, *Monday morning*, *early Monday morning* are not just assigned as *Monday*. Each of them will be mapped to a time period, which is very useful for modeling refined temporal information described in text, and capturing details of geographic dynamics.

4. Sentences with multiple spatial terms and a single temporal reference

It is also possible for more than one spatial reference to exist with only one temporal term in a sentence, for example, *the remnants of Tropical Depression Ana continued to drop heavy rain across Hispaniola and Cuba on Tuesday*. For these cases, multiple locations are linked with one temporal term. The left and right context information of the spatial terms *Hispaniola* and *Cuba* are checked in GATE in order to match the locations with a temporal reference. In this example, *Tuesday* is extracted and assigned to *Hispaniola* and *Cuba* respectively.

5. Sentences with multiple spatial and multiple temporal references

This last case is often the most complex where multiple spatial references and multiple temporal references are given in the same sentences. Here two types of cases are considered for assigning temporal expressions to spatial terms.

For the first case, punctuation is used to assist with assigning the temporal terms to spatial terms. The system checks for specific punctuation, such as commas in the left and right contexts of spatial terms. For example, *some services in Philadelphia were restored Tuesday, and Southeastern Pennsylvania was scheduled to resume service Wednesday morning, according to a SEPTA statement.* In this example, *Philadelphia* is parsed as the first spatial term. Starting from the left context, if there is a comma in the left context of *Philadelphia*, checking is stopped and the temporal expressions that have been detected are assigned to the location respectively. When the left context checking is finished, the search moves to the right context of the location term and applies the same rationale. In this example, there is no punctuation mark on the left of *Philadelphia* and no temporal information is detected, so checking moves to the right context using the same rationale. In the right context of *Philadelphia*, there is a comma and only *Tuesday* is detected as a temporal term to assign to *Philadelphia*. After finishing checking both context of the first spatial term, the search moves to the next spatial term. Therefore, *Tuesday* is assigned to *Philadelphia*, and *Wednesday morning* is assigned to the second spatial term, *Southeastern Pennsylvania*.

The second case is for sentences that do not contain any commas, for example, *hurricane appeared on track to hit Destin and Panama City Beach late Sunday or early Monday.* In this example, no punctuation marks exist except a period at the end of the sentence. *Destin* and *Panama City Beach* are identified as spatial references. The extracted temporal terms are *late Sunday* and *early Monday*. The left and right context information for each spatial reference is checked to help assign the temporal expressions appropriately. In this case, each of these temporal expressions is assigned to *Destin*. Then the second

spatial reference *Panama City Beach* is combined with every temporal term, e.g., *Panama City Beach* with *late Sunday*, *Panama City Beach* with *early Monday*.

It is also possible for text documents to present only spatial information and not contain *any* temporal information. In these cases, the document date in the headline of each news article is assigned to spatial locations mentioned in the article, providing some temporal information. When combining spatiotemporal information, additional processing is undertaken to manage temporal information. In this work, the handling of temporal expressions is extended with additional processing to standardize temporal information for mapping the temporal ordering of events. For example, the duration of events is assigned to a discrete time period, e.g., *early Thursday* is assigned a period between 5:00 am and 11:00 am on Thursday. The published date of the news reports also serves as a criterion to compare temporal expressions (e.g., Thursday, tomorrow, the next day) in the documents so that these expressions may be ordered temporally and assigned with appropriate dates. All temporal information needs to be converted to a standard time format (i.e., YYYY-MM-DD hh: mm: ss) in the geodatabase for further processing. The data is then ready for geocoding and mapping.

2.3.3 Exporting results for geocoding and geovisualization

All results are exported in a database that includes a set of tuples $\{S_{ID}, S_E, S_G, X, Y, T\}$, where (1) S_{ID} is the id number of S_E (a spatial term). GATE marks each annotated term with a specific ID number after the results are sorted. The spatial terms can be mapped based on their ID number that determines the order of positions according to where spatial terms occur in text documents; (2) S_E corresponds to the spatial term; (3) S_G describes spatial granularity of the spatial term (e.g., city, state, or county name); (4) X and Y represent x, y coordinates for the location of the term S_E ; and (5) T refers to the associated temporal information (e.g., 08/09/2012 12:00con:00). After all records are stored in ArcGIS, the locations are mapped with an x, y display function.

Using Yahoo's geocoding API, geographic coordinates are assigned to each extracted location (<http://www.gpsvisualizer.com>). In this research, for generalized (i.e., high-level) geographic references including states, counties, and regional geographic entities (e.g., southwest Florida, Riley county), geocoding selects a central point of reference. Researchers are investigating techniques for improved handling of vague places (Delboni et al. 2007; Jones et al. 2008; Vasardani et al. 2013; Chasin et al. 2014). This aspect remains a challenge for geographic information retrieval. For spatial references that are at a finer granularity, such as cities, towns, neighborhoods, or street references (e.g., City Island, New York or West Street, Manhattan), the gazetteers can be parsed, and locations geocoded on maps to show where events are occurring. The mapping of extracted spatiotemporal information is implemented using ESRI's ArcMap 10.1

2.4 Case study: tornadoes in Oklahoma, US.

April 14-16, 2012

The dataset for this case study includes Web news reports on tornadoes that impacted a large portion of the US Midwest, especially Oklahoma, during April 2012. Twenty articles about this event published from April 14 to April 17 have been obtained from <http://www.cnn.com/> and processed for their spatial and temporal content. Initially each Web news report is ordered by its published date after spatiotemporal parsing. In this way, the date of the document can provide a time frame for the phenomena being analyzed. Each news article includes a header specifying its published date that is used to distinguish documents and sort them in an order that they are published. Therefore, document date can be used to give a basic temporal order to the extracted spatial locations. However, temporal ordering of events can be improved further by extracting temporal expressions contained in the text content. The temporal detail, in addition to the extracted spatial information enriches our understanding of the hazard dynamics. Specifically, rules of assigning appropriate spatial information and temporal information to events, according to how this

information are presented in the documents, have been developed. Two approaches will be applied to the data set: 1) only use extracted spatial information plus document date information (Figure 2.4); 2) extract both spatial and temporal information and employ the combined spatiotemporal information (Figure 2.5). With comparing the results of these two approaches, we test to determine whether additional spatiotemporal changes and trends that would remain otherwise unknown are revealed through the application of the second approach.

For the first approach, each document in the dataset includes a header giving the published date of the article, and spatial information is extracted using GATE, allowing locations retrieved from the text to be ordered according to the different document dates (i.e., April 14-17). A map shows a portion of the central United States and the Great Lakes region that were impacted by the tornadoes with the highest number of events occurring in *Kansas* (Figure 2.4). Currently, for generalized geographic references in the articles, such as states, counties, and regional geographic entities, geocoding selects the centroid in the area. In this case study, these types of locations mentioned in the documents, for example, *Kansas*, *western Kansas*, *northern Oklahoma*, *Nebraska*, and *Iowa* etc., are represented on maps with a unique symbol ● to highlight that the extracted locations of these events are uncertain. For all other locations, however, the gazetteers capture the explicit spatial references that can be parsed and geocoded on maps directly (Figure 2.4).

For approach 2, spatial and temporal information is automatically extracted from text. In this case, some of the locations identified in Figure 2.4 where only the document date is used, are now associated with dates that are different to the document date, and a greater range of dates is represented. For example, *Norman, Oklahoma*, extracted from the news report on April 14, is also associated with events occurring earlier on April 13th (Figure 2.5). Some states, such as *Kansas*, *Iowa*, *Nebraska*, *Oklahoma*, reported on April 15th, are revealed to be actually associated with events that occurred on April 14th. It's also possible to detect some locations that are associated with events that will happen in

the future (e.g., events that will occur on April 15th were reported on April 14th). After extracting and processing temporal references from each document and using these in conjunction with spatial references, the locations displayed on the map represent a temporal sequence of events that is more detailed and realistic than when only document dates are used. In this way, it is possible to distinguish information about current or past movements vs. future movements by extracting both spatial and temporal information about events from the documents.

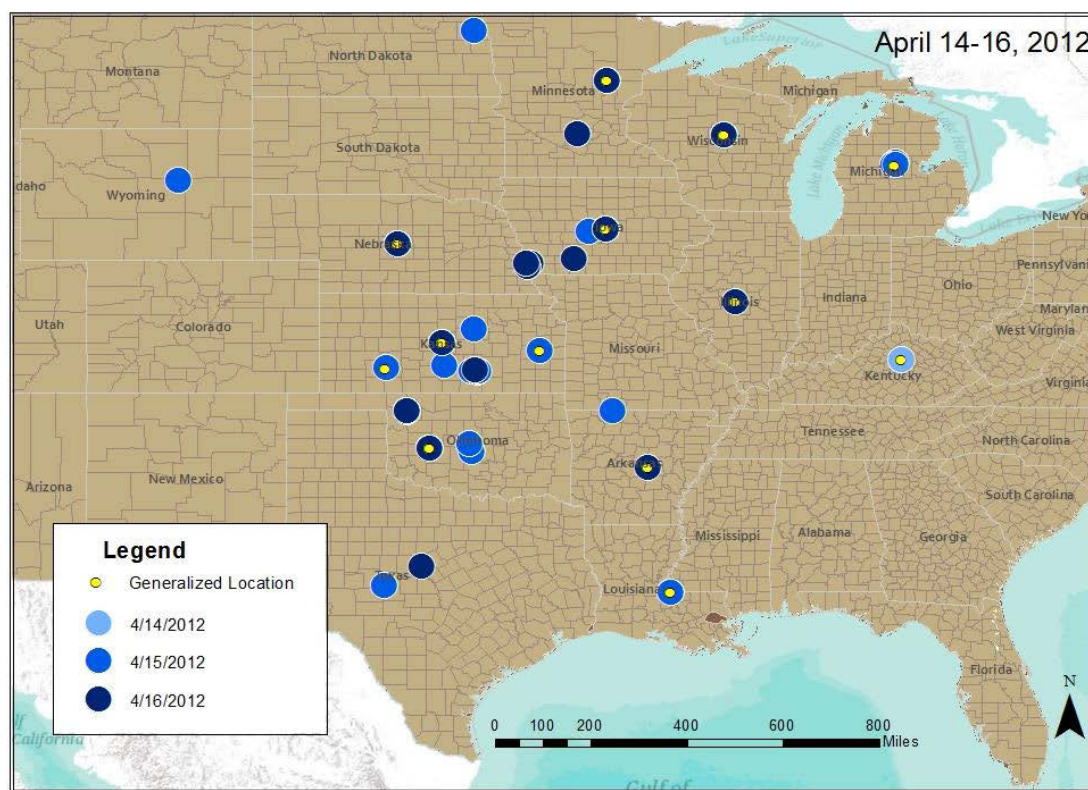


Figure2.4 Locations automatically extracted from news documents about tornadoes that struck Oklahoma, US in April 2012

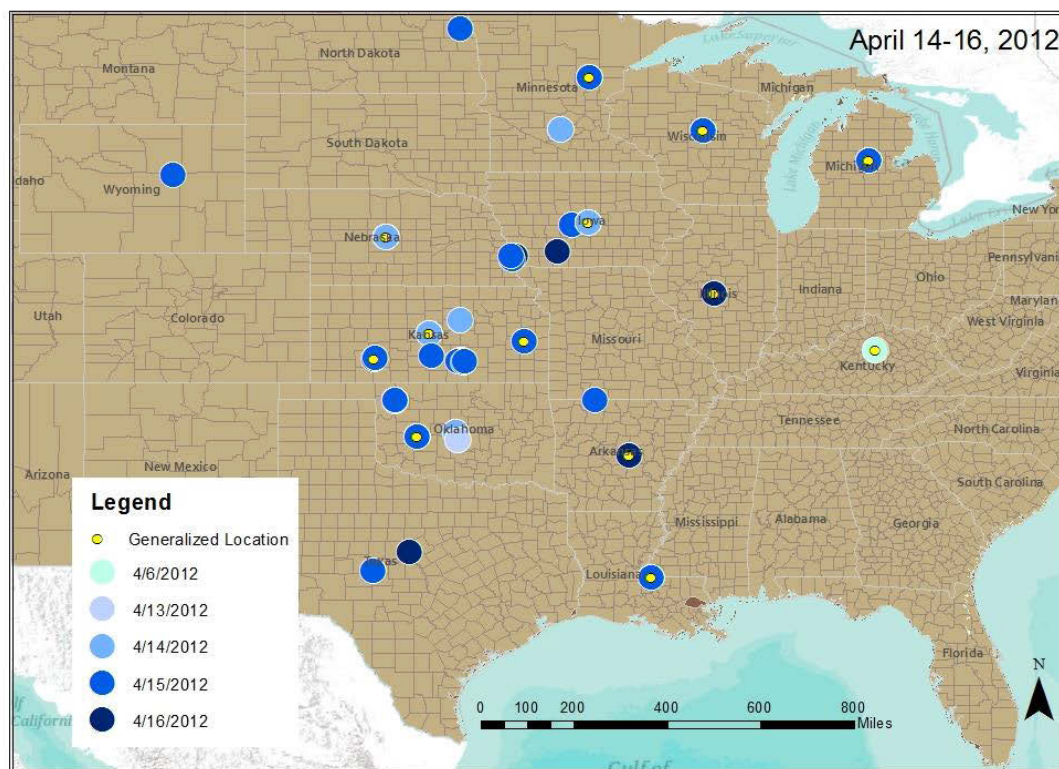


Figure 2.5 Locations extracted from news reports about tornadoes in April 2012. Temporal information refines the spatial pattern of locations affected by the tornadoes associating events with more days as compared to the map in Figure 2.4

This research demonstrates that spatial and temporal information presented in Web news reports can be extracted to retrieve a temporal ordering of extracted hazard events and correctly assign locations and times to these events. Geographic dynamics about hazards is captured through automated spatiotemporal information extraction, and the extracted results reveals possible spatiotemporal patterns that remain otherwise unknown in text document datasets. There are challenges too since different documents often contain varying degrees of spatial and temporal detail, that can lead to uncertain descriptions of change. The degree, to which generalized locations, for example, are included, is still an open research question. News articles are helpful to develop and test our approach and this

work may be especially promising for obtaining information about local dynamics where text-based descriptions may be one of the few available sources of information about events that are happening in a local area. It should be noted that our methods do not test for correctness of a news article, i.e., whether the reporting of a hazard accurately depicts the hazard. Rather, our work focuses on the processing of information in the documents revealing spatiotemporal details that can be mapped and further compared or analyzed for correctness as needed.

2.5 Evaluation

This research presents a set of steps used to extract and combine spatial and temporal information about hazard events from Web news reports to capture the dynamics of the events. To assess how well the extraction system performs, an evaluation is undertaken. In IR and NLP, the most commonly used evaluation approach is to compute *precision* and *recall* statistics over a set of evaluation data (Manning et al. 2008). *Precision* refers to how many of the extracted results are correct. The higher precision, the fewer errors are contained in the extracted results. *Recall* indicates the numbers of items that should have been detected, and records how many are effectively extracted. The highest recall value (i.e., 1) indicates the results that need to be extracted are actually all extracted (Sitter et al. 2004). In this evaluation, we extend the traditional evaluation metrics of using precision and recall to measure the performance of our approach:

STprecision = the number of correctly resolved spatiotemporal references / the total number of spatiotemporal references that the system attempts to resolve,

STrecall = the number of correctly resolved spatiotemporal references/the total number of all references.

To compute the STprecision and STrecall values, automatically processed results by the system are compared with a golden standard in order to acquire the numbers of correctly resolved spatiotemporal references, incorrectly resolved spatiotemporal

references, and missing spatiotemporal references. We recruited human subjects to provide a golden standard for this evaluation (Clark et al. 2010). Five volunteers were recruited to manually process spatial and temporal references and assign spatial and temporal references to an event. The volunteers were trained before they conducted the evaluation tasks by providing two sample news reports along with instructions of how to recognize events, and spatial and temporal information. In the two training documents, spatial and temporal information have been annotated by a human expert. The instructions provide predetermined criteria and examples for volunteers so they understand the definition of spatial and temporal information, and the assignment tasks. After training, volunteers were assigned 10 CNN news reports to process. Each volunteer manually annotated spatial and temporal terms from the evaluation data, and assigned the spatial to relevant temporal results based on the context of hazard event information given in the news reports. The results of manual annotation of spatial and temporal information, as well as the combination of spatiotemporal information were saved in csv.data by volunteers.

It is recognized that human subjects might not be consistent in their judgments of spatial and temporal references contained in the news reports, as well as how to detect spatial and temporal information and assign it to an event. Therefore, Kappa statistics are applied to the results by the human evaluators to achieve the golden standard (Viera and Garrett 2005; Manning 2008). This involves comparing the results from each volunteer. An acceptable standard for assessing the results obtained from the manual test is where either the results by all volunteers agree, or four out of five of the results agree (Viera and Garrett 2005). Results where three out of five or two out of five agree are required to be rechecked by all volunteers, and results with 0% agreement are excluded. In summary, the manually-derived spatiotemporal assignments, and the results obtained from automatically processing the text using the system, are each compared with the standard to get the precision and recall values.

Table 2.1 compares the results for precision and recall based on human performance with the performance results for the algorithms (i.e., the system). For the 10 news reports, there are a total of 126 events with associated spatiotemporal references in the golden standard. The average results processed by human subjects are shown in this table: 116 correct references, 10 incorrect references, and 8 missed references. This results in precision and recall values of 0.92 and 0.94 respectively for the manual annotation task. For the same 126 processed events with spatiotemporal references, the system performed with 108 correct references, 18 incorrect references, and 15 missed spatiotemporal references. Based on this performance, precision and recall are calculated as 0.86 and 0.88 respectively. The precision and recall values (0.86 and 0.88) achieved by our method is acceptable compared to other precision and recall values in the field of information extraction, for example, the precision (0.756) and recall (0.4135) values from GATE using ANNIE on spatial information retrieval (Clough 2005), and average evaluation results (precision 77.6 and recall 86.1) using benchmark data to process spatial and temporal information (Strotgen et al. 2010). Again, our evaluation does not test for correctness of a news article, i.e., whether the reporting of a hazard accurately depicts the hazard. Rather, our work focuses on the processing of information in the documents revealing spatiotemporal details that can be mapped and further compared or analyzed for correctness as needed.

The precision and recall values are also computed to evaluate the performance of each class of rule for assigning spatiotemporal information, i.e., the five cases discussed in this chapter (Table 2.2). For the 10 documents, the case of one spatial term and one temporal term is relatively more common than the other cases (34% of cases). The highest precision value occurred for the case where only spatial information appears in a sentence (i.e., 0.91). However, the recall value of this case is lower than the other cases (i.e., 0.86). The highest recall value occurs with the case of one spatial term and one temporal term. Compared to the other cases, the case of multiple spatial and multiple temporal terms in a

sentence has the lowest precision and recall values. This is perhaps not so surprising given that this is the most complex case. Further testing is necessary to confirm the patterns observed here, but these evaluation results suggest the methods can lead to appropriate spatiotemporal assignments.

Table 2.1 Manually annotated and automatically processed precision and recall results based solved spatiotemporal references

Spatio-Temporal References	Correct References		Incorrect References		Missed References		Precision		Recall	
	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto
126	116	108	10	18	8	15	0.92	0.86	0.94	0.88

Table 2.2 Precision and recall results based on the assigning rules

Type of Assignment	%of cases	Precision	Recall
Only spatial (information)	23%	0.91	0.86
One spatial and one temporal (information)	34%	0.90	0.95
Multi spatial and one temporal (information)	20%	0.85	0.92
One spatial and multi temporal (information)	16%	0.84	0.86
Multi spatial and multi temporal (information)	7%	0.80	0.79

2.6 Conclusions

This chapter combines principles from the field of GIR and NLP with ongoing work in the field of GIScience where there is an interest in capturing geographic dynamics. In this study, a set of steps created a framework that can be employed for extracting spatial and temporal information about hazard events from Web news reports. The main

contribution of this chapter is that the spatial and temporal information presented in Web news reports be extracted to retrieve a temporal ordering of extracted hazard events and correctly assign locations and times to these events. In this work, spatial and temporal gazetteers were created for supporting spatial and temporal references matching in GATE. Algorithms were developed to automatically assign appropriate spatial to temporal information based on how they presented in text documents. Five possible cases for how spatial and temporal information can occur in a document were identified, including the possibility that only spatial information is available in a sentence, sentences where events are associated with one spatial and one temporal reference, one spatial term and multiple temporal references, multiple spatial terms and a single temporal reference, and finally, with multiple spatial and multiple temporal references. The method of automatically extracting and combining spatiotemporal information reveals how geographic dynamics over space and time can be automatically retrieved from web news reports.

A case study based on web news reports describing tornadoes that impacted a large portion of the US Midwest during April 2012 is used to illustrate this research. Tornado events were represented using GIS based extracted temporal and spatial details about the events described in the news reports. This research shows how spatial and temporal information from Web news reports can be extracted to retrieve a temporal ordering of hazard events. The work involves designing spatial and temporal gazetteers, applying rules for assigning extracted spatial and temporal event information, and geocoding the spatial data so that spatial and temporal information about hazard events can be mapped using GIS. In addition to spatial and temporal information, semantic information about events (e.g., information about the type of event) adds rich meaning to extracted spatiotemporal information. The next chapter investigates the role of ontologies and semantic information retrieval for spatiotemporal events described in Web news reports.

CHAPTER 3

USING A HAZARD ONTOLOGY FOR SEMANTIC INFORMATION RETRIEVAL

3.1 Introduction

In the second chapter, spatial and temporal information are automatically captured from text documents to reveal the dynamics of hazards (e.g., the event sequence of storm events) described in web news reports. Semantic information, that assigns a particular meaning to spatial and temporal information, is also considered an important component for supporting the understanding of spatiotemporal patterns of hazard phenomena. In this research, semantic information in Web documents is defined as domain-related events (e.g., *airport closed* or *electricity shortage*) associated with hazards and the higher-level classes to which events belong, e.g., *hazard impact* is a more abstract class (superclass) of *airport closed* or *electricity shortage*. Making spatiotemporal information more meaningful by adding semantics (e.g., *hazard recovery* and *hazard response*) improves an understanding about the dynamic patterns of hazard outbreaks (e.g., tornado impacts with respect to time and space) for Web users. Users can directly visualize spatial and temporal trends of various events and map semantics at multiple levels of detail that are not explicitly described in text documents. While the fields of NLP and GIR have contributed solutions for helping users to find information based on their interests, the possibility of automatically tracking semantic changes relating to spatiotemporal events in Web news reports is still a challenge.

In this research, an approach is presented for automatically extracting semantic information from hazard-related Web news reports. This research investigates the role of ontologies as a key component in the process of semantic information retrieval. Ontologies deal with the nature of the phenomena, focusing on the organization of reality (Welty 2003; Kalfoglou and Schorlemmer 2003). In information science, ontologies have been applied

as a tool for knowledge management and knowledge representation. They use formal methods to represent entities, attributes, relationships, and values for a specific domain (Noy 2004; Wiegand and Garcia 2007; Sawsaa and Lu 2012). Well-developed ontologies can serve as a standard for conceptualizing and understanding domains of interest, and ontologies enable data sharing and semantic interoperability (Cruz and Xiao 2005; Schorlemmer and Kalfoglou et al. 2008; Fernandez et al. 2011; Cruz et al. 2013).

Ontologies can also play an important role in GIR, by providing a knowledge base that supports semantic understanding of text and improves search results as well as extracted information (e.g., geographic information). Associations between different semantics can be achieved by applying ontologies with classification schemes and hierarchies (Kemp et al. 2007; Jones and Purves 2008; Kontopoulos et al. 2013). Ontologies can also be applied for the disambiguation of geographic names and improving gazetteer interaction in GIR systems (Volz et al. 2007; Janowicz and Kebler 2008; Machado et al. 2011). Applying ontology in GIR applications can help to capture natural facts and related information from a human perspective and represent extracted text information using relations in the ontology. Events can also be represented in ontologies (Allen and Ferguson 1994; Worboys and Stewart Hornsby 2004; Imran et al. 2013; Lee et al. 2013).

This chapter presents methods for automatically extracting spatiotemporal and semantic information from hazard-related Web news reports. The research questions addressed in this chapter are:

- Is it possible to integrate gazetteers and ontologies in order to contribute to semantic information retrieval over multiple granularities of hazard information?
- Can ontologies be used to provide semantic information that supports a GIR process?

- How does mapping semantic information associated with hazard events contribute to an understanding of event dynamics?

For this work, we are interested in capturing the spatiotemporal patterns of hazard-related events as well as this associated semantics from texts in order to track the occurrences of natural hazards from different perspectives. A hazard-based ontology has been built to assist the semantic information retrieval process, especially with the automatic detection of semantics from news articles about hazards and represent the hidden relationships between the events over space-time using GIS. In this way, events associated with hazards or other dynamic happenings can be automatically extracted and represented and is particularly useful if a large set of documents multiple analyzed for content. The semantic information retrieval provides richness to maps through a story-telling approach using both natural and human perspectives.

3.2 Related Work

Ontologies have been imported into the GIR process to facilitate the retrieval of heterogeneous geographic information from texts, and knowledge representation and reasoning (Buscaldi et al. 2006; Saggion et al. 2007; Jones and Purves 2008; Machado et al. 2011; Kontopoulos et al. 2013; Buscaldi et al. 2014). Ontologies reflect different relations that hold among entities, for example, entities are arranged into superclasses and subclasses related by *is-a* relations in order to classify entities into subgroups, or to capture *part-whole* relationships through *part-of* relations. Ontologies have also been applied to capture events, e.g., biomedical events (Hu et al. 2011) or business events (Saggion et al. 2007; Arendarenko and Kakkonen 2012). Ontologies can refer to upper-level concepts or domain-based (i.e., lower level) concepts. An upper-level ontology is a model that captures high-level abstractions of entities in the world such as *Region*, *Event*, *Physical Object* and *Feature* (e.g., DOLCE and BUFO) (Guarino 1998). Upper ontologies benefit GIR by formalizing and integrating extracted information relating to

high-level semantics (Gangemi et al. 2002). Domain ontologies on the other hand, capture features that relate to a particular domain (e.g., medicine, indoor or outdoor) and specialize the concepts in the top-level ontology (Guarino 1998; Stewart et al. 2013). Domain ontologies are especially useful for processing and reasoning over text content, and enhance semantic information construction in information extraction systems.

Ontologies for geographic phenomena were introduced as an important component of naive geography in GIScience (Egenhofer and Mark 1995). Geographic ontology is described as a representation that consists of geospatial concepts, categories, relations, and processes, and with their interrelations at different resolutions (Mark et al. 2001). Geographic ontologies not only describe location names, but also spatial concepts including topologies, measurements, and spatiotemporal variation (Egenhofer and Mark 1995). Using ontologies, geographic domain knowledge that is relevant for conceptual modeling can be formalized to facilitate the sharing of geographic information and improved data modeling.

The SPIRIT project (SPatially-aware Information Retrieval on the InTernet) incorporates a domain ontology, a geographic ontology, footprints, and spatial indexing to create a spatial search engine in which geospatial semantic differences are distinguished (Purves et al. 2007; Jones and Purves 2008). The developed geographic ontology contains actual and alternative place names, place types, spatial footprints (i.e., geometric extent), and spatial relationships (i.e., part-of), and is applied to recognize place names and support disambiguation of place name expression in user queries. In this work, the domain ontology included non-spatial concepts related to an application, for example, tourism, useful for capturing “what” in a spatial query. For a project on marine environmental management (Kemp et al. 2007), an ontology framework is developed to assist GIR by mapping between alternative spatiotemporal classifications. Three ontologies were integrated to capture the semantics of spatial, temporal and

thematic dimensions drawn from two heterogeneous fishery databases. The thematic or domain ontology provides a hierarchical structure of concept terms and also links between concepts in one domain (i.e., the fishery) with other domains (e.g., biology).

For this research, an ontology that is suited for information extraction relating to natural hazard events especially severe storms (hurricanes, tornadoes, blizzards, etc.) was developed. The ontology is used to provide support for analyzing hazard-related events in natural language and perform semantic information retrieval. The ontology includes hazard-related topics from not only a natural perspective (e.g., geological hazards, hydrological hazards, etc.), but also from a human perspective, where the impacts of hazards (e.g., airport closings, flight cancellations, closed roads, power outages, etc.) and the human responses to the hazard (e.g., evacuations, power restored) are relevant for extracting information about hazards from text.

3.3 Constructing a Hazard Ontology

To support semantic information retrieval tasks involving events relating to major natural hazards such as severe storms, an ontology is created that includes key concepts describing both the natural perspective of hazards as well as the human perspective. An open source toolkit, NeOn (<http://neon-toolkit.org/>), is used to construct the hazard ontology. This platform was chosen for its flexible environment that supports easy importing, editing, visualizing, and exporting ontologies. The ontology is created from authoritative sources on hazards, e.g., the US Federal Emergency Management Agency (<http://www.fema.gov/>) and the US National Weather Service (<http://www.weather.gov/>), existing ontologies (e.g., OpenCyc (<http://sw.opencyc.org/>), GeoNames (<http://www.geonames.org/>), and terms extracted from the news document sets during training. The ontology is organized as a hierarchy where each class has one superclass and a set of subclasses, and classes are linked by *is-a* and *part-of* relations. The ontology

consists currently of approximately 250 classes relating to hazards and is presented in Appendix I.

At the highest level in the ontology, there are the most abstract classes, in this case, *Object* and *Happening* (Figure 3.1). Upper-level class *Object* describes high-level categories of entities that are important for natural hazards. This class has two subclasses, *Agent*, *Vehicle*, *Place* and *Time*. *Happening* refers to the dynamic processes that are common to all hazards. *Event*, is a subclass of *Happening*, and is associated with subclasses *NaturalHazard* and *MeteorologicalEvent*. Four subclasses are subsumed by class *NaturalHazard* including *GeologicHazard*, *ClimateHazard*, *HydrologicHazard*, and *WildfireHazard*. *MeteorologicalEvent* subsumes weather classes including *Precipitation*, *Storm* and *Wind*.

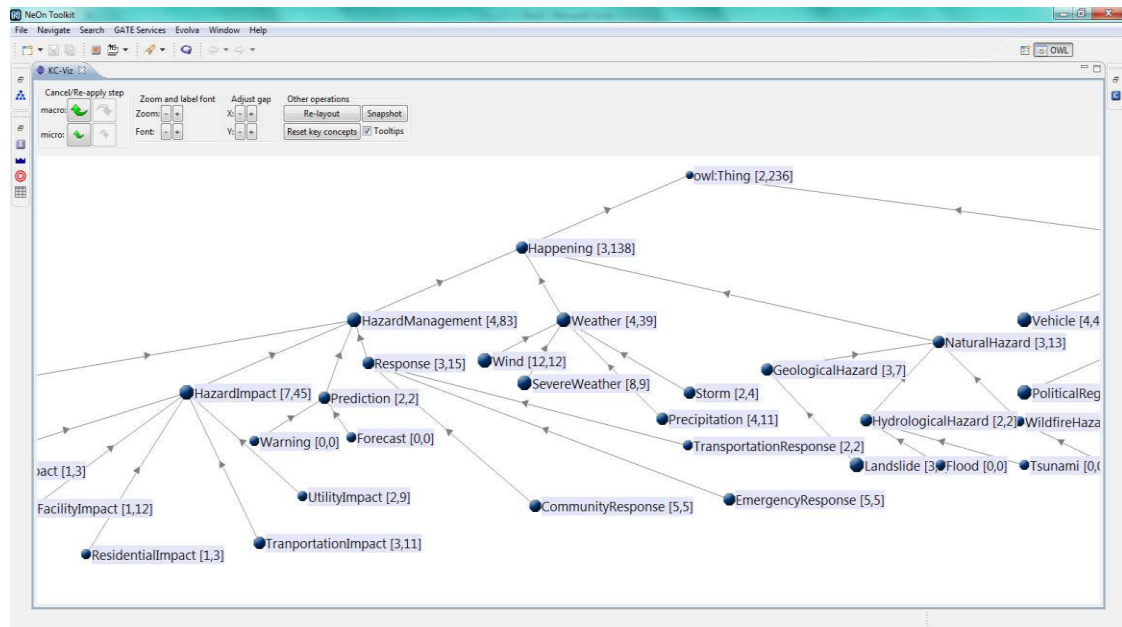


Figure 3.1 A partial view of classes in the hazard ontology created using NeOn

HazardManagement is related to class *NaturalHazard* through *part-of* relation, and captures the human aspects associated with a natural hazard. This class has four subclasses, *Prediction*, *HazardImpact*, *HazardResponse*, and *HazardRecovery*. *HazardImpact* has subclasses *CommercialImpact*, *FacilityImpact*, *IndividualImpact*, *ResidentialImpact*, and *PoliticalImpact*. For example, Hurricane Sandy occurred during the 2012 US Presidential election, and numerous electioneering events were affected (typically cancelled) as a result of the severe weather. These events are modeled as *PoliticalImpact* events. Subclasses of *HazardResponse* include *CommunityResponse*, *EmergencyResponse*, and *TransportationResponse*. *Recovery* subsumes *CommericalRecovery*, *FacilityRecovery*, *ResidentialRecovery*, *TransportationRecovery*, and *UtilityRecovery*.

3.4 Integrating the Semantic Gazetteers with the Hazard Ontology

In order to prepare Web news reports for the information retrieval process, the content of the news reports is annotated with respect to spatial, temporal and hazard-related information as described in Chapter 2 of this dissertation. For this example, the principal information extraction engine uses three different gazetteers. The information extraction task is implemented using GATE 8.0 (<http://gate.ac.uk/download/>).

In this work, three gazetteers were created to store different kinds of vocabulary as commonly found in news reports on hazards. Spatial and temporal gazetteers were created as described in Chapter 2 for capturing spatiotemporal information from texts. In this chapter we describe the process of creating a semantic gazetteer. For the semantic gazetteer, a set of 180 training documents of hazard news reports is used to collect samples of hazard-related terms or phrases based on different hazard topics. For training, there is a range to the number of articles used, however, it is not uncommon to see 70% of a data set used for training purposes (Resnik and Lin 2010). In this work, 60% of the CNN news documents on blizzards, hurricanes, flooding, tornadoes, and wildfires that will be used for

this research were processed to capture vocabulary relating to hazards. Hazard-related references were manually annotated and stored in a set of .lst files in the semantic gazetteer to correspond with different classes in the hazard ontology, e.g., *storm.lst*, *precipitation.lst*, *hydrologicalhazard.lst*, *communityresponse.lst*, *hazardimpact.lst*, etc. Each list file contains a set of related hazard events as identified in the documents during training. Rules have been implemented in the JAPE transducer (a Java Annotation Pattern Engine) in GATE to support semantic reference matching between the extracted terms and phrases from the texts with the gazetteer.

To link the terms from the semantic gazetteer with classes in the hazard ontology, the ontology is imported into GATE as a language resource, an OWLIM ontology (Figure 3.2). An ontology API in GATE that uses several plugins, such as OntoGazetteer, is used to link the terms stored in the semantic gazetteer with classes in the ontology. In this way, gazetteer terms at different granularities can be related using the ontology. To build the connection between the gazetteer and ontology classes in OntoGazetteer, a mapping file is created. The mapping file is used to connect lists in the semantic gazetteer and classes in the hazard ontology. For example, *precipitation.lst:http://gate.ac.uk/hazardontology.owl:Precipitation* links the list of different precipitation-related terms in the semantic gazetteer to the class *Precipitation* in the ontology. Given the relations (e.g., *is-a*) specified in the ontology, terms in gazetteers are associated with different granularities in the ontology. This allows for reasoning over the ontological relations. For example, assume the term *supercell* is annotated during training in one of the text documents, and that this term also exists in the ontology as a subclass of *Thunderstorm*. Given that *Thunderstorm* is a *Storm*, it can be inferred that *Supercell is-a Storm* after the processing in GATE.



Figure 3.2 A view of the GATE interface for Importing Ontology

3.5 Spatiotemporal and Semantic Information

Retrieval for Natural Hazard

Incorporating hazard-related semantics along with spatiotemporal information extraction provides a story-telling approach to mapping that involves both natural and human perspectives about hazards. The process of automated spatiotemporal and semantic information extraction from text documents uses the hazard ontology and the spatial,

temporal and semantic gazetteers to map the results of spatiotemporal information extraction, i.e., the different kinds of events (e.g., hazard events, response events and impact events) that unfold over space and time. These events, i.e., the actual phrases from the news reports as well as higher-level semantics derived from the ontology, can be visualized and represented at different spatial granularities. The resulting framework integrates text linguistic processing (in this case using GATE for the tokenizer, sentence splitter, part of speech tagger, etc.), information extraction processing using the developed gazetteers, an ontology, geocoding, and geovisualization using GIS (Figure 3.3).

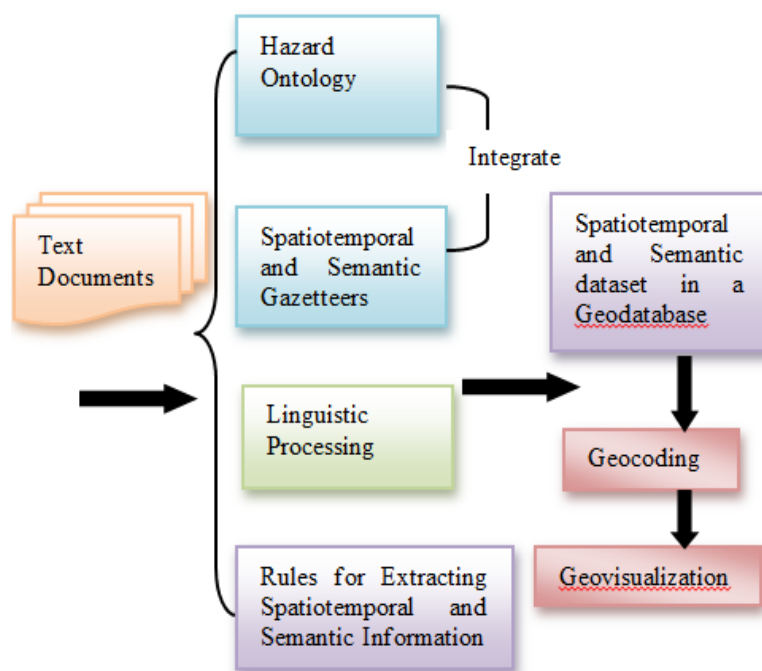


Figure 3.3 The process for incorporating ontologies into spatiotemporal and semantic information retrieval for hazard events

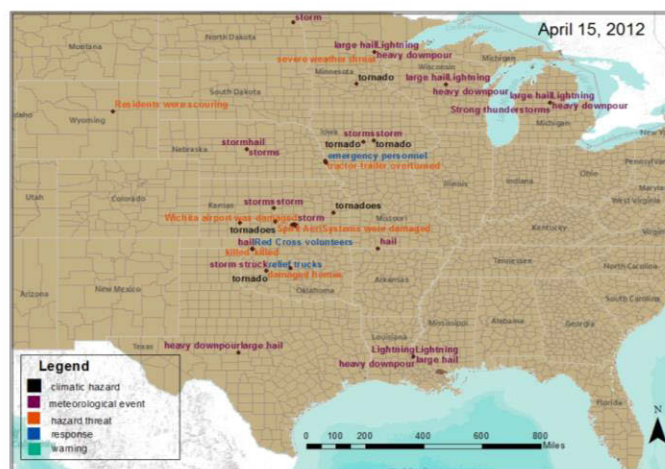
The extraction process involves automated parsing of spatial, temporal and semantic terms from text documents, processing these terms with rules developed and

implemented in GATE, and saving the extracted results to a geodatabase (Stewart Hornsby and Wang 2013). Rules include assigning proper temporal and semantic terms to locations according to the possible spatiotemporal and semantic information represented in a sentence (Wang and Stewart 2014). For example, for the sentence *More than 2,000 Westar Energy Customers in Riley County were without power Saturday evening*, ‘Saturday evening’ and ‘without power’ are annotated as temporal and event information respectively and are assigned to *Riley County* after processing. The event ‘without power’ was matched with terms in the *PowerShortage.lst* file in the semantic gazetteer. In the ontology, *PowerShortage* is a subclass of *UtilityImpact*, and *UtilityImpact* is a subclass of *HazardImpact*. In this case, ‘without power’ is associated with the higher-level class *HazardImpact* through linking the semantic gazetteer and the hazard ontology. ‘Saturday evening’, ‘Riley County’, ‘without power’ and *HazardImpact* are processed as a set (along with all other extracted terms) to a geodatabase.

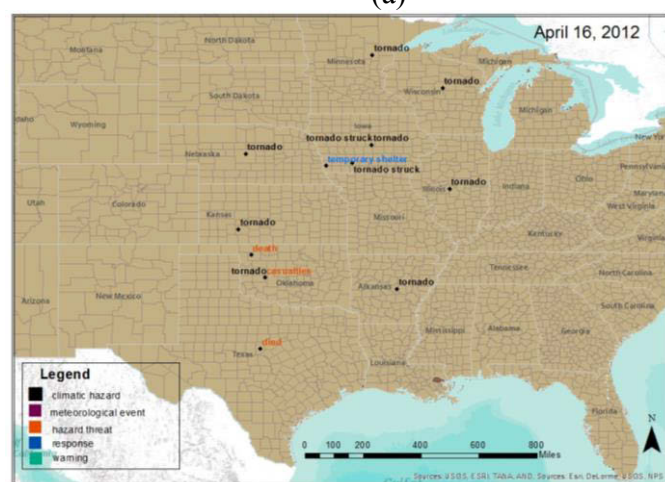
Using these methods, spatial, temporal and semantic information about events are automatically extracted from Web text documents. Ontologies bolster semantic processing by linking annotated terms from the news articles to the different kinds of semantics about hazard events that they represent, and model the events at different granularities. In this way, event semantics (e.g., hazard impact events, response events, or recovery-related events) can be represented on a map. This affords important opportunities for human-environment applications as the spatiotemporal evolution of a set of events can be tracked and new insights revealed about the dynamics and different kinds of reported events in document collections that might otherwise be unknown.

3.6 Case Studies

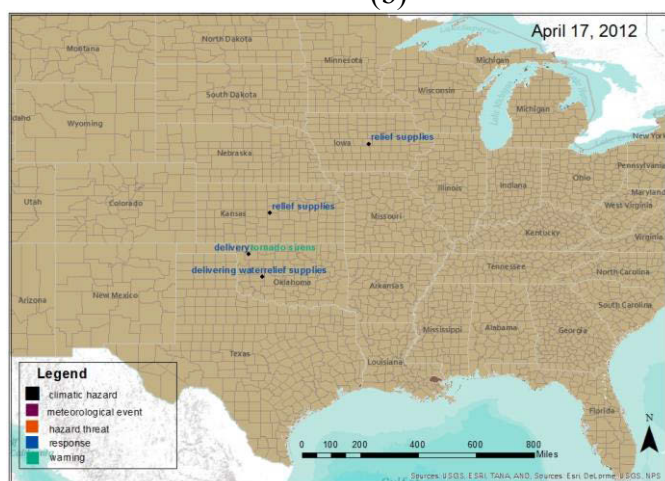
Two case studies are presented to demonstrate the methods developed in this chapter for extracting semantic information about hazard events over space-time. The



(a)



(b)



(c)

Figure 3.4 Mapping hazard-related events from the news reports about tornadoes that struck Oklahoma, US (a). April 15, 2012 (b). April 16, 2012 (c). April 17, 2012

evolution of reported events can be tracked, visualized and analyzed on maps using this approach for semantic information retrieval.

3.6.1 The first case study--tornadoes in the Midwest in April, 2012

In this first case study, the data set applied in Chapter 2 about tornadoes impacted the Midwest US in April 2012 is used. After processing, the text terms are mapped to represent storm-related events (Figure 3.4). The spatiotemporal pattern of terms is used to understand more about the dynamic tornado-related events across space. The map reveals that *large hail*, *lightning*, *heavy downpour*, and *strong thunderstorm* were associated with parts of *Texas*, *Louisiana*, *Wisconsin*, *Michigan* and *Minnesota*. *Facilities* were impacted by *tornadoes*. For example, *Wichita airport* was damaged in *Kansas*. In addition, *relief activities* were triggered in *Nebraska*, *Iowa*, *Kansas*, and *Oklahoma*. For example, trucks were being sent out for rescue in *Oklahoma*. It is possible to follow the sequence of these different elements of a hazard (e.g., *hurricane*, *response*, *recovery*) automatically based on analysis of text. Making sense of spatiotemporal information through extracted semantic information contributes additional understanding to the dynamic pattern of hazard outbreaks such as *tornadoes* for users. Event-related terms extracted from multiple documents are represented on maps, and the dynamics of events are represented in a temporal order. Users can directly visualize spatial and temporal characteristics associated with different events described in text documents without going through each document.

The results can also be presented at a higher level of abstraction (Figure 3.5). In this case, terms are presented at the granularity of super classes based on the hazard ontology. For example, *lightning*, *downpour*, *hail*, and *storm* are generalized to the class: *MeteorologicalEvent* while *twister* and *tornado* are represented at the higher level, *ClimaticHazard*. The map shows how the five classes vary spatially as per the report on April 15th, 2012. Based on this representation and for this document date, it is possible to see which areas are still experiencing *storms* (i.e., *MeteorologicalEvent*) and where

Response is beginning to occur. Event-related terms can be automatically generalized to upper-level classes using *is-a* relations, and these events can be represented on maps using upper-level abstractions. Events can be summarized and represented at different granularities to derive spatiotemporal patterns on maps.

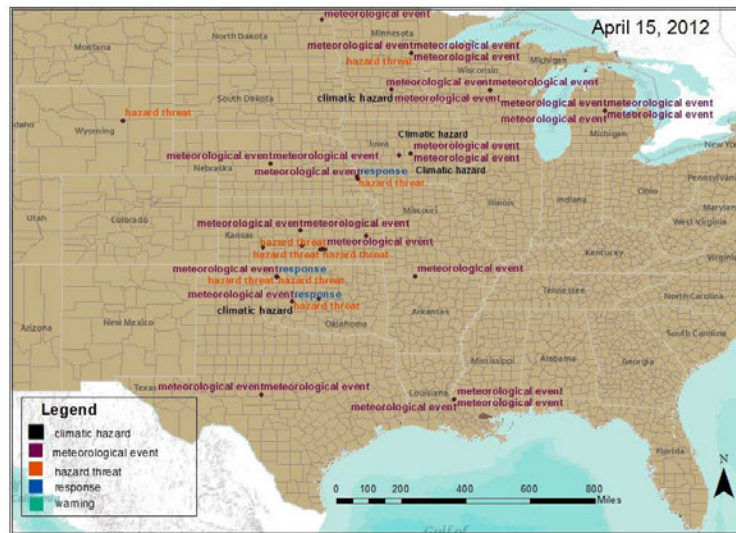
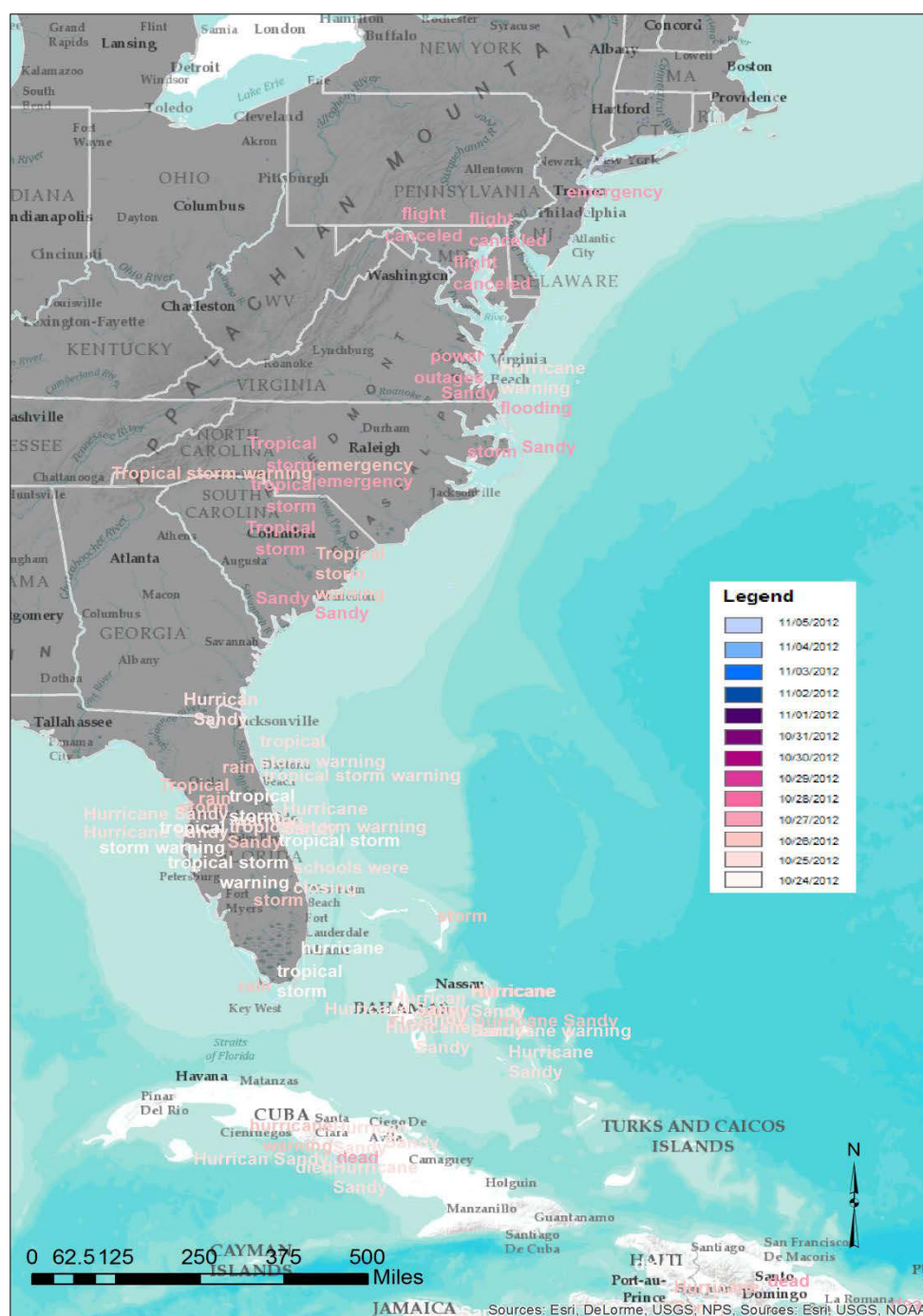


Figure 3.5 Semantics extracted from a news report about tornadoes that struck Oklahoma, US on April 15, 2012

3.6.2 The second case study-- hurricane Sandy from October 24 to November 4 2012

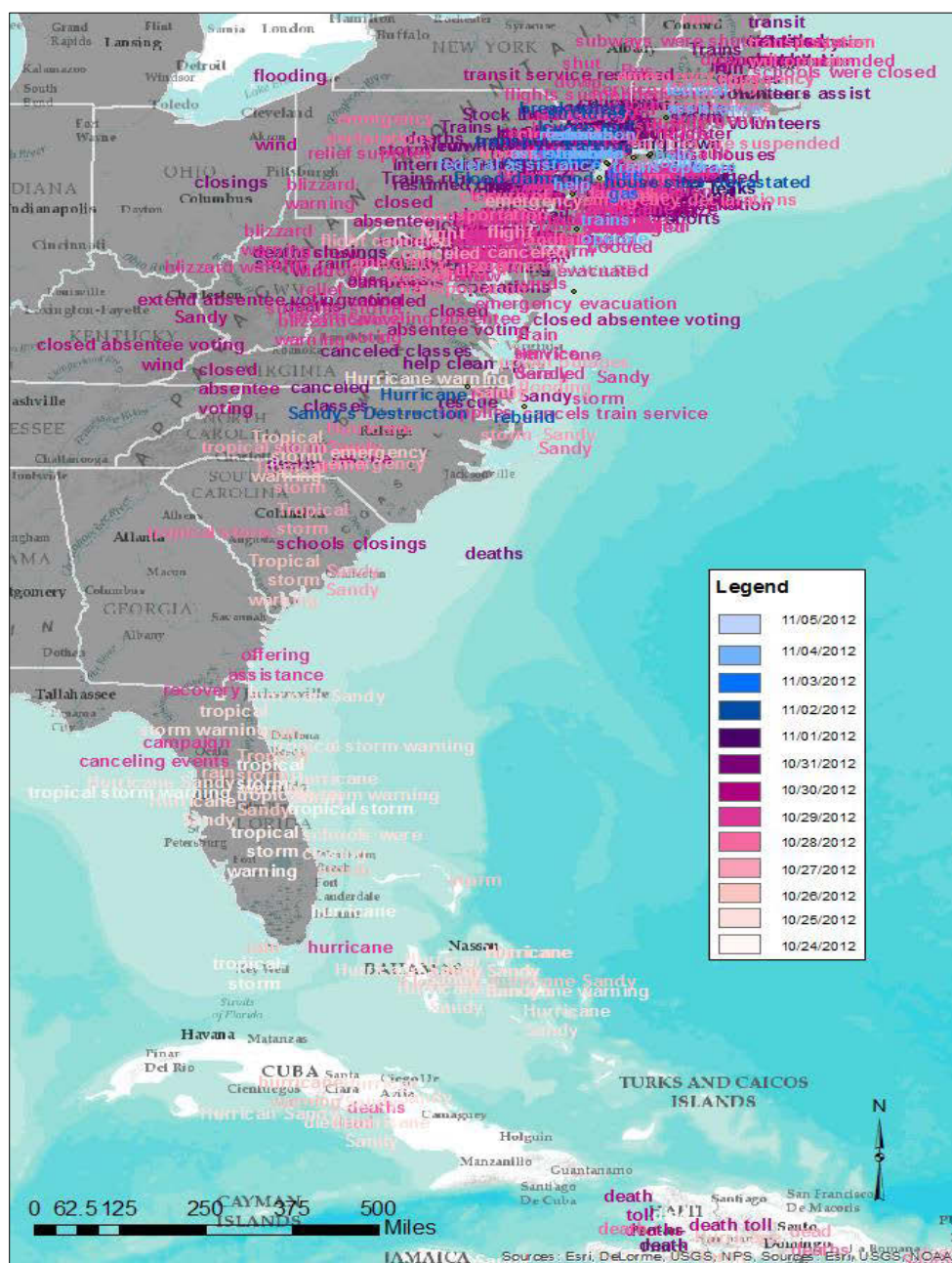
The second case study to be discussed is based on a set of 50 CNN news reports collected from October 24-November 4, 2012 about Hurricane Sandy, a very severe weather event that hit the east coast of the US. Each document is analyzed for past, present, and future information about hazard events. The range of dates associated with the events described in the article frequently goes beyond the document date for each report, as articles routinely referred to both past and future happenings.



(a)

Figure 3.6 Extracted events relating to Hurricane Sandy from Web news reports early in the storm time period (a) Oct 24-Oct 27, 2012 (b) Oct 24-Nov 04, 2012

Figure 3.6 Continued



(b)

As a result of text processing, early reports (Oct 24 to Oct 27) show described events located as far south as the *western Caribbean Sea, Jamaica, Cuba, and the Bahamas* (Figure 3.6a). The associated temporal information is represented using graduated colors.

Most of the earliest events reported in the tracked period, were located in Jamaica, Cuba, Bahamas, and Florida. By October 27th, events had moved north along the coast to locations, including *Charleston, South Carolina, Cape Hatteras, North Carolina, Mount Airy, and Maryland*. After processing all 50 of the Web news reports, mapped events generally shifted from south to north as far up as *Maine* in the northeast (Figure 3.6b). Inland events were also reported, for example, events in *West Virginia, Ohio and Pennsylvania*. Event-related terms extracted from all 50 news reports are represented in Figure 3.6b, and the temporal pattern of events is captured using graduated colors to represent temporal order, light pink–dark red–dark blue–light blue where light pink refers to the earliest dates of extracted event information, October 24, 2012, and light blue is associated with the latest date extracted, November 5, 2012. This map illustrates the focus of reporting on the highly populated regions of New York and New Jersey around October 30th, 2012. The text extraction process shows the transition of the storm from *tropical storm* to *hurricane* in the afternoon of Oct24th.

With the addition of ontology-based semantic processing, it is possible to extract and retrieve more information about the kinds of events occurring during the hurricane. Applying the ontology makes it possible to represent the extracted texts according to the different classes of events. An analysis based on a subset of 12 news articles published on October 30th, a peak day during the hurricane Sandy, shows events classified as *HazardImpact, HazardResponse* and *HazardRecovery* (Figure 3.7). For example, *power failed, flights cancelled, and lost homes* are specializations of the class *HazardImpact* while

emergency evacuation and *firefighters battled blaze* activities signify aspects of *HazardResponse*. The results of the analysis capture the fact that *power outages* were a key event for facilities impacted by Hurricane Sandy as reported in *Manhattan*, *Queens*, and *Staten Island* in *New York*, and *Newark* in *New Jersey*. *Response*-related events were

triggered in *Breezy Point*, *Brooklyn*, and *Staten Island* in *New York*, with *emergency evacuation* in *Moonachie*, *Queens*, and *Roosevelt* in *New York*. Events associated with *Recovery* were reported in *Manhattan* and *Syosset*. To understand the spatiotemporal pattern of classified events with respect to the population density of the U.S. in 2012 obtained from ESRI's 2012 Updated Demographics¹, the results are mapped over space and time against high population areas. The extracted hazard-related events are clustered in major cities with high population density, such as *Manhattan*, *Brooklyn*, and *Queens* in *New York*. Based on the results, more spatiotemporal and semantic queries can be conducted to detect inherent facts that cannot be directly retrieved from the texts. For example, which locations in *New York City* did Sandy impact from the *noon of Oct 29th* until the *morning of the 30th*? The spatial restriction for this query is *New York City*, while the semantic restriction is *HazardImpact*. The temporal restriction, *Oct29th afternoon to Oct 30th morning* is between 10/29/2012 12:00pm-10/30/2012 8:00am. The results returned that satisfy these three conditions include the boroughs of *Queens* and *Manhattan* in *New York City*.

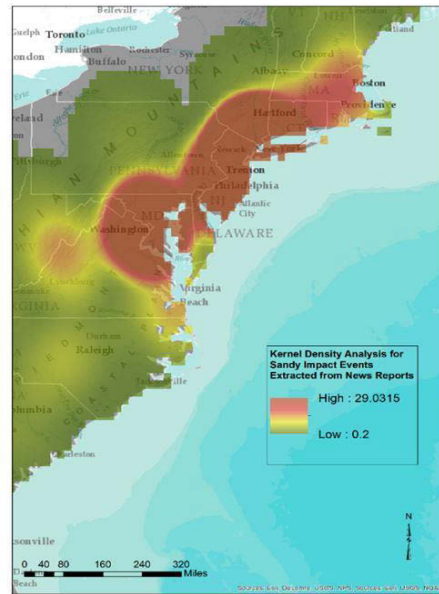
Hazard impact events extracted from the 50 news reports are also mapped using kernel density analysis in order to show the spatial distribution pattern of hazard impact events (Figure 3.8a). Kernel density is a way of estimating the intensity of events by generating a smooth surface using a quadratic kernel function (Silverman 1986; Li et al. 2013). Two parameters are used in the kernel density analysis: kernel search radius (bandwidth) to calculate density and cell size for the output raster data. The kernel search radius was 55km given the shorter of the width or height of the results extent in the output spatial reference divided by 30 (Silverman 1986), and the cell size was 6 km given the shorter of the width or height of the results extent in the output spatial reference, divided by 250 (Silverman 1986). The kernel search radius of 55km is to avoid creating a map that

¹ <http://www.arcgis.com/home/item.html?id=a18f489521ba4a589762628893be0c13>

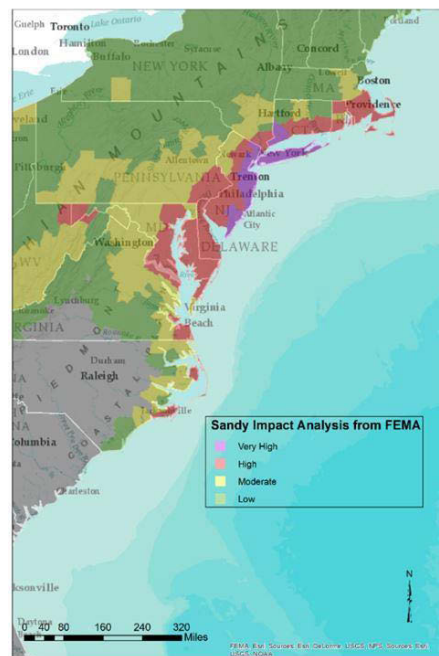
is too smooth or too fuzzy to interpret. The cell size of 6 km was used to show fine detail. The areas with high density of hazard impact events are represented with the darkest hue of red, such as the east coast of *New York, New Jersey, Philadelphia, Virginia, Maryland, Baltimore, and Washington*. As shown in Figure 3.8b, Hurricane Sandy impact data obtained from a FEMA Modeling Task Force² is mapped for the areas impacted by Hurricane Sandy with four different levels (i.e., *very high, high, moderate, and low*). The similarity of the spatial patterns between the map with the extracted *HazardImpact* events and the impact analysis map from FEMA shows our extracted results about hazard impact are promising. However, there are some places on both figures, such as Washing DC, associated with different impact levels. One of the possible reasons is that a large amount of human events related to hazard impact, such as school closed, airline canceled, power outage, road closed, etc. are associated with these area (might not be captured by FEMA). For Figure 3.8b, the analysis is based on a composite of surge, wind, precipitation and snow data from FEMA to assess the impact. The event information retrieved from text documents can be used in combination with other sources of data (e.g., FEMA impact data) to enhance the understanding of hazard events, for example, to estimate the number of different events reported in association with Hurricane Sandy. In addition, integrating data from different sources also allows us to discover additional information, such as *find all recovery events in the very high impact areas (FEMA) from the early morning of Oct 30th to late afternoon on November 2nd*.

In this research, two case studies demonstrated above illustrate the work for automated semantic and spatiotemporal information extraction through integrating a hazard ontology and a semantic gazetteer. Applying this approach, hazard-related semantics are automatically extracted from text documents, and can be represented on

² <http://fema.maps.-arcgis.com/home/webmap/viewer.html?webmap=307dd522499d4a44a33d7296a5da5ea0>



(a)



(b)

Figure 3.8 Hazard impact (a) Kernel density analysis for hazard impact-related events extracted from 50 CNN news reports; (b) Hurricane Sandy impact analysis (11/8/2012-4/18/2013) from FEMA Modeling Task Force (MOTF)

maps with events described at multiple granularities to show their spatiotemporal patterns. Incorporating semantics in spatiotemporal information extraction also provides richness to maps using a “story-telling” perspective from both natural and human angles.

3.7 Evaluation

The approaches demonstrated in this chapter show how spatial, temporal and semantic information can be extracted from Web documents to represent the dynamics of hazard events. An evaluation has been undertaken to test the automatic association of extracted events with the appropriate location, time and semantic. The evaluation was conducted using 10% of the Hurricane Sandy data set (10 news reports). A criterion for the evaluation of spatiotemporal semantic extraction of events is extended based on adapting the precision and recall evaluation metrics in Chapter 2.

- STSPrecision = the number of correctly resolved spatiotemporal semantic references / the number of spatiotemporal semantic references that the system or users attempt to resolve;
- STSRecall = the number of correctly resolved spatiotemporal semantic references / the number of all such references.

For this evaluation, to compute the STSPrecision and STSrecall values, automatically processed results by the system are compared with a golden standard in order to acquire the numbers of correctly resolved spatiotemporal semantic references, incorrectly resolved spatiotemporal semantic references, and missing spatiotemporal semantic references. Human subjects were recruited to provide a golden standard for this evaluation using kappa statistics (Manning 2008; Clark et al. 2010). For human evaluators, five volunteers were used to manually process the evaluation data. Each volunteer was trained to process spatiotemporal and semantic information from text documents by providing two sample news reports with instructions and examples. Each volunteer manually annotated spatial, temporal, and semantic terms from the 10 news reports, and

assigned the annotated terms to events based on the context in the text documents. The results are expressed as a set of vectors of the combination of spatial, temporal, and semantic information (i.e., a set of {spatial, temporal, semantic} vectors) stored in a .csv database file. To obtain the golden standard, the results from each volunteer were compared. In this evaluation, the golden standard consists of the results with at least 80% agreement (4 out of five volunteers). Results that are lower than 80% agreement are required to be rechecked and discussed with the tester, to decide whether they should be included or excluded from the golden standard. Results with 0% agreement are excluded. Manually derived spatiotemporal semantic information sets, and the results obtained from automatically processing the text are each compared with the golden standard. The number of correct references, incorrect references, and missed references for the users and the system are determined.

The results for precision and recall based on human performance and the performance results for the algorithms (Table 3.1). These results are derived according to the golden standard (113 sets of spatial, temporal and semantic information). For the human evaluation, there were 103 correct sets of references, 10 incorrect sets of references, and 10 missed references in average. This precision and recall values are calculated as 0.91 and 0.91 respectively. The system performed with 94 correct references, 19 incorrect references, and 21 missed references. Based on this performance, precision and recall are calculated as 0.83 and 0.82 respectively. This evaluation is based on the sets of spatiotemporal and semantic references, and the results for system performance are acceptable (0.83 for precision and 0.82 for recall).

In this evaluation, the quality of extracted results was also checked. The geographic locations (i.e., 648 for the total number) can be grouped into two different classes: explicit locations (i.e., places that are equal or finer in spatial granularity as *cities* and can be geocoded to potentially an address location) and generalized locations (e.g., Southern New

Table 3.1 Manually annotated and automatically processed precision and recall results based on numbers of solved spatiotemporal semantic references

Spatio-Temporal Semantic References	Correct References		Incorrect References		Missed References		Precision		Recall	
	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto
113	103	94	10	19	10	21	0.91	0.83	0.94	0.82

Jersey or East Coast, where a centroid location is used). The results contain 262 explicit locations and 386 generalized locations. It is worth noting that for the case study, the analysis was undertaken with CNN articles. It is probably the case that detailed geographic locations (e.g., explicit street names or addresses) are less likely to be reported in these news reports than generalized locations. Moving to a different, regional or local news source would be expected increase the relative amounts of fine-grained location information.

To further evaluate the approach, the methods for spatiotemporal semantic extraction were applied to a text report of U.S. seasonal drought assessment from June to August, 2013³. The text document was obtained from the Climate Prediction Center, National Weather Service (<http://www.cpc.ncep.noaa.gov>). Spatiotemporal information and drought-related events were extracted from the report and then geocoded using the methods described above. A point density analysis was conducted based on the two main classes in the ontology that are associated with the geocoded locations (Figure 3.9a): drought and non-drought. The drought area (i.e., pink area in Figure 3.9a) is mostly concentrated on the west side of the US while the non_drought area is primarily located on the east side. In addition, a U.S. seasonal drought forecast map from May to August, 2013,

3

http://www.cpc.ncep.noaa.gov/products/expert_assessment/sdo_archive/2013/sdo_jja13-rev_text.shtml

Web news reports. The extracted results are mapped using ArcGIS to represent the spatiotemporal patterns of events. Events can be categorized into different classes according to their semantic properties and formalized in an ontology. Ontological relations allow us to model these event types at different granularities. This affords important opportunities for human-environment applications as semantic retrieval affords a more detailed understanding of the nature of reported events, and combining this with a spatiotemporal perspective offers important new insights about the dynamics of reported events in document collections. Incorporating semantics in spatiotemporal information retrieval provides insight into facts that cannot be directly retrieved from texts, for example, what kinds of events were reported over a certain time window or in a certain region. The approach has been applied to two case studies of 20 CNN news reports on Tornadoes in the Midwest in April 2012 and 50 CNN news reports on Hurricane Sandy from October to early November 2012. The spatial and temporal characteristics associated with the hazard-related events described in texts can be directly visualized without going through each individual document. The results also provide new details about the kinds of events occurring during the hazard, patterns of change, and spatiotemporal trends for hazard events as well as an understanding of events at multiple spatial and temporal granularities.

The approach presented in this chapter shows the series of steps that can be followed to integrate gazetteers and ontologies in order to contribute to semantic information retrieval of hazard information. Ontologies developed in this research provides semantic information about natural hazards that supports a GIR process. Mapping semantic information associated with hazard events contributes to an understanding of event dynamics from different perspectives (human-related activities vs natural phenomena).

Online news articles have been useful to develop and test this research and this approach may be especially promising for obtaining information about local dynamics where text-based descriptions may be one of the few available sources of information about

events that are happening in a local area. The next chapter develops an extension of the research applied to twitter feeds that adds real-time spatiotemporal semantic information.

CHAPTER 4

EXTRACTING SPATIOTEMPORAL EVENTS FROM TWEETS

4.1 Introduction

Twitter, one of the most prevalent social networking and micro-blogging services, enables 140 maximum characters for each tweet and allows more than 250 million users to send out or share real-time events happening around the world every day (Ozdikis et al. 2013). Twitter is designed to work as a “micro” version of blogs or news reports. One of the most important advantages of Twitter is the rapid information transmission via the Internet (Signorini et al. 2011; Lau et al. 2014). Research results indicate that outbreak news is often disseminated on Twitter first before being reported by public media (Kaplan and Haenlein 2011). As the major social networking platform nowadays, Twitter becomes a valuable and rich revenue for mining the “real-time Web” (MacEachren et al. 2011; Pak and Paroubek 2012; Schuurman 2013; Leetaru et al. 2013; Wang 2013; Tsou and Leitner 2013).

News, especially about natural hazards (e.g., Hurricane Sandy in October 2012, blizzards of February 2013 in the eastern US, or dust storms in Utah in March 2014) tweeted by individual users or official agencies (e.g., FEMA) draws immediate attention to a great number of web users around the world. Unexpected events, particularly hazard-related events, often do not provide people with enough time to prepare. Quick response and rescue plays a major role in preventing even more serious hazard impacts. To this end, the real-time information spread by Twitter is critical for many applications, including the hazard response (Sakaki et al. 2010; Crooks et al. 2013). Real-time tweets can be used to extract not only temporal information, but also the spatial and semantic information about hazards for better understanding of hazard-related events. More importantly, hazard responses can be organized more efficiently using the extracted information (Goodchild and Glennon 2010, Sakaki et al. 2010; Vieweg et al. 2010; Crooks et al. 2013).



Figure 4.1 A tweet about an event

Tweets commonly portray spatial and temporal information about events (Figure 4.1). The spatial, temporal, and semantic data in tweets are helpful for event pattern detection and spatiotemporal queries. In this chapter, an approach for automatically extracting hazard-related spatial, temporal and semantic information from tweets is presented (Figure 4.2). Spatial information is detected by analyzing three *input signatures* of a tweet: 1) attached geographic coordinates (longitude, latitude) from GPS-enabled mobile phones or tablets; 2) user profile locations, and 3) embedded locations in the text content of tweets. Temporal information is associated with the tweeting time, as well as the possible temporal expressions presented in tweet contents. Semantic information, as we defined in Chapter 3, refers to the kinds of hazard events and their associated meanings, for example, *airport closed*, *winter weather advisory*, or *power shortage*, and their general abstractions (modeled as upper level classes) that are not explicitly described in text documents, such as *hazard alert*, *hazard impact*, *response*, or *hazard recovery*. An important contribution of this research is the approach for combining the features from Twitter (GPS, user profiles, tweeting time, and tweet content) to automatically assign appropriate spatial and temporal information to semantic information, for which the details will be discussed later in this chapter.

This chapter focuses on developing a new approach for automatically extracting spatial, temporal and semantic information from tweets that is implemented in a Java environment with an open source toolkit and API: Twitter POS Tagger

(<https://gate.ac.uk/wiki/twitter-postagger.html>) and GATE API (<http://jenkins.gate.ac.uk/job/GATE-Nightly/javadoc/index.html>). The research provides an overview of the emerging opportunities for harvesting spatiotemporal and semantic content. Several related data handling topics, such as spatial filtering, and data normalization, are discussed during the information analyzing process. For this approach, the analyzed results from tweets are combined with the results from processing Web news reports by employing the information

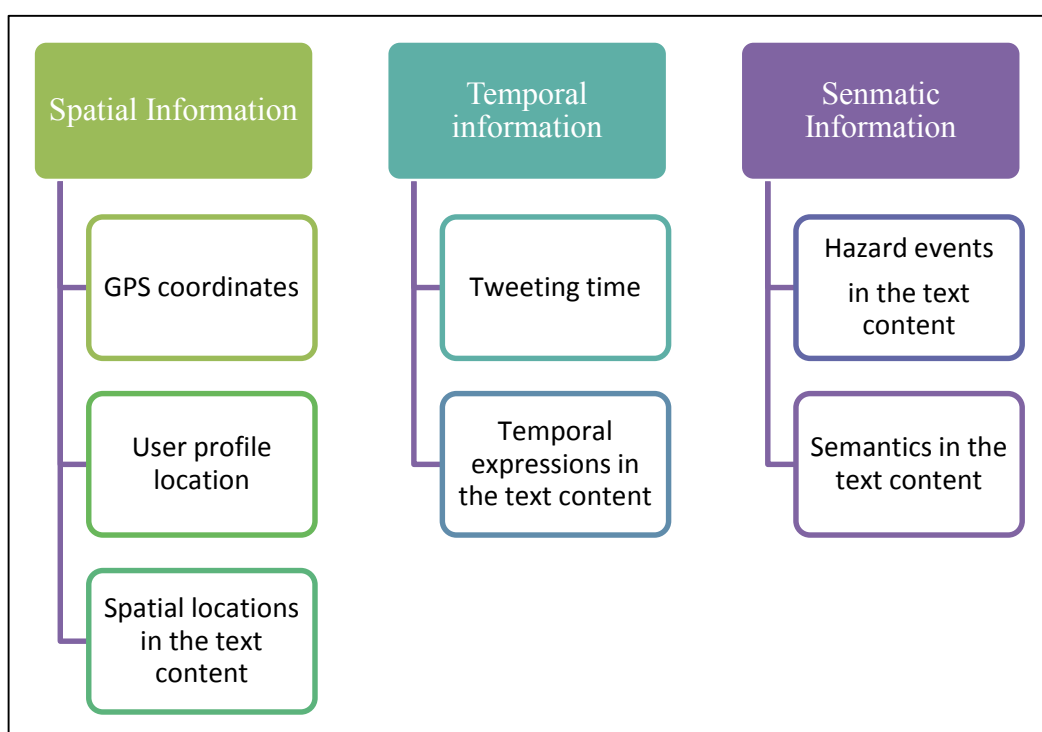


Figure 4.2 Spatial information, temporal information, and semantic information from Twitter utilized in this research

extraction approaches in Chapter 2 and Chapter 3. The research questions addressed in this chapter are:

- Can the same methods for spatial, temporal and semantic information extraction developed for Web news reports, be applied to tweets? What additional processing approaches are needed?
- What is the potential role of data extracted through processing Twitter feeds with respect to the results already achieved from processing Web news reports?

A case study of snowstorms that occurred in the southeastern United States during January 2014, is used to examine the key issues of applying the approach discussed in this chapter to tweets and the potential role of the approach for extracting information about hazard dynamics. In order to evaluate the approach, a data set of 27, 0000 tweets for three days, Jan 27, 28, and 29, 2014 were collected using the Twitter Streaming API and Python procedures. Although the case study is based on a winter storm scenario, it is possible to apply the approach for other hazard events, such as hurricanes, tornadoes, or flooding.

4.2 Related Work

Research on mining Twitter data is rapidly expanding. Common research topics based on Twitter include news topic detection (Sankaranarayanan et al. 2009; Yang and Rim 2014), sentiment analysis (Liu 2010; Pak and Paroubek 2010; Nielsen 2011; Bollen et al. 2011), disease spreading estimation (Aramaki et al. 2011; Dredze et al. 2013), and natural hazard detection (Goodchild and Glennon 2010; Crook et al. 2013; Chen et al. 2014). The main challenge for Twitter data analysis is the high ratio of noises (i.e., abbreviations, slang and spelling errors) contained in the unstructured text (Ritter et al. 2011; Neubig et al. 2011; Saif et al. 2012). Therefore, the traditional NLP techniques, such as named entity recognition (NER) and part-of-speech (POS) tagging, needed to be improved to apply on the complex text content of Twitter (Corvey et al. 2010; Neubig et al. 2011).

TwitterStand is a web-based news processing system that aims to identify breaking news through Twitter and visualize their geographic locations on map (Sankaranarayanan

et al. 2009; Jackoway et al. 2011; Teitler et al. 2014). In TwitterStand, tweets are collected using Twitter Services, such as Seeders and Search, and are separated into news and junk through training a Naive Bayes classifier, a machine learning based approach for classifying text documents into two different categories (Mitchell 1997). Identified news-related tweets published in recent three days are clustered in groups based on different topics. NER and POS approaches are combined to extract geographic information from Twitter messages (Sankaranarayanan et al. 2009).

Sentiment analysis is another common Twitter research topic that captures the level of public interest and topics that involve strong sentiments by Twitter users (Liu 2010; Pak and Paroubek 2010; Nielsen 2011; Bollen et al. 2011; Saif et al. 2012). Semantics contained in tweets are an important component for sentiment analysis (Saif et al. 2012). By analyzing semantic information contained in the tweet messages, Twitter users' emotions are monitored through sentiment analysis. A corpus with positive, neutral, and negative sentiment references is used as a sentiment classifier with supervised machine learning techniques including Naive Bayesian, Maximum Entropy, Support Vector Machine, and NLP techniques, such as POS tagging, to determine the emotion for each tweet (Pak and Paroubek 2010; Nielsen 2011). Twitter users' sentiments can be traced and tracked for political election results estimation (Tumasjan et al. 2011; Tsou et al. 2013), or stock market trend prediction (Bollen et al. 2009).

In public health studies, researchers are interested in retrieving geographic location information from Twitter for disease tracking (Lamb et al. 2013; Dredze et al. 2013). Garmen, a system to obtain structured location information (e.g., country, state, county, and city) from Twitter, is applied in public health to improve influenza surveillance (Dredze et al. 2013). The spatial information that Garmen detects from Twitter are based on GPS coordinates from mobile and user profile information. Various models are utilized in epidemics research on Twitter to measure risk factors and track diseases, for example, a

linear regression model (Culotta 2010), a terms co-occurrence model (Quincey and Kostkova 2010), and a topic aspect model (Paul and Dredze 2011).

Research shows that analyzing Twitter for hazards events is faster than analyzing other official announcements because of its real-time nature (Hughes and Palen 2009). The earthquake events in Japan 2010 were detected from Twitter, and locations associated with the earthquake were estimated (Sakaki et al. 2010). Each Twitter user who tweeted an earthquake-related event is treated as a sensor. The time and geolocation information provided by the sensor were assigned with a higher probability to reduce the noise and uncertainty that caused by Twitter data. To estimate the locations of the earthquake, Bayesian filters, such as Kalman and Particle filters, were applied. In addition, semantic analysis was conducted to classify tweets into a positive class and a negative class, based on the content (Sakaki et al. 2010). The final results were sent as immediate alerts to communities to provide the local people more time to response to the hazard impact. Similar to this research, Twitter can be treated as a distributed sensor system to support monitoring of natural events (Crooks et al. 2013). Each Twitter user is considered as a mobile sensor for detecting the relevant geographic events (e.g., an earthquake that occurred on the east coast of the US on August 23, 2011) through an analysis of the content of their tweets.

Effectively extracting spatiotemporal and semantic information from Twitter is a challenging task. Currently, most studies focused on hazard-related tweets choose geographic coordinates from GPS-enabled devices (Fuchs et al. 2013; Li et al. 2013) as their inputs for events detection, for example, latitude and longitude “34.06824, -81.1569”, however, only 2% of tweets have GPS coordinates (Dredze et al. 2013). In addition, the accuracy of location estimation should not be dependent only on the coordinates. For example, it’s possible that a user who attached his/her current location to a tweet could refer to an event that happens in a different area. Besides the coordinates, locations can be detected from user profiles (Crooks et al. 2013). Twitter users usually register their

residential locations in the user profile (Sakaki et al. 2010). However, research shows 30% users enter valid geographic locations for their user profiles (Cheng et al. 2013). The spatial information in user profiles is associated with different granularities, where 20% are restricted to major cities. 70% of users either leave the profile location blank or enter nongeographic information (Cheng et al. 2013). Our work in this chapter will extend the research in this dissertation by automatically analyzing the spatiotemporal and semantic characteristics of hazard events expressed in tweets, and to extend previous work that focused solely on extractions from Web new reports to a media where real-time tweeted information is also available.

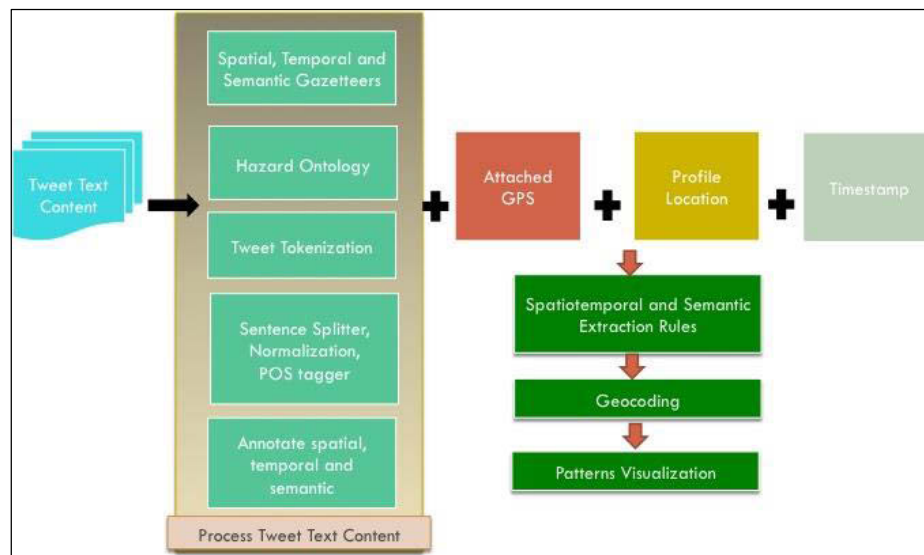


Figure 4.3 A framework for processing four features of Twitter

4.3 Processing Tweets using NLP and GIR

This research presents a framework to process Twitter for detecting spatial, temporal and semantic information. The key components of the framework are shown in

Figure 4.3 and include 1) crawling the tweets on the Web and parsing features of the tweets; 2) analyzing the spatial, temporal, and semantic information from the content of tweet messages using NLP techniques; 3) extracting spatial and temporal information in the tweets that are directly related to the semantic information by following a set of rules developed for this research.

4.3.1 *Crawling tweets and parsing features of the tweets*

Tweets are crawled from the data provider (Twitter server) for a continuous time period (e.g., 3 days) by applying the Twitter Streaming API (<https://dev.twitter.com/docs/api/streaming>). A retrieving request is submitted to the HTTP server to establish a connection between the Twitter server and front-end Twitter users (Figure 4.4). The streaming connection process is responsible for handling the retrieving requests. Once the connection is built and a request is received, the Twitter server opens a streaming connection to receive streamed tweets from the front-end Twitter users. Each tweet is stored in a local user database with JavaScript Object Notation (JSON) format. In this research, we are interested in certain features including the *content of tweet messages*, *tweeting time*, *user profile locations*, and any *attached GPS data* from Twitter. Rules are created to parse these four features, and the parsed elements of each tweet are stored in the local database with the following pattern:

```
{
  "Text Content": "University will be closed Wednesday due to winter weather.",
  "Time": "Tue Jan 29 04:24:45 +0000 2014",
  "Profile Location": "Georgia Tech ",
  "GPS": {"X": 33.74362, "Y": -84.374}
}
```

As discussed in the previous section, Twitter users can enter geographic information at different granularities in their profiles, for example, “Wisconsin”, “Greenville, NC”, “Central Texas”, “US”, or “\u00dcT: 36.72517 -76.33804”. It’s also possible to find non-geographic information, such as user ID, a Web link, or a brief

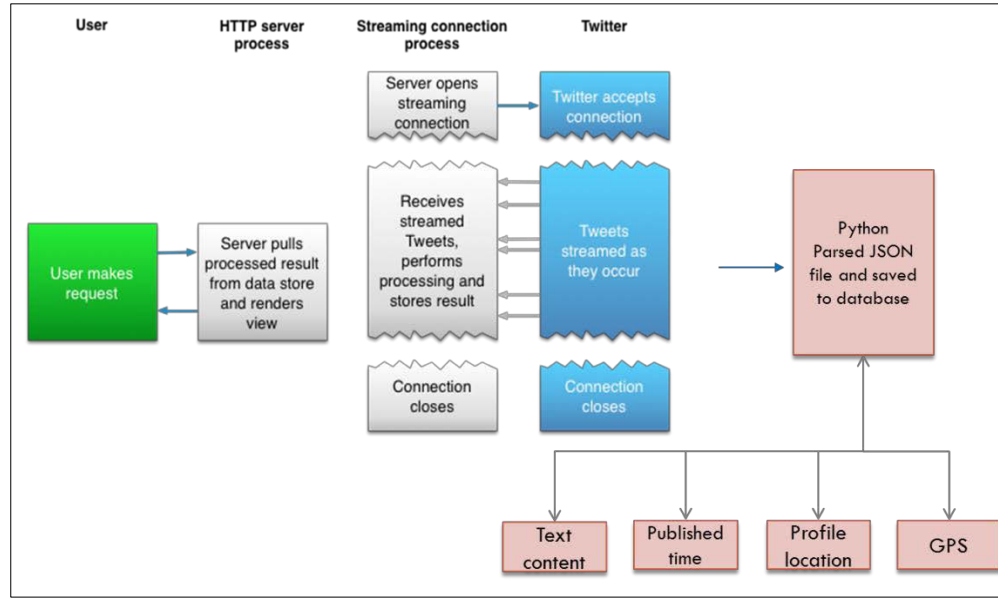


Figure 4.4 Crawling Tweets using Twitter Streaming API

biography, in the users' profiles. However, some of the profile information is not relevant, for example, "chasing dream in the clouds" or "Six feet under the stars". Therefore, a filtering procedure is applied to exclude any non-geographic information that exists in the users' profiles. Since our study is focused on events in the United States, any profile locations outside the United States are filtered during Twitter crawling.

In addition, data normalization is an important step for information retrieval from tweets. Since larger cities contain populations with high densities, it's possible for more Twitter activities to occur in these cities, rather than other (smaller) locations. Also re-tweeted tweets leads to a large redundancy of the same spatiotemporal and semantic results, as well as ambiguities when the assignment of spatial and temporal information to the appropriate hazard events is undertaken. In order to exclude the spatial pattern that is correlated with larger population densities and high numbers of re-tweeted tweets, data

normalization is conducted to minimize dependency and redundancy of results from tweets using population and numbers of retweets.

4.3.2 Extracting spatial, temporal, and semantic information from the text content of tweets

In the previous subsection, it was stated that four features of tweets are parsed, including the text messages, tweeting time, users' profiles and GPS coordinates. In this section, an approach for automatically extracting spatial, temporal, and semantic information from the text content of tweets through the application of a set of NLP techniques will be illustrated.

In order to handle the text content of tweets, the spatial, temporal, and semantic gazetteers developed in Chapter 2 and Chapter 3 will be used to annotate the tweets with respect to spatiotemporal information as well as the hazard-related semantics. In addition, the hazard ontology from Chapter 3 is once again linked with the semantic gazetteer to assign various semantic meanings to the extracted hazard-related events. In this way, the events can be clustered into different higher-level classes, e.g., *HazardImpact*, *HazardWarning*, *HazardResponse*, and *HazardRecovery*, for hazard pattern detection according to the needs of the users.

However, the irregular structure of tweets and noise contained in tweet messages pose challenges for information extraction (Ritter et al. 2011; Andrienko et al. 2013). Available research shows that the traditional NLP annotating accuracy drops significantly in tweets, because gazetteers with formal references cannot solve Twitter's wide range of named entity types, such as misspelling, slang, and jargon (Ritter et al. 2011; Derczynski et al. 2013). Lacking large annotated text sources based on tweets might result in poor extraction performance results (with very low precision and recall values). The performance of the extraction is likely to be improved by reducing the proportion of "noisy" and "unknown" vocabulary or terms that exist in tweets. A machine learning-based

POS tagger can be used for applying the extraction tasks to new data with labeled training through supervised learning (Derczynski et al. 2013). Consequently, a GATE twitter POS tagger is adopted in this research for enhancing the performance of the extraction tasks and handling, for example, slang and noise. In addition to this adopted tagger, an additional Twitter semantic gazetteer is created to specifically support the extraction of hazard-related events from Twitter. To create this gazetteer, a data corpus of tweets with 20k tokens is collected from November 2013 to March 2014 based on a set of keywords, for example, “snowstorm”, “blizzard”, “hurricane”, and “tornado”. The hazard events (e.g., *storm brewing*, *delay opening*, *no school*, *curfew*, and *postpone game*) from this data corpus are manually annotated, and stored in a set of .lst files as a twitter semantic gazetteer. The contents of this gazetteer corresponds with the different classes in the hazard ontology, e.g., *utilityimpact.lst*, *hydrologicalhazard.lst*, *communityresponse.lst*, etc. Each list file contains different hazard-related topics events as identified in the tweets during training. Finally, rules have been implemented in Java using the GATE API to support spatial, temporal, and semantic references matching from the gazetteers and extraction from the tweets text content.

The approach used to process the content of tweets, includes several steps: linguistic processing involving tweet tokenization, sentence splitting, applying gazetteers and ontologies, and tagging using a twitter part-of-speech tagger. Then, our method stores the extracted spatial, temporal and semantic information in a local database (Figure 4.5).

4.3.3 Rules for extracting spatial information

Locations detected correctly from tweets determine the accuracy of the analysis results. As we claimed in section 4.3.1, spatial information may exist in tweet contents, attached GPS coordinates, or user profiles. However, not all the spatial information given in the tweets are directly related to the hazard events. It is not trivial to extract the useful

Algorithm 1 Parse Text Content of Tweets**Input:**Text Content T_C **Output:**result $T (T_{\text{spatial}}, T_{\text{temporal}}, T_{\text{event}}, T_{\text{semantic}})$

Begin

for each text content of tweet **do** **parse** (T_C) rule(spatiotemporal and semantic extraction)(T_C) result $T \leftarrow \text{add}(T_{\text{spatial}}, T_{\text{temporal}}, T_{\text{event}}, T_{\text{semantic}})$ **end for****return** result T **Figure 4.5** Parse the spatial, temporal and semantic content of tweets

and accurate location information from the raw tweets. To this end, a routine is derived that combines the GPS data, user profiles, and spatial information in the tweet contents together for better representation of location information.

The following cases illustrate how spatial information is systematically assigned to a related hazard event when:

- 1) only the spatial information in the text content is available;
- 2) only the GPS data is available;
- 3) only the user profile location is available;
- 4) both the GPS data and the user profile location are available;
- 5) both the GPS data and spatial information in the text content are available;
- 6) both the spatial information in the text content and the user profile location are available;
- 7) the GPS data, user profile location, and spatial information in the text content are available.

An algorithm is developed for each of cases above to determine how the spatial information is assigned to the hazard event in a tweet. Concerning the best spatial information candidate, we order the three features according to their priorities: the explicit spatial information present in the tweet content > GPS coordinates > the user profile location.

1. When spatial information is explicit in the text content

This is the situation in which the spatial information is provided solely in the text content, for example,

```
{ "Text Content": "Flight delays at Detroit Metro Airport. ",
  "Time": "Wed Jan 29 09:46:02 +0000 2014",
  "Profile Location": null,
  "GPS": {"X": null, "Y": null}.
```

We assign the available spatial information in the text content to the hazard event. In this example, *Detroit Metro Airport* is assigned to the *Flight delays*.

2. When only the GPS data is available

When the GPS data are available for a tweet, this information is assigned to the detected event directly. For example,

```
{ "Text Content": "Due to the Winter Storm school canceled this evening! Be safe! "
  "Time": "Tue Jan 28 13:36:35",
  "Profile Location": null,
  "GPS": {"X": 32.51861071, "Y": -87.83853678}.
```

In this example, the hazard event of *school canceled* is assigned to geographic coordinates directly.

3. When only the user profile location is available

In this case, the user profile contains the only spatial information available and this information is assigned to the event. For example,

{"Text Content": "Winter weather causes two fatalities",

"Time": "Wed Jan 29 14:45:18 +0000 2014",

"Profile Location": "*Montgomery Alabama*",

"GPS": {"X": null, "Y": null}.

In this example, no spatial information is contained in the text content, and no coordinates are attached. Therefore, the user profile location is treated as the only spatial information associated with the event. In the user profiles, locations can be associated with different granularities (e.g., geographic coordinates, schools, airports, rivers, street names, cities, counties, or states) including explicit references (e.g., Charleston, South Carolina) or vague reference (e.g., south GA). For this study, we only consider profile spatial information designated with the granularity at city level (e.g., Austin TX) or finer than the city level (e.g., x, y coordinates, airports, street addresses, etc.). Spatial information with coarser granularity than city level or vague information, such as US, West Michigan, will be filtered. In this case, *Montgomery Alabama* is assigned to the event of *two fatalities*.

4. When both of GPS coordinates and spatial information in the text content are available

This case is more complicated than the previous cases especially if more than one kind of information is available in a tweet contains different information. GPS coordinates are point-based geographic locations. The spatial information contained in the text content, similar to the user profile locations, might be associated with different granularities (e.g., geographic coordinates, street names, cities, counties, states, etc.) and explicit or vague reference (e.g., Detroit Metro Airport or south GA). We need to first obtain the spatial relationships between the two features (i.e., where the GPS data is regarded as a point

feature A, and the spatial reference in the text content is regarded as geometry B), and then determine which spatial information should be assigned to the hazard event.

If the GPS data equals to the spatial reference (i.e., geographic coordinates) extracted from the text content (the two features are completely coincident), the GPS coordinates are assigned to the hazard event. Otherwise, the spatial relationships (e.g., within) between these two features are needed to determine which feature should be considered as a better candidate. If the geometry B is a polygon-based feature (e.g., cities), and the point A is contained within the boundary region of B, the GPS data is assigned to the event (Berke and Shi 2009). If the point A is not contained within the boundary region of B (i.e., the two geographic features refer to two different places), the spatial information in the text content is regarded the mostly related geographic feature to the hazard event, and the spatial information in the text content will be assigned to the event. If the geometry B is a line-based feature (e.g., streets or rivers) and point feature A is on geometry B, A is assigned to the event. If A is not on B, B is assigned to the event. For example,

```
{"Text Content": "Who woulda thought class would get cancelled in Atlanta for winter weather? ",
```

```
"Time": "Tue Jan 29 04:24:45 +0000 2014",
```

```
"Profile Location": null,
```

```
"GPS":{"X": 39.73617,"Y": -84.1751}}.
```

In this case, *Atlanta* and 39.73617, -84.1751 are parsed as spatial information. Comparing *Atlanta* and the GPS coordinates, it is found that the GPS coordinates (located in *Dayton, OH*) are not contained in *Atlanta*. The spatial reference in the text content will be considered as a candidate to be assigned to the hazard event. Therefore, *Atlanta* is assigned to the event *class would get cancelled*.

5. When both of GPS data and profile location are available

It is possible for both the GPS data and user profile location to be available in a tweet. In this case, the GPS data will always be selected as the candidate for assigning to the event. For example,

```
{ "Text Content": "Trying to stay home during this snow\ice storm.",
  "Time": "Jan 29 04:27:12 +0000 2014",
  "Profile Location": "North Carolina",
  "GPS": {"X": 35.62556, "Y": -78.3286}.
```

In this case, the geographic coordinates are contained in *North Carolina*. Therefore, the geographic coordinates are determined as the candidate for this tweet.

6. When both spatial information in the text content and profile location are available

In this case, no matter if these two kinds of spatial information are identical or not, the spatial information in the text content is always selected as the candidate. In the case where the two kinds are not identical, the profile location can be considered when the spatial information in the text content is ambiguous (e.g., one geographic name can be referred to difference places). In this case, if the spatial information in the text content is contained within the location in the user profile, then the profile location can be added to the spatial information in the content to assign to the event. The profile location provides a “boundary region” for reducing the ambiguity caused by the vague spatial information. For example, if *Springfield* is the only spatial information detected in the tweet content, and *Illinois* is detected in the user’s profile, then *Illinois* can be used as a reference for distinguishing *Springfield, Illinois* than the other 32 *Springfield* around the United States. For example,

```
{ "Text Content": "Winter Weather NO SCHOOL TOMORROW (Thursday) at Fox
Creek",
  "Time": "Wed Jan 29 20:03:13 +0000 2014",
```


"Profile Location": "*North Augusta*",
 "GPS": "X": null, "Y": null}.

Fox Creek and *North Augusta* are annotated as spatial information in this example. There are four different places and eight different rivers named as *Fox Creek* in the US. With the lack of other geographic references in the tweet content, the user's profile location is considered. Combining *Fox Creek* and *North Augusta*, the geographic location (*Fox Creek, North Augusta*) is assigned to the event *NO SCHOOL*.

7. When GPS coordinates, profile location, and spatial information in the text content are available

This is the most challenging case. The spatial information in the text content is compared with the GPS coordinates first. If they are identical, the GPS information is assigned to the hazard event. If not, the spatial relationship between these two features need to be identified by the system to determine which feature should be considered as a better candidate (similar to Case 4). If the GPS coordinates are *not contained within* (polygon geometry) or *not on* the spatial reference (line geometry), the spatial information in the content is compared with the profile location, and the rules discussed in Case 6 can be applied to select the best candidate. For this case, it's common to see the situation where two features or all the three features refer to the same place. For example,

{"Text Content": "Winter weather update: # *Birmingham* main Library will be CLOSED on Wed. Jan. 29",

"Time": "Jan 29 04:28:09+0000 2014",

"Profile Location": "*Birmingham* ",

"GPS": "X": 33.52042343, "Y": -86.80743526}

In this example, the GPS data is contained in *Birmingham*, so *GPS* is assigned to the event *Library will be CLOSED*. For another example,

{"Text Content": "*Atlanta* was upgraded to a winter storm warning at 3:38am Tuesday",

"Time": "Wed Jan 29 04:29:10 +0000 2014",
 "Profile Location": "*Atlanta*",
 "GPS": "X": 40.40250900, "Y": -79.98393819},

In this example, *Atlanta* annotated in the tweet content is compared with the GPS coordinates 40.40250900, -79.98393819 (located at Pittsburgh, PA). It is found that the coordinates are not contained within *Atlanta*. The profile location is identical to the spatial information in the text content, and then *Atlanta* is chosen as the candidate. It's also possible that all the three features refer to different places, respectively. For example,

{"Text Content": "Some students in *GA* still stranded in schools after tonight's #snowstorm. National Guard called in to rescue drivers. <http://t.co/V9VG5yneW86>",
 "Time": " Wed Jan 29 04:28:18 +0000 2014",
 "Profile Location": "*Austin TX*",
 "GPS": "X": 37.15973206, "Y": -84.1103043900}.

In this example, *GA*, *Austin TX*, and 37.15973206, -84.1103043900 are annotated as the spatial information. The GPS data is not contained in *GA*, so the profile location *Austin TX* is compared with *GA*. It is found that they also belong to different states. Therefore, the spatial information of the text content, *GA*, is selected as the candidate for assigning to the hazard event.

Since the spatial information in the tweet is given priority, the algorithm is used to assign appropriate spatial information to the related events (summarized in Figure 4.6). We discuss how well these rules perform in an evaluation in Section 4.5. In the next section, we discuss how to assign temporal information to the hazard events.

4.3.4 Rules of extracting temporal information

Two types of temporal data are considered in this research: 1) the tweet timestamp, and 2) the temporal expressions contained in tweet contents. If no temporal information is

Algorithm 2 Assigning Spatial Information

Input: Record R_n (GPS Location S_{GPS} , Profile Location $S_{profile}$, Published Time $T_{published}$, Tweet Content T_C , Tweet location $T_{spatial}$, Tweet time $T_{temporal}$, Tweet event T_{event} , Tweet Semantic $T_{semantic}$)

Output: T_C , S_{GPS} , $S_{profile}$, $T_{published}$, $T_{spatial}$, $T_{temporal}$, T_{event} , $T_{semantic}$, Final location $F_{Spatial}$, Final temporal $F_{temporal}$

```

Begin
  for each  $R_n$  do
    process  $R_n$ 
    if only  $S_{GPS}$  existing, then
      assign  $S_{GPS}$  to  $F_{Spatial}$ 
    if only  $T_{spatial}$  existing, then
      assign  $T_{spatial}$  to  $F_{Spatial}$ 
    if only  $S_{profile}$  existing, then
      assign  $S_{profile}$  (filtering) to  $F_{Spatial}$ 
    if  $S_{GPS}$  and  $T_{spatial}$  existing, then
      compare  $S_{GPS}$  and  $T_{spatial}$ 
      if  $S_{GPS} == T_{spatial}$ , then
        assign  $S_{GPS}$  to  $F_{Spatial}$ 
      if  $S_{GPS} != T_{spatial}$ , then
        if  $S_{GPS}$  is contained within  $T_{spatial}$ , ||  $S_{GPS}$  is on  $T_{spatial}$  then
          assign  $S_{GPS}$  to  $F_{Spatial}$ 
        else if
          assign  $T_{spatial}$  to  $F_{Spatial}$ 
    if  $S_{GPS}$  and  $S_{profile}$  existing, then
      assign  $S_{GPS}$  to  $F_{Spatial}$ 
    if  $T_{spatial}$  and  $S_{profile}$  existing, then
      compare  $T_{spatial}$  and  $S_{profile}$ 
      if  $T_{spatial}$  is contained within  $S_{profile}$  then
        assign ( $T_{spatial}$ ,  $S_{profile}$ ) to  $F_{Spatial}$ 
      else if
        assign  $T_{spatial}$  to  $F_{Spatial}$ 
    if  $S_{GPS}$ ,  $T_{spatial}$ ,  $S_{profile}$  existing, then
      compare  $S_{GPS}$  and  $T_{spatial}$ 
      if  $S_{GPS} == T_{spatial}$ , then
        assign  $S_{GPS}$  to  $F_{Spatial}$ 
      if  $S_{GPS} != T_{spatial}$ , then
        if  $S_{GPS}$  is contained within  $T_{spatial}$ , ||  $S_{GPS}$  is on  $T_{spatial}$  then
          assign  $S_{GPS}$  to  $F_{Spatial}$ 
        else if
          compare  $T_{spatial}$  and  $S_{profile}$ 
          if  $T_{spatial} == S_{profile}$  then
            assign  $T_{spatial}$  to  $F_{Spatial}$ 
          if  $T_{spatial} != S_{profile}$ , then
            if  $T_{spatial}$  is contained within  $S_{profile}$  then
              assign ( $T_{spatial}$ ,  $S_{profile}$ ) to  $F_{Spatial}$ 
            else if
              assign  $T_{spatial}$  to  $F_{Spatial}$ 

```

Figure 4.6 Algorithm of assigning spatial information

found in the tweet text contents, the timestamp associated with a tweet is assigned to the hazard event. If any temporal information has been detected from the text content, we compare it with the tweeting time to determine which temporal information is more closely related to the hazard events.

Algorithms developed in Chapter 2 and Chapter 3 are applied to this research for handling the temporal information. For example,

{"Text Content": "Due to the Winter Weather Storm our library will be closed *tomorrow!*
Be safe! "

"Time": "*Tue Jan 28 10:30:35*",

"Profile Location": null,

"GPS":{"X": 30.34566798, "Y": -89.16351993},

Tomorrow and *Tue Jan 28 10:30:35* are parsed as different temporal expressions and *tomorrow* is assigned to the event *library will be closed*. The timestamp of the tweet serves as a criterion to compare temporal expressions (e.g., Thursday, tomorrow, the next day) in the text content. In this way, these expressions may be ordered temporally and assigned with appropriate dates. In this case, *tomorrow* is assigned with a period between *12:00 am and 11:59 pm on Wednesday*. All temporal information needs to be converted to a standard time format (i.e., YYYY-MM-DD hh: mm: ss) in the geodatabase for being further processed.

With the heuristics developed in this section, spatiotemporal and semantic information from hazard-related tweets are automatically extracted. In the next section, these rules are illustrated through a case study of severe winter weather for the eastern US in January 2014. A set of analyses will be conducted based on the extracted results to interpret the geographic dynamics (underlying hazard patterns over space and time) in a GIS environment.

4.4 Case Study-Snowstorm in January 2014

In this section, a case study using snowstorm-related tweets is presented. The goal of the case study is to understand: (1) how spatiotemporal and semantic information can be extracted from a set of tweets using the process and rules described above and what can be learned about hazard event dynamics? (2) How can information from news reports be combined with information from Twitter to represent hazard dynamics?

4.4.1 Tweet data

The dataset in this study includes 270,000 tweets about winter storms that occurred in Southeastern of the United States during the week of January 27, 2014. Each tweet is time-stamped, and covers the time period from January 27 to January 29, 2014. Figure 4.7 shows an overview of locations that are related to the snowstorms in the dataset. The frequency of the locations mentioned in the tweets is represented by graduated symbols. In the data set, 1% of users attached GPS coordinates, and 64% of users listed location information in their profiles. Applying the spatial filtering to the profile information, it is found that 87% of the profiles were associated with geographic information. Among the profile locations, 68% of the data are restricted to major cities or even at finer spatial granularity (e.g., street level). After processing the text of the tweets, it was found that 30% of the tweets contain spatial information in their text content.

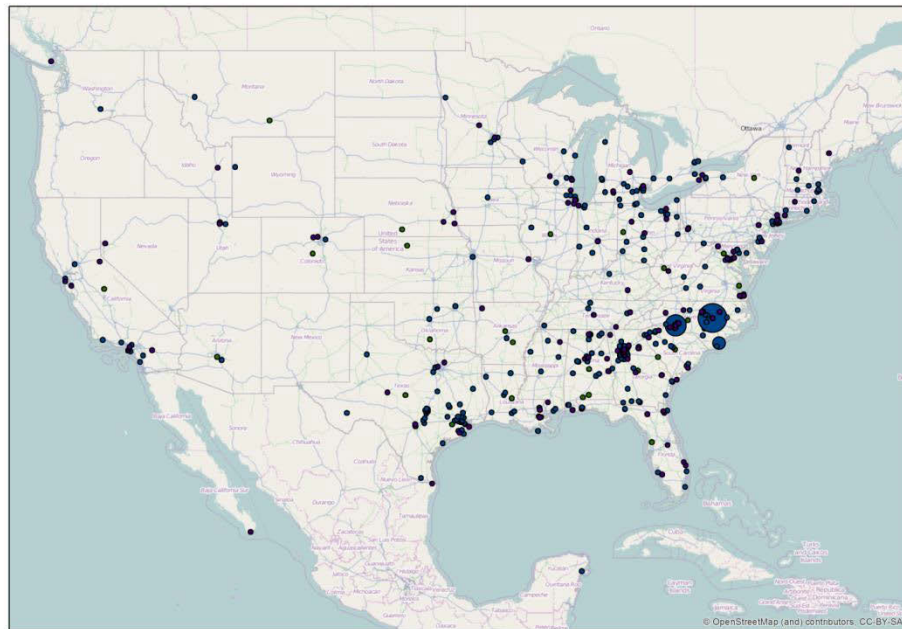
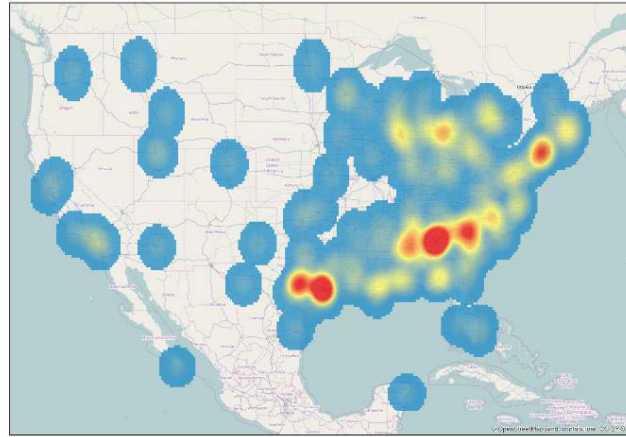


Figure 4.7 An overview of the spatial distribution of the snow storm-related tweets from Jan 27-29, 2014

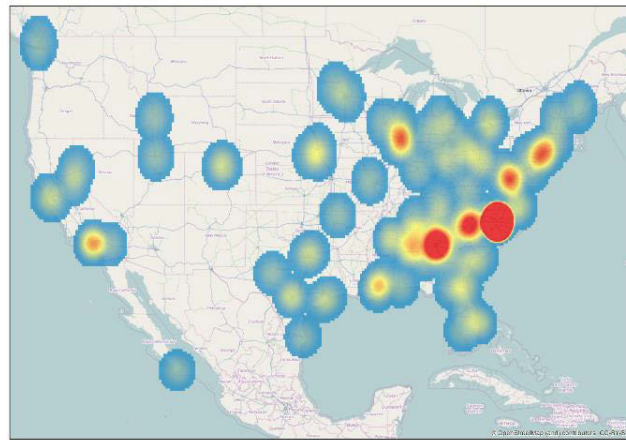
4.4.2 Spatiotemporal clustering of tweets

In order to detect the hazard patterns that are embedded in the extracted spatiotemporal and semantic information from Twitter, kernel density analysis is conducted, and the information extracted from Web news reports for this time period are integrated with the extracted information from tweets.

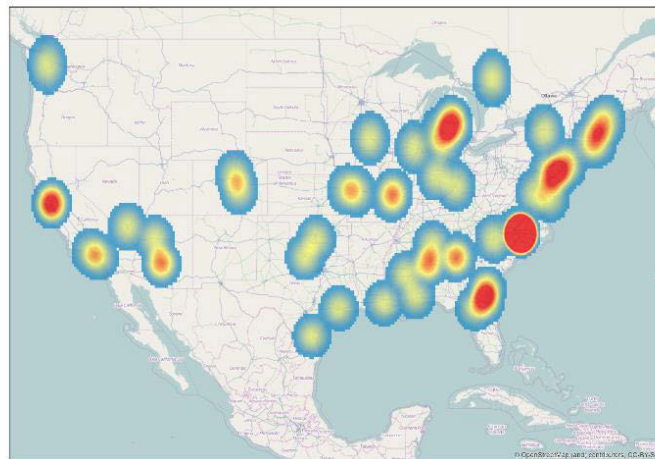
To better explore patterns of geographic dynamics from Twitter, a kernel density-based clustering analysis is conducted on the set of snow storm-related tweets using the geocoded locations at different times (Jan 27, 28, and 29, 2014). Kernel density analysis is a way of estimating the intensity of points by generating a smooth surface using a quadratic kernel function (Silverman 1986). The kernel density analysis used two parameters: a kernel search radius (bandwidth) of 110 km to calculate density, given the shorter of the width or height of the results extent in the output spatial reference divided by 30, and the cell size of 13 km for the output raster data given the shorter of the width or height of the results extent in the output spatial reference, divided by 250. The kernel search radius of 110 km is to avoid creating a map that is too smooth or too fuzzy to interpret. The cell size of 13 km was used to show fine detail. The patterns of events vary spatially per day (Figure 4.8a-c). The areas with highest density of events are represented with the darkest hue of red. Figure 4.8a shows a map with two major clusters, located primarily in the southern gulf coast region of the US, such as *New Orleans, LA, Houston, Edinburg, and Austin, TX*, and also, *Birmingham and Tuscaloosa, AL, and Atlanta, GA*, on Jan 27th 2014. As shown in Figure 4.8b, the area with high density shifts to the main cities along the east coast on Jan 28th, 2014, such as *Columbia, SC, Charlotte, Clayton and Greenville, NC, Washington DC*, and parts of *Pennsylvania*. In Figure 4.8c, there are four main clusters of high density along the east coast extending to the northeastern corner of the US. One cluster is located in the *Great Lakes region*, such as *Fairmont and Mankato, MN, and Amery, WI*, and another cluster is located in *Los Angeles, CA*.



(a)



(b)



(c)

Figure 4.8 Kernel density clustering based on snowstorm related events extracted from tweets (a) on Jan 27, 2013; (b) Jan 28, 2013; and (c) Jan 29, 2013

4.4.3 Interpreting spatiotemporal hazard patterns based on semantics from tweets

Applying ontologies as part of the information extraction process allows us to represent the extraction results according to the different kinds of events (Andrienko et al. 2013). In this research, the hazard ontology developed in Chapter 3 is used to categorize or classify the event-related terms from tweets according to the different upper-level classes in the hazard ontology. These events can be represented on maps at different levels of abstraction, such as *severe weather*, *power outage*, *school closed* or *airline canceled*. Modeling at multiple granularities offers supports for better summarization and representation of spatiotemporal patterns on maps. Making sense of spatiotemporal information through extracted semantic information reveals details about the dynamic aspects of hazards for users. For example, the spatiotemporal pattern of terms is used to understand more about the dynamic *hazard impact*-related events across space (Figure 4.9). Text terms related to *hazard impact* events, for example, *flights canceled*, *school is closed*, *accidents*, *icy roads* and *classes canceled*, can be mapped over the three days (Figure 4.9). The map reveals that hazard impact-related events that are associated with parts of the southern states along the gulf coast, western side of the US, the Great Lakes region, as well as the cities on the east coast. Schools were impacted by *snowstorms*, for example, many schools in Maryland, Washington DC, New York and New Jersey were closed.

4.4.4 Integrating extracted information from news reports with tweeted information

In the previous subsection, it was demonstrated that events in tweets can be abstracted to different upper-level classes (e.g., hazard impact) using the hazard ontology. Twitter provides a real-time data source associated with individuals reporting about their local events and provides more detailed temporal information as contributed by individual users (on-the-ground eyewitnesses) who are actually experiencing the hazard events. This

is in contrast to the information extracted from news reports that provide a more overall understanding of hazard events, possibly for an extent that is broader. Twitter becomes an important resource for possibly helping to augment the coarser spatiotemporal information provided in the traditional media, for example, Web news articles. Web news reports provide hazard-related information from a more macroscopic perspective. For Web news documents, hazard-related events can be summarized with less noise contained in the text documents, but potentially at a lower level of granularity. In addition, web news articles may not appear until some hours after an event while tweets could well be sent during the onset of a hazard. For this reason, Web news reports becomes a useful resource for verifying and reducing the noise impacts within a large amount of tweets. The combination of the two results provides a formal and rich data source to support hazard patterns

is in contrast to the information extracted from news reports that provide a more overall understanding of hazard events, possibly for an extent that is broader. Twitter becomes an important resource for possibly helping to augment the coarser spatiotemporal information provided in the traditional media, for example, Web news articles. Web news reports provide hazard-related information from a more macroscopic perspective. For Web news documents, hazard-related events can be summarized with less noise contained in the text documents, but potentially at a lower level of granularity. In addition, web news articles may not appear until some hours after an event while tweets could well be sent during the onset of a hazard. For this reason, Web news reports becomes a useful resource for verifying and reducing the noise impacts within a large amount of tweets. The combination of the two results provides a formal and rich data source to support hazard patterns

detection. The two different sources can be merged to complement each other, and this can be viewed as a step towards accessing the value of better harvesting information from social media feeds for event detection and analysis. With this approach, extracted information from tweets and news reports can be combined and utilized by stakeholders or users with an interest in the hazards. The inherent patterns of extracted results can be analyzed for better understanding the hazard dynamics.

Events (represented as spatiotemporal points) can be aggregated with density-based clustering techniques according to their locations. In addition to the hazard patterns detected from Twitter, a corpus of news reports from CNN is analyzed using the approaches discussed in Chapters 2 and 3 for augmenting the Twitter results. In this research, five news reports, collected from the *CNN* Web site about the same severe snowstorms from Jan 28th-30th, 2014, are processed to obtain spatiotemporal and semantic information using the approaches developed in Chapter 3. Analysis shows that 69% of the events extracted from these reports are *HazardImpact*, 13% of the events refer to *HazardResponse*, and 18% are associated with *HazardWeather*. Hazard impact events from the news reports, e.g., *airport closed*, *flights canceled*, *bus abandoned*, and *vehicle stranded*, are clustered on the map based on their locations (Figure 4.10). The highest density of hazard impact events is primarily concentrated in cities, such as *Atlanta*, *Brookhaven*, *Dunwoody*, and *Snellville* in *Georgia* and *Birmingham* and *Leeds* in *Alabama* during these days. Hazard impact events from Web news, and overlapping areas are located near *Gulf Coast*, and the *Great Lakes region*. There are more locations provided by tweets that are affected by snowstorm events (Figure 4.10). The tweets provide more fine-level spatial information and additional kinds of events being experienced by individuals that would not have been documented in the Web news reports. The event information retrieved from Twitter is used in combination with information extracted from news reports to enhance the understanding of hazard events from different spatiotemporal granularities

(e.g., macro-level and micro-level). For example, to estimate earliest occurring time and major impacted areas in association with different *hazard impact* events.

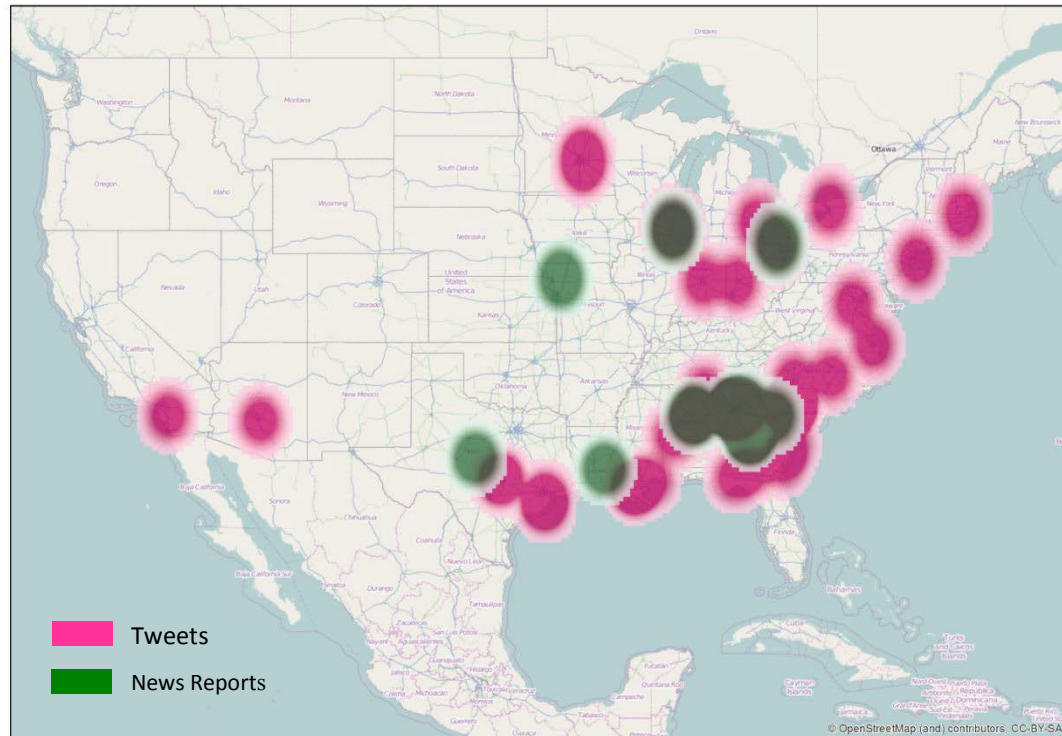


Figure 4.10 Clustering of hazard-events from CNN news reports and tweets

Integrating data from tweet data sources and also online news articles and exploiting GIS query functions supports the discovery of information that is not explicitly described in texts, but can be derived through spatiotemporal queries, for example, “*find all snowstorm response-related events in Raleigh, North Carolina on early Friday morning, January 30th*”. Another example is, “*find all airports affected by the snow storm on January 29th*”. The query results return 8 airports, including *Hartsfield-Jackson Atlanta International Airport, Louis Armstrong New Orleans International Airport, Houston Bush*

Intercontinental Airport, Cleveland Hopkins International Airport, Chicago O'Hare Airport, Detroit Metro Airport, William P. Hobby Airport and Johnston County Airport.

4.5 Evaluation

In this chapter, the approach shows how spatiotemporal and semantic information can be automatically extracted from Twitter. To assess how well the system performs, evaluations have been undertaken both qualitatively and quantitatively. The dataset for the evaluation for the system was based on a collection of 18,869 tweets about a New York building explosion that occurred in the morning of March 21st. A smaller set of 200 tweets was analyzed atomically by system and manually by human evaluators.

As a first step, a criterion for the evaluation of spatiotemporal semantic extraction of events is adopted to calculate the STSPrecision and STSrecall values described in Chapter 3.

- STSPrecision = the number of correctly assigned spatiotemporal semantic references / the number of spatiotemporal semantic references that the system or users attempt to assign;
- STSRecall = the number of correctly assigned spatiotemporal semantic references / the number of all such references.

For this evaluation, automatically processed results by the system are compared with a golden standard in order to acquire the numbers of correctly resolved spatiotemporal semantic references, incorrectly resolved spatiotemporal semantic references (combination set of spatial, temporal and semantic), and missing spatiotemporal semantic references from tweets. Five volunteers were used to provide a golden standard for this evaluation. All volunteer were trained to manually process spatiotemporal and semantic information from randomly selected 200 tweets by providing instructions and seven examples. The seven examples correspond to the cases discussed in section of this chapter. Each volunteer

manually annotated spatial, temporal, and semantic references and assigned the annotated terms to events based on the context in the evaluation data. The results were stored in a .csv database file. To obtain the golden standard, the results from each volunteer were compared. An acceptable standard for assessing the results obtained from the manual test is whether the results by *all* volunteers agree, or whether four out of five of the results agree. Results that correspond to three out of five or two out of five are required to be rechecked and discussed with the tester, to decide whether they should be included or excluded from the .csv file. Results with 0% agreement are excluded. Manually derived spatiotemporal semantic information sets, and the results obtained from automatically processing the tweets are each compared with the standard. The number of correct references, incorrect references, and missed references for the users and the system are determined.

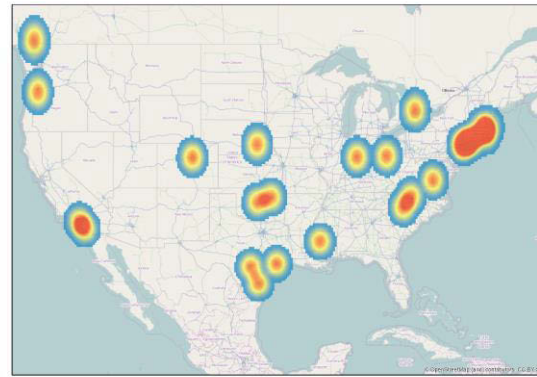
The results of precision and recall are calculated based on human performance and the system performance using our algorithms (Table 4.1). The average results annotated by human subjects are shown in this table. These results are based on 200 processed spatiotemporal semantic references in the tweets. For this evaluation, the average numbers were 191 correct references, 9 incorrect references, and 8 missed references. This results in precision and recall values of 0.96 and 0.96 respectively. For the same 200 processed spatiotemporal semantic references, the system performed with 172 correct references, 28 incorrect references, and 38 missed spatiotemporal and semantic references. Based on this performance, precision and recall are calculated as 0.86 and 0.81 respectively. Compared to the human performance, the recall value of the system is relatively low because of Twitter's wide range of named entity types, such as misspelling, slang, and jargon, that gazetteers with formal references cannot solve (Ritter et al. 2011; Derczynski et al. 2013). The irregular structure of tweets and noises contained in tweet messages pose challenges for this information extraction task (Ritter et al. 2011; Andrienko et al. 2013). With respect

to precision and recall results, further work is still required to improve the quality of spatiotemporal and semantic information extraction. For example, research on disambiguation and noise reduction in tweets. Methods are needed to reduce the ambiguity of the semantic information in tweet text contents.

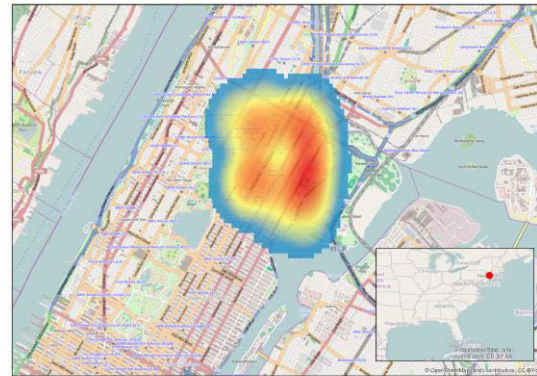
Table 4.1 Manually annotated and automatically processed precision and recall results based on numbers of solved spatiotemporal semantic references from Tweets

Tweets	Correct References		Incorrect References		Missed References		Precision		Recall	
	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto
200	191	172	9	28	8	38	0.96	0.86	0.96	0.81

To further evaluate the performance of our methods for spatiotemporal semantic assignments, another evaluation is conducted to check the quality of extracted results. In this data set, 40 tweets are attached with GPS coordinates. We generate two maps to investigate the regions that are associated with the explosion event based on the 40-geotagged tweets. Figure 4.11a shows the spatial distribution only based on the 40 GPS coordinates. The events are scattered over the US. Figure 4.11b illustrates the clustering of locations that are directly associated with the explosion events based on the approach discussed in this chapter (rules of assigning appropriate spatial information to the event based on tweet feeds, GPS locations and user profile locations). In Figure 4.11a, while in Figure 4.11b, the events are concentrated in Manhattan, NY, which reveals the ground truth of the events more accurately. The results show our extracted results overcome some noise caused by the process of assigning spatial and temporal information to the hazard events, which improves the results of analyses for exploring the patterns of hazards.



(a)



(b)

Figure 4.11 Spatiotemporal clustering of explosion events (a) spatiotemporal clustering of geotagged tweets using GPS coordinates; (b) spatiotemporal clustering of same data using the approach proposed in this chapter

4.6 Discussions and Conclusions

This work investigates how the retrieval of data from Twitter is processed and a set of methods for automatically extracting spatial, temporal and semantic information from tweets. First, the use of Twitter Streaming API for parsing four features of hazard-related tweets from the Web including GPS, user profile locations, tweeting timestamp, and text of tweet messages is discussed. During the process of parsing, a filter is applied to exclude the non-geographic information from the profiles. Then we demonstrated our methods to automatically process spatial, temporal and semantic information from the text of tweets

by applying NLP techniques and GATE. The results show that the same methods for spatial, temporal and semantic information extraction developed for Web news reports are useful for tweets. Spatial, temporal and semantic information detected from the tweet text messages are combined with other three features, with algorithms developed in this research to assign appropriate spatial information and temporal information to the hazard events. Seven cases were discussed for determining spatial information based on how the spatial and temporal information is presented in the four features of a tweet.

Data normalization and density analysis are applied on the extracted results to show the inherent spatiotemporal patterns of the retrieved hazard-related events from Twitter to exclude the spatial autocorrelation patterns. Events associated with upper level classes in the hazard ontology are mapped over space and time. Patterns of hazard-related events were detected through clustering the spatiotemporal information into groups based on semantics, revealing the evolution of the severe weather events monitored over time.

The method demonstrated through this work supports multi-scale, spatiotemporal event analysis of hazards, in this case, snowstorm, extracting information on severe weather, hazard warnings, house damage, cancelled flights, and emergency response, from tweets. In addition, this research also demonstrate the benefits of augmenting the results from processing Web news reports with data extracted through processing Twitter posts. Specifically, event information extracted from Web news reports can be integrated with analyzed results of tweets about the same events to enhance the understanding of hazards that is implicitly described in the texts, for example, identification of the impact area of the hazard-related events.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Discussion and conclusions

This dissertation explores three broad research topics related to representing geographic dynamics, the study of change or movement events with spatial and temporal characteristics based on an understanding of the principle functions of relevant forces and their relationships over space and time (Yuan and Stewart Hornsby 2008). This work centers on automated spatiotemporal and semantic information extraction about hazard events from Web news reports. Chapters 2, 3 and 4 can be regarded as three standalone journal papers. The research topics covered by the three chapters are related to each other, and are presented in a sequential way. This dissertation makes a unique contribution of bridging GIScience, geographic information retrieval, and natural language processing. This research applies geographic information retrieval and natural language processing techniques to extract spatiotemporal and semantic information automatically from Web documents, to retrieve information about patterns of hazard events that are not explicitly described in the texts.

The research begins with an investigation of methods for automatically extracting spatial and temporal information about hazards, for example, tornadoes, from Web news reports. A set of rules is developed to combine the spatial and temporal information contained in the reports based on how this information is presented in text in order to capture the dynamics of hazard events (e.g., changes in event locations, new events occurring) as they occur over space and time. The methods provide a systematic way to process spatial and temporal information from Web documents. Spatial and temporal gazetteers are introduced as key elements for spatial and temporal information annotation. A case study using CNN web news reports about tornadoes that hit the US Midwest, during April 2012 is presented in Chapter 2 to illustrate the steps discussed in this chapter. The

results of an evaluation comparing the results of volunteers with the results from the system processing show that the heuristics used for combining spatial and temporal information automatically lead to appropriate spatiotemporal assignments for the hazard events in the five cases. This research demonstrate that the capability of information extraction of an NLP system. GATE can be extended for GIR tasks through the use of spatial and temporal gazetteers, and presents a set of rules for assigning appropriate spatial and temporal information to events, based on the context of the text documents instead of processing spatial and temporal information in isolation. This work demonstrates that it is possible to retrieve a temporal ordering of extracted hazard events and correctly assign locations and times to these events.

Chapter 3 presents methods for retrieving semantic information about hazard events from documents. This work supports the understanding of hazard phenomena from two perspectives. One perspective relates to natural phenomena associated with hazards (e.g., *drought, dust storms, earthquakes, and landslides*) and the other relates to human activities associated with hazards (e.g., *transportation response, facility recovery, and donations*). This research develops and integrates a hazard ontology with a semantic gazetteer to support semantic information extraction from Web text documents. Two case studies using CNN reports on tornadoes in the Midwest in April, 2012 and Hurricane Sandy in November 2012 are undertaken. The major contributions presented in Chapter 3 are: creating a semantic gazetteer and hazard ontology for automated semantic information extraction of hazard events; integrating a semantic gazetteer and the hazard ontology in order to perform semantic information retrieval over multiple granularities of hazard information, including domain-related events (e.g., *power outage*) as well as higher level classes to which the events belong (e.g., *hazard impact* or *hazard response*); and demonstrating how mapping semantic information associated with events contributes to an understanding of the nature of geographic dynamics of hazard events.

Using the methods presented in Chapter 2 and 3, an approach for automatically extracting spatial, temporal, and semantic information from Twitter messages (tweets) is introduced in Chapter 4. Since tweets represent shorter, but more current information about hazards and how they are impacting a local area, key information about hazards can be retrieved through extracted spatiotemporal and semantic information from tweets. In this chapter, a set of rules for assigning appropriate spatial and temporal information to hazard events in tweets are presented using four features (text content of tweets, tweeting time, GPS data, and user profile locations). This work also investigates how hazard information extracted from tweets can be integrated with the information extracted from news reports and what the result of this combination offers. Web news reports become a resource for verifying and reducing the noise impacts within a large amount of tweets. On the other hand, Twitter provides a real-time data source associated with possibly finer-scale geographic locations and more accurate temporal information from individual users who are discussing the hazard events as they occur. In this way, the two different sources are merged to complement each other, and this is a step towards a more integrated approach for harvesting information from social media feeds for event detection and analysis. The case study for this work is a collection of 270,000 tweets about snowstorms in Southeastern of the United States from the week of January 27, 2014. The results show that the same methods for spatial, temporal and semantic information extraction developed for Web news reports are useful for tweet messages along with some additional processing such as assigning spatial information to the events based on the 7 cases discussed in this chapter. Evaluations have been undertaken both qualitatively and quantitatively to assess the performance of the system as compared to volunteers who manually process the tweets using the approaches presented in this chapter. The system performed to 86% for precision and 81% recall during these evaluations. The major contributions of this chapter are: a new framework for automatically extracting spatial, temporal and semantic information from tweets; algorithms for assigning appropriate spatial information and temporal information

to the hazard events in tweets; and augmenting the results from processing Web news reports with data extracted through processing Twitter posts to enhance the understanding of hazards that is implicitly described in the texts.

5.2 Future Work

Future efforts will be directed toward improving and extending the gazetteers. The spatial gazetteer can be extended in future studies to handle complex spatial expressions and different (e.g., topological) relationships among geographic locations, for example, “fuzzy” regions and intra-urban place names, e.g., *along the Mississippi river*, or *in and around Iowa City*. Extracting the richness of temporal information is also considered a significant research challenge for temporal information extraction. More complex temporal expressions from the text documents, such as event-anchored temporal information, for example, *two days before the flooding*, need to be considered for extending the temporal gazetteer. Also, the size of gazetteers affects the quality of the extracted results (Pasley et al. 2007; Lieberman and Samet 2012). Small gazetteers may restrict the ability to parse references from text documents (missing references), while large gazetteers slow the parsing process and may result in increased ambiguity. Therefore, further research is necessary to improve gazetteers for geographic information retrieval.

Some challenges with respect to geocoding also remain since generalized or vague locations are commonly included in text documents. New methods for translating these places into relevant event locations on a map are still needed. Geocoding not only relies on transforming addresses, place names, geographic entities, zip codes into x, y coordinates to benefit computer processing, but also includes identifying the spatial relationships between them. For the future research, more geocoding techniques, such as parcel geocoding and segmentation geocoding can be considered to represent line-based geographic features, such as the *Iowa River*, or generalized geographic features (polygon-based), *south of eastern Iowa*. The improvement of geocoding techniques benefits the

precision of the visualization representing the extracted information from text documents.

With respect to geographic information extraction based on tweets, future work is still required to improve the quality of spatiotemporal and semantic information extraction from tweets, for example, research on disambiguation and noise reduction in tweets. For example, the tweet “Both my nieces are sick. I swear they're taking turns coughing up a storm\n#nosleepforme” is not about natural hazards per se. Methods are needed to analyze contexts in tweets to reduce the ambiguity of the semantic information in the text content. Other social media, such as Facebook, Foursquare, and Google+ can also be considered for retrieving spatiotemporal and semantic information. In addition, although this research is mainly focused on hazard scenarios, the methods can be also extended and generalized to apply to other domain-related text documents, for example, criminal events with terrorists, political campaigns, and disease outbreak tracking.

REFERENCES

- Amitay, E., Har'El, N., Sivan, R., and Soffer, A. (2004) Web-a-where: geotagging web content. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sheffield, United Kingdom, 273-280.
- Andrienko, G., Andrienko, N., Bosch, H., Ertl, T., Fuchs, G., Jankowski, P., and Thom, D. (2013) Thematic patterns in georeferenced tweets through space-time visual analytics. *Computing in Science and Engineering*, 15(3), 72-82.
- Aramaki, E., Maskawa, S., and Morita, M. (2011) Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 1568-1576.
- Arendarenko, E. and Kakkonen, T. (2012) Ontology-based information and event extraction for business intelligence. In *Proceedings of the 15th International Conference of Artificial Intelligence: Methodology, Systems, and Applications, AIMSA 2012*, Varna, Bulgaria, 7557: 89-102.
- Berke, E.M. and Shi, X. (2009) Computing travel time when the exact address is unknown: a comparison of point and polygon ZIP code approximation methods. *International Journal of Health Geographics*, 8(23):1-9.
- Bollen, J., Mao, H., and Zeng, X. (2011) Twitter mood predicts the stock market. *Journal of Computer Science*, 2(11): 1-8.
- Bosley, J. C., Zhao, N.W., Hill, S., Shofer, F.S., Asch, D. A., Becker, L.B., and Merchant, R.M. (2013) Decoding twitter: surveillance and trends for cardiac arrest and resuscitation communication. *Resuscitation*, 84(2): 206-212.
- Buscaldi, D., Rosso, P., and Peris, P. (2006) Inferring geographical ontologies from multiple resources for geographical information retrieval. In *Proceedings of the 3rd ACM Workshop on Geographic Information Retrieval*, Seattle, WA, USA, 52-55.
- Buscaldi, D., Bessagnet, M. N., Royer, A., and Sallaberry, C. (2014) Using the semantics of texts for information retrieval: a concept- and domain relation-based approach. In *Proceedings of the 17th East European Conference on Advances in Databases and Information Systems: New Trends in Databases and Information Systems*, Genoa, Italy, 241: 257-266.
- Chen, J., Chen, H., Zheng, G., Pan, J.Z., and Wu, H. (2014) Big smog meets web science: smog disaster analysis based on social media and device data on the web. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, Republic and Canton of Geneva, Switzerland, 505-510.
- Cheng, Z., Caverlee, J., and Lee, K. (2010) You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, Toronto, Ontario, Canada, 759-768.

- Chasin, R., Woodward, D., Witmer, J., and Kalita, J. (2014) Extracting and displaying temporal and geographical entities from articles and historical events. *The Computer Journal*, 57(3): 403-426.
- Chowdhury, G. (2003) Natural language processing. *Annual Review of Information Science and Technology*, 37 (1): 51-89.
- Clark, A., Fox, C., and Lappin, S. (2010) *The Handbook of Computation Linguistics and Natural Language Processing*, Wiley_Blackwell, ISBN: 978-1-4051-5581-6.
- Clough, P. (2005) Extracting medtadata for spatially aware information retrieval on the internet. In *Proceedings of the 2005 workshop on Geographic Information Retrieval*, Bremen, Germany, 25-30.
- Corvey, W., Vieweg, S., Rood, T., and Martha, P. (2010) Twitter in mass emergency: what nlp techniques can contribute. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, Stroudsburg, PA, USA, 23-24.
- Crooks, A., Croitoru, A., Stefanidis, A., and Radzikowski, J. (2013) # Earthquake: twitter as a distributed sensor system. *Transactions in GIS*, 17(1): 124-147.
- Croitoru, A., Crooks, A., Radzikowski, J., and Stefanidis, A. (2013) Geosocial gauge: a system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science*, 27(12): 2483-2508.
- Cruz, I. F. and Xiao, H. (2005) The role of ontologies in data integration. *Engineering Intelligent Systems for Electrical Engineering and Communications*, 13(4): 245.
- Cruz, I. F., Ganesh, V. R., and Mirrezaei, S. I. (2013) Semantic extraction of geographic data from web tables for big data integration. In *Proceedings of the 7th Workshop on Geographic Information Retrieval*, Orlando, FL, USA, 19-26.
- Culotta, A. (2010b). Towards detecting influenza epidemics by analyzing twitter messages. In *Proceedings of the First Workshop on Social Media Analytics, SOMA'10*, NY, USA, 115-122.
- Delboni, T. M. and Borges, K. A. (2007) Semantic expansion of geographic web queries based on natural language positioning expressions. *Transactions in GIS*, 11(3): 377-397.
- Densham, I. and Reid, J. (2003) A geo-coding service encompassing a geo-parsing tool and integrated digital gazetteer service. In *Proceeding of the HLT-NAACL 2003 workshop on Analysis of geographic references*, Stroudsburg, PA, USA, 79-80.
- Derczynski, L. Ritter, A., Clarke, S., and Bontcheva, K. (2013) Twitter part-of-speech tagging for all: overcoming sparse and noisy data. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Hissar, Bulgaria, 198-206.
- Dredze, M., Paul, M.J., Bergsma, S., and Tran, H. (2013) Carmen: a twitter geolocation system with applications to public health. In *Processing of AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*, Bellevue, WA, 20-24.

- Dodge, S., Weibel, R., and Lautenschutz, A. K. (2008) Taking a systematic look at movement: developing a taxonomy of movement patterns. *Information Visualization*, 7(3): 240-252.
- Egenhofer, M. J. (2002) Toward the semantic geospatial web. In *Proceedings of the 10th ACM international symposium on Advances in Geographic Information Systems*, New York, NY, USA, 1-4.
- Egenhofer, M. J. and Mark, D. M. (1995) Navie geography. In Frank, A. U., and Kuhn, W., editors, *Spatial Information Theory: A Theoretical Basis for GIS. Lecture Notes in Computer Sciences*, 988:1-15.
- Eisenstein, J., O'Conner, B., Smith, N.A., and Xing, E.P. (2010) A latent variable model for geographic lexical variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 1277-1287.
- Fernandez, M., Cantador, I., Lopez, V., Vallet, D., Castells, P., and Motta, E. (2011) Semantically enhanced information retrieval: an ontology-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9 (4): 434-452.
- Fraser, A. and Marcu, D. (2013) Semi-supervised training for statistical word alignment. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA, 769-776.
- Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S., and Stange, H. (2013) Tracing the German centennial flood in the stream of tweets: first lessons learned. In *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, Orlando, FL, USA, 31-38.
- Gao, X.Y., Cao, J., He, Q., and Li, J. T. (2013) A novel method for geographical social event detection in social media. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, New York, NY, USA, 305-308.
- Garbin and Mani (2005) Disambiguating toponyms in news. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 363-370.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002) Sweetening ontologies with DOLCE. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management*, London, UK.
- Gelernter, J. and Mushegian, N. (2011) Geoparsing messages from microtext. *Transactions in GIS*, 15(6): 753-773.
- Guarino, N. (1998) Formal ontology and information systems. In (ed): Formal ontology in information systems. In *Proceedings of the 1st International Conference on Formal Ontology in Information Systems*, Trento, Italy, 3-15.
- Goodchild, M.F. and Hill, L.L. (2008) Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22(10): 1039-1044.

- Goodchild, M.F. and Glennon, J.A. (2010) Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3): 231-241.
- Habib, M.B. and Keulen, M. (2012) Improving toponym extraction and disambiguation using feedback loop. *Web Engineering*, 7387: 439-443.
- Imran, M., Elbassuomi, S., Castillo, C., Diaz, F., and Meier, P. (2013) Extracting information nuggets from disaster-related messages in social media. In *Proceedings of the 10th International ISCRAM Conference*, Baden-Baden, Germany, 1-10.
- Jackoway, A., Samet, H., and Sankaranarayanan, J. (2011) Identification of love news event using twitter. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, New York, NY, USA, 25-32.
- Janowicz, K. and Keßler, C. (2008) The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science*, 22(10): 1129-1157.
- Janowicz, K., Scheider, S., Pehle, T., and Hart, G. (2012) Geospatial semantics and linked spatiotemporal data-past, present, and future. *Semantic Web*, 3(4): 321-332.
- Joliveau, T., Ultsch, J., Royer, A., Sallaberry, C., Gaio, M., Beorchia, S., and Ny, P.A.L. (2011) Toward the spatial and temporal management of documents: the geotopia platform. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 46(3): 160-169.
- Jones, C.B., Purves, R.S., Clough, P.D., and Joho, H. (2008) Modelling vague places with knowledge from the web. *International Journal of Geographical Information Science*, 22(10): 1045-1065.
- Jones, C.B. and Purves, R. (2008) Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3): 219-228.
- Kalfoglou, Y. and Schorlemmer, M. (2003) Ontology mapping: the state of the art. *The Knowledge Engineering Review*, 18(1): 1-31.
- Kaplan, A. H. and Haenlein, M. (2011) The early bird catches the news: nine things you should know about micro-blogging. *Business Horizons*, 54(2): 105-113.
- Karimzadeh, M., Huang, W., Wallgrun, J. O., Hardisty, F., Pezanowski, S., Mitra, P., and MacEachren, A. M. (2013) GeoTxt: a web API to leverage place references in text. In *Proceeding of the 7th Workshop on Geographic Information Retrieval*, Orlando, FL, USA, 72-73.
- Kemp, Z., Tan, L., and Whalley, J. (2007) Interoperability for geospatial analysis: a semantics and ontology-based approach. In *Proceedings of the Eighteenth Australian Database Conference*, Ballarat, Victoria, Australia, 63: 83-92.
- Klippel, A., MacEachren, A. M., Mitra, P., Turton, I., Zhang, X., Jaiswal, A., Soon, K., Oyler, J., and Li, R. (2008) Geographic analysis of linguistically encoded movement patterns: a contextualized perspective. *GIScience 2008*, Park City, Utah, USA, 113-117.

- Kontopoulos, E., Berberidis, C., Dergiades, T., and Bassiliades, N. (2013) Ontology-based sentiment analysis of twitter posts. *Expert System with Applications*, 40(10): 4065-4074.
- Kornai, A. (2005) MetaCarta at GeoCLEF 2005. In *Proceedings of the Working Notes of the CLEF Workshop*, Vienna, Austria, 21-23.
- Kwan, M.P. (2000) Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8(1): 185-203.
- Kwan, M.P. and Ding, G.X. (2008) Geo-narrative geographic information systems for narrative analysis in qualitative and mixed-methods research. *The Professional Geographer*, 60(4): 443-465.
- Kwan, M.P., Richardson, D., Wang, D., and Zhou, C. (2014) *Space-Time Integration in Geography and GIScience*. Springer, Dordrecht.
- Lamb, A., Paul, M.J., and Dredze, M. (2013) Separating fact from fear: tracking flu infections on twitter. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, USA, 789-795.
- Larson, R.R. and Shaw, R. (2009) Mapping life events: temporal and geographic context for biographical information. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, Pittsburgh, PA, USA, 471-472.
- Lau, C. H. (2014) Detecting News Topics from Microblogs Using Sequential Patterns Mining. PhD thesis, Queensland University of Technology, http://eprints.qut.edu.au/68159/1/Cher%20Han_Lau_Thesis.pdf
- Laube, P., Dennis, T., Forer, P., and Walker, M. (2007) Movement beyond the snapshot—dynamic analysis of geospatial lifelines. *Journal of Computers, Environment and Urban Systems*, 31(5): 481-501.
- Lee, C. H., Wu, Ch. H., Yang, H. Ch., Wen, W. Sh., and Chiang, Ch. Y. (2013) Exploiting online social data in ontology learning for event tracking and emergency response. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Niagara Falls, ON, Canada, 1167-1174.
- Lee, S.S. and Wong, N.S. (2011) The clustering and transmission dynamics of pandemic influenza A (H1N1) 2009 cases in Hong Kong. *Journal of Infection*, 63(4): 274-280.
- Leetaru, K., Wang, S. W., Cao, G.F., Padmanabhan, A., and Shook, E. (2013). Mapping the global twitter heartbeat: the geography of twitter. *First Monday*, 18(5): doi:10.5210/fm.v18i5.4366.
- Leidner, J. L. and Lieberman, M.D. (2011) Detecting geographical references in the form of place names and associated spatial natural language. *SIGSPATIAL Special*, 3(2): 5-11.

- Li, L., Goodchild, M., and Xu, B. (2013) Spatial, temporal and socioeconomic patterns in the use of twitter and flicker. *Cartography and Geographic Information Science*, 40(2): 61-77.
- Li, Z., Yang, C., Sun, M., Li, J., Xu, C., and Huang, Q. (2013) A high performance web-based system for analyzing and visualizing spatiotemporal data for climate studies. *Web and Wireless Geographical Information Systems*, 7820: 190-198.
- Lieberman, M. D., Samet, H., Sankaranarayanan, J., and Sperling, J. (2007) STEWARD: demo of spatio-textual extraction on the web aiding the retrieval of documents. In *Proceedings of the 7th National Conference on Digital Government Research*, Philadelphia, PA, USA, 300-301.
- Lieberman, M.D., Samet, H., and Sankaranarayanan, J. (2010) Geotagging with local lexicons to build indexes for textually-specified spatial data. In *Proceedings of the 26th IEEE International Conference on Data Engineering*, Long Beach, CA, 201-212.
- Lieberman, M.D. and Samet, H. (2012) Adaptive context features for toponym resolution in streaming news. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, Portland, OR, USA, 731-740.
- Ling, X. and Weld, D. S. (2010) Temporal information extraction. In *Proceedings of the Twenty Fifth National Conference on Artificial Intelligence*, San Francisco, USA.
- Lingad, J., Karimi, S., and Yin, J. (2013) Location extraction from disaster-related microblogs. In *Proceedings of the 22nd international conference on World Wide Web companion*, Geneva, Switzerland, 1017-1020.
- Liu, B. (2010) Sentiment analysis and subjectivity. *Natural Language Processing Handbook*, Boca Raton, FL: Chemical Rubber Company Press.
- MacEachren, A.M., Robinson, A., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J., and Mitra, P. (2011) Geo-twitter analytics: applications in crisis management. In *Proceedings of 25th International Cartographic Conference*, Paris, France, 1-8.
- Machado, I. M. R., Alencar, R. O. D., Campos, R. D. O., and Clodoveu, A., D. (2011) An ontological gazetteer and its application for place name disambiguation in text. *Journal of the Brazilian Computer Science*, 17(4): 267-279.
- Mark, D., Skupin, A., and Smith, B. (2001) Features, objects, and other things ontological distinctions in the geographic domain. *Lecture Notes in Computer Science*, 2205: 489-502.
- Mani, I., Doran, C., Harris, D., Hitzeman, J., Quimby, R., Richer, J., Wellner, B., Mardis, S., and Clancy, S. (2010) SpatialML: annotation scheme, resources, and evaluation. *Lang Resources & Evaluation*, 44 (3): 263-280.
- Mani, I., Anderson, D. and Hitzeman, J. (2006) A framework for interring spatial locations and relationships from text. *National Center for Geographic Information & Analysis (NCGIA) Digital Gazetteer Research and Practice Workshop*, <http://ncgia.ucsb.edu/projects/nga/docs/mani-paper.pdf>

- Manning, C.D., Raghavan, P., and Schütze, H. (2008) Introduction to Information Retrieval. Cambridge: Cambridge University Press.
- Miller, H. J. (1991) Modeling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems*, 5(3): 281-301.
- Mitchell, T. M. (1997) Machine Learning. New York, NY: McGraw-Hill Science/Engineering/Math.
- Neubig, G., Matsubayashi, Y., Hagiwara, M., and Murakami, K. (2011) Safety information mining what can nlp do in a disaster. In *Proceedings of the Fifth International Joint Journal Conference on Natural Language Processing*, Chiang Mai, Thailand, 965-973.
- Nielsen, F. A. (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, Heraklion, Crete, Greece, 93-98.
- Noy, N.F. (2004) Semantic integration: a survey of ontology-based approaches. *Special Interest Group on Management of Data Record*, 33(4): 65-70.
- Hughes, A. L. and Palen, L. (2009) Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3): 1471-4825.
- Pan, R. K., Kaski, K., and Fortunato, S. (2011) World citation and collaboration networks: uncovering the role of geography in science. *Scientific reports* 2: 902.
- Pak, A. and Paroubek, P. (2010) Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 1320-1326.
- Pasley, R.C., Clough, P. D., and Sanderson, M. (2007) Geo-tagging for imprecise regions of different sizes. In *Proceedings of the 4th ACM workshop on Geographical information retrieval*, Arlington, VA, USA, 77-82.
- Paul, M. and Dredze, M. (2011) You are what you tweet: analyzing twitter for public health. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Catalonia, Spain, 265- 272.
- Pham, T.P., Ng, H.T., and Lee, W.S. (2005) Word sense disambiguation with semi-supervised learning. *Computational Linguistics and Intelligent Text Processing*, 3406: 238-241.
- Price, C., Nakagaki, N., Hitt, K., and Clawges, R. (2003) Mining GIRAS: improving on a national treasure of land use data. In *Proceedings of the 23rd ESRI International User Conference*, Redlands, CA, USA, 11.
- Purves, R.S., Clough, P., Jones, C.B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A.K., Vaid, S., and Yang, B. (2007) The design and implementation of spirit: a spatially aware search engine for information retrieval on the internet. *International Journal of Geographical Information Science*, 21(7): 717-745.

- Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., and Lazo, M. (2003) The timebank corpus. In *Proceedings of Corpus Linguistics*, Lancaster, UK, 647-656.
- Pustejovsky, J., Lee, K., Bunt, H., and Romary, L. (2010) Iso timeml: an international standard for semantic annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Valletta, Malta, 19-21.
- Quincey, E. and Kostkova, P. (2010) Early warning and outbreak detection using social networking websites: The potential of twitter. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 27: 21–24. Springer Berlin, Heidelberg: Springer.
- Raubal, M., Miller, H. J., and Bridwel, S. (2004) User-centred time geography for location-based services. *Human Geography*, 86(4): 245-266.
- Resnik, P. and Lin, J. (2010) Evaluation of NLP Systems. *The Handbook of Computational Linguistics and Natural Language Processing*, 57, 271-295. Oxford, UK: Wiley-Blackwell.
- Ritter, A., Clark, S., Mausam, and Etzioni, O. (2011) Name entity recognition in tweets: an experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK, 1524-1534.
- Rushton, G., Armstrong, M.P., Gittler, J., Greene, B. R., Pavlik, C.E., West, M.M., and Zimmerman, D.L. (2010) Geocoding Health Data: the Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Boca Raton, FL: Chemical Rubber Company Press.
- Saggion, H., Funk, A., Maynard, D., and Bontcheva, K. (2007) Ontology-based information extraction for business intelligence. *The Semantic Web Lecture Notes in Computer Science*, 4825: 843-856.
- Saif, H., He, Y., and Alani, H. (2012) Semantic sentiment analysis of twitter. In *Proceedings of 11th International Semantic Web Conference, ISWC 2012*, Boston, MA, USA, 7649: 508-524.
- Sankaranarayanan, J. H., Teitler, B. E., Lieberman, M. D., and Sperling, J. (2009) Twitterstand: news in tweets. In *Proceedings of 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, WA, USA, 42-51.
- Sakaki, T., Okazaki, M., and 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*, Raleigh, NC, USA, 851-860.
- Saquete, E. and Pustejovsky, J. (2011) Automatic transformation from tides to timeml annotation. *Lang Resources and Evaluation*, 45 (4): 495-523.
- Sawsaa, A. F. and Lu, Joan (2012) Developing a domain ontology of information science (OIS). In *Proceedings of the International Conference on Information Society*, London, UK, 447-452.

- Scanfeild, D. Scanfeild, V., Larson, E. L. (2010) Dissemination of health information through social networks: twitter and antibiotics. *American Journal of Infection Control*, 38(3): 182-188.
- Scharl, A., Stern, H., and Weichselbraun, A. (2008) Annotating and visualizing location data in geospatial web applications. In *Proceedings of the First International Workshop on Location and the Web*, Beijing, China, 65-68.
- Schorlemmer, W. M. and Kalfoglou, Y. K. (2008) Institutionalising ontology-based semantic integration. *Applied Ontology*, 3(3): 131-150.
- Schuurman, N. (2013) Tweet me your talk: geographical learning and knowledge production 2.0. *The Professional Geographer*, 65(3): 369-377.
- Shaw, S.L. and Yu, H. B. (2009) A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space. *Journal of Transport Geography*, 17(2): 141-149.
- Shi, G. and Barker, K. (2011) Extraction of geospatial information on the web for GIS applications. In *Proceedings of the 10th IEEE International Conference on Cognitive Informatics*, Banff, Alberta, Canada, 41-48.
- Signorni, A., Segre, A. M., and Polgreen, P. M. (2011) The use of twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS ONE*, 6(5): 1-10.
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Sitter, A.D., Calders, T., and Daelemans, W. (2004) A formal framework for evaluation of information extraction. <http://www.wis.win.tue.nl/tcalders/pubs/DESITTERTR04.pdf>.
- Stefanidis, A., Crooks, A., and Radzikowski, J. (2013) Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78 (2): 319-338.
- Stewart Hornsby, K. and Li, N. C. (2009) Conceptual framework for modeling dynamic paths from natural language expressions. *Transactions in GIS*, 13(1): 27-45.
- Stewart Hornsby, K. and Wang, W. (2010) Representing dynamic phenomena based on spatiotemporal information extracted from web documents. In *Proceedings of the Sixth International Conference on Geographic Information Science*, Zurich, Switzerland. http://www.giscience2010.org/pdfs/paper_131.pdf.
- Stewart, K. and Wang, W. (2013) The Role of Spatiotemporal Theories for Information Retrieval and Modeling Human-Environment Interactions. *COSIT 2013 Workshop*, Scarborough, North Yorkshire, United Kingdom.
- Stewart, K., Fan, J. C., and White, E. (2013) Thinking about space-time connections: spatiotemporal scheduling of individual activities. *Transactions in GIS*, 17(6): 791-807.
- Stock, K., Pasley, R.C., Gardener, Z., Brindley, P., Morley, J., and Cialone, C. (2013) Creating a corpus of geospatial natural language. In *Proceedings of the Eleventh International Conference on Spatial Information Theory*, COSIT 2013, Scarborough, UK, 279-298.

- Strotgen, J., Gertz, M., and Popv, P. (2010) Extraction and exploration of spatiotemporal information in documents. In *Proceedings of the 6th workshop on Geographic Information Retrieval*, GIR 2010, Zurich, Switzerland, 1-8.
- Strotgen, J. and Gertz, M. (2010) TimeTrails: a system for exploring spatio-temporal information in documents. In *Proceedings of the Very Large Data Bases Endowment*, Singapore, 3(2): 1569-1572.
- Teitler, B. E., Lieberman, M. D., Panozzo, D., Sankaranarayanan, J., Samet, H., and Sperling, J. (2008) NewsStand: a new view on news. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Irvine, California, USA, 18: 1-10.
- Teitler, B.E., Sankaranarayanan, J., Samet, H., and Adelfio, M. D. (2014) Online document clustering using GPUs. *New Trends in Databases and Information Systems, Advances in Intelligent Systems and Computing*, 241: 245-254.
- Theruaykt, M. and Claramunt, C. (2013) On the representation of life trajectories. *COSIT 2013 Workshop*, Scarborough, North Yorkshire, United Kingdom.
- Tsou, M. H. and Leitner, M. (2013) Visualization of social media: seeing a mirage or a message. *Cartography and Geographic Information Science*. 40(2): 55-60.
- Tsou, M. H., Kim, I.H., Wandersee, S., Lusher,D., An, L., Spitzberg, B., Gupta, D., Gawron, J., Smith, J., Yang, J. A., and Han, S. Y. (2014) Mapping ideas from cyberspace to real space: visualizing the spatial context of keywords from web page search results. *International Journal of Digital Earth*, 7(4) 316-335.
- Tsou, M. H., Yang, J. A., Lusher, D., Han, S., Spitzberg, B., Gawron, J. M., Gupta, D., and An, L. (2013) Mapping social activities and concepts with social media (Twitter) and web Search engines (Yahoo and Bing): a case study in 2012 U.S. presidential election. *Cartography and Geographic Information Science*, 40(4), 337-348.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2011) Election forecasts with twitter: how 140 characters reflect the political landscape. *Social Science Computer Review*, 29 (4): 402–418.
- Yang, M.C. and Rim, H.C. (2014) Identifying interesting twitter contents using topical analysis. *Expert Systems with Applications*, 41(9): 4330-4336.
- Yuan, M. (2013) Some “missing” ideas about time and space-time in GIScience. *COSIT 2013 Workshop*, Scarborough, North Yorkshire, United Kingdom.
- Yuan, M. and Stewart Hornsby, K. (2008) Computation and Visualization for Understanding Dynamics in Geographic Domains-a Research Agenda. London, New York: Taylor and Francis Group.
- Yu, H. B. (2006) Spatio-temporal GIS design for exploring interactions of human activities. *Cartography and Geographic Information Science*, 33(1): 3-19.
- Vasardani, M., Winter, S., and Richter, K.F. (2013) Location place names from place descriptions. *International Journal of Geographical Information Science*, 27(12): 2509-2532.

- Viera, A.J. and Garrett, J.M. (2005) Understanding interobserver agreement: the kappa statistic. *Family Medicine*, 37(5): 360-363.
- Vieweg, S., Hughes, A.L., Starbird, K., and 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*, Atlanta, GA, USA, 1079-1088.
- Volz, R., Kleb, J., and Mueller, W. (2007) Towards ontology-based disambiguation of geographical identifiers. In *Proceedings of WWW2007 Workshop I3: Identity, Identifiers, Identification, Entity-Centric Approaches to Information and Knowledge Management on the Web*, Banff, Canada. http://ceur-ws.org/Vol-249/submission_132.pdf.
- Wang, S. (2013) CyberGIS: blueprint for integrated and scalable geospatial software ecosystems. *International Journal of Geographical Information Science*, 27(11): 2119-2121
- Wang, W. and Stewart, K. (2013) Extracting spatiotemporal and semantic information from documents. In *Proceedings of the Seventh Workshop on Geographic Information Retrieval*, Orlando, Florida, USA, 17-18.
- Wang, W. and Stewart, K. (2014) Creating spatiotemporal semantic maps from web text documents. In: M-P. Kwan, D. Richardson, D. Wang, and C. Zhou, *Space-Time Integration in Geography and GIScience: Research Frontiers in the US and China*, Springer, Dordrecht.
- Welty, C. (2003) Ontology research. *AI Magazine*, 24(3): 11-12.
- Wiegand, N. and Garcia, C. (2007) A task-based ontology approach to automate geospatial data retrieval. *Transaction in GIS*, 11(3): 355-376.
- Worboys, M. F. and Stewart Hornsby, K. (2004) From objects to events: GEM, the geospatial event model. In *Proceedings of the Third International Conference on Geographic Information Sciences*, Adelphi, MD, USA, 3234: 327-344.
- Woodruff, A. G. and Plaunt, C. (1994) GIPSY: automated geographic indexing of text documents. *Journal of the American Society for Information Science*, 45(9): 645-655.
- Zhang, Q., Jin, P., Lin, S., and Yue, L. (2012) Extracting focused locations for web pages. *Web-Age Information Management Lecture Notes in Computer Science*, 7142:76-89.

APPENDIX

```

<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdfs2 "rdfs:" >
  <!ENTITY owl "http://gate.ac.uk/owlim#" >
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY psys "http://proton.semanticweb.org/protonsys#" >
  <!ENTITY pext "http://proton.semanticweb.org/protonext#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY protont "http://proton.semanticweb.org/2005/04/protont#" >
  <!ENTITY protonu "http://proton.semanticweb.org/2005/04/protonu#" >
  <!ENTITY protons "http://proton.semanticweb.org/2005/04/protons#" >
  <!ENTITY protonkm "http://proton.semanticweb.org/2005/04/protonkm#" >
]>
<rdf:RDF xmlns="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.w3.org/2000/01/rdf-schema"
  xmlns:owlim="http://gate.ac.uk/owlim#"
  xmlns:protons="http://proton.semanticweb.org/2005/04/protons#"
  xmlns:protonu="http://proton.semanticweb.org/2005/04/protonu#"
  xmlns:protont="http://proton.semanticweb.org/2005/04/protont#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:psys="http://proton.semanticweb.org/protonsys#"
  xmlns:rdfs2="rdfs:"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:pext="http://proton.semanticweb.org/protonext#"
  xmlns:protonkm="http://proton.semanticweb.org/2005/04/protonkm#">
  <owl:Ontology rdf:about="http://www.owl-ontologies.com/unnamed.owl"/>
  <!--
  //////////////////////////////////////
  //
  // Annotation properties
  //
  //////////////////////////////////////
  -->
  <!--
  //////////////////////////////////////
  //
  // Object Properties
  //
  //////////////////////////////////////
  -->
  <!-- http://gate.ac.uk/owlim#Dispatch -->
  <owl:ObjectProperty rdf:about="&owlim;Dispatch">
    <label>Dispatch</label>
  </owl:ObjectProperty>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#DevelopsInto -->

<owl:ObjectProperty rdf:about="&rdfs;DevelopsInto"/>

<!-- http://www.w3.org/2000/01/rdf-schema#Partof -->
<owl:ObjectProperty rdf:about="&rdfs;Partof">
  <rdf:type rdf:resource="&owl;TransitiveProperty"/>
  <range rdf:resource="&rdfs;HazardManagement"/>
  <domain rdf:resource="&rdfs;HazardManagement"/>
  <range rdf:resource="&rdfs;NaturalHazard"/>
  <domain rdf:resource="&rdfs;NaturalHazard"/>
</owl:ObjectProperty>
<!--
////////////////////////////////////
//
// Classes
//
////////////////////////////////////
-->
<!-- http://gate.ac.uk/owlim#Drought -->
<owl:Class rdf:about="&owlim;Drought">
  <label>Drought</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Earthquake -->
<owl:Class rdf:about="&owlim;Earthquake">
  <label>Earthquake</label>
  <subClassOf rdf:resource="&rdfs;GeologicHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Flood -->
<owl:Class rdf:about="&owlim;Flood">
  <label>Flood</label>
  <subClassOf rdf:resource="&rdfs;HydrologicHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Happening -->
<owl:Class rdf:about="&owlim;Happening"/>
<!-- http://gate.ac.uk/owlim#IceStorm -->
<owl:Class rdf:about="&owlim;IceStorm">
  <label>Icestorm</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Landslide -->
<owl:Class rdf:about="&owlim;Landslide">
  <label>Landslide</label>
  <subClassOf rdf:resource="&rdfs;GeologicHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Mudflow -->
<owl:Class rdf:about="&owlim;Mudflow">
  <label>Mudflow</label>
  <subClassOf rdf:resource="&owlim;Landslide"/>

```

```

</owl:Class>
<!-- http://gate.ac.uk/owlim#Mudslide -->
<owl:Class rdf:about="&owlim;Mudslide">
  <label>Mudslide</label>
  <subClassOf rdf:resource="&owlim;Landslide"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Object -->
<owl:Class rdf:about="&owlim;Object"/>
<!-- http://gate.ac.uk/owlim#Rockslide -->
<owl:Class rdf:about="&owlim;Rockslide">
  <label>Rockslide</label>
  <subClassOf rdf:resource="&owlim;Landslide"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Tornado -->
<owl:Class rdf:about="&owlim;Tornado">
  <label>Tornado</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Tsunami -->
<owl:Class rdf:about="&owlim;Tsunami">
  <label>Tsunami</label>
  <subClassOf rdf:resource="&rdfs;HydrologicHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Volcanic_eruption -->
<owl:Class rdf:about="&owlim;Volcanic_eruption">
  <label>Volcanic_eruption</label>
  <subClassOf rdf:resource="&rdfs;GeologicHazard"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#Weather -->
<owl:Class rdf:about="&owlim;Weather">
  <label>Weather</label>
  <subClassOf rdf:resource="&owlim;Happening"/>
</owl:Class>
<!-- http://gate.ac.uk/owlim#volcanic_gas_release -->
<owl:Class rdf:about="&owlim;volcanic_gas_release">
  <label>volcanic_gas_release</label>
  <subClassOf rdf:resource="&owlim;Volcanic_eruption"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema# -->
<owl:Class rdf:about="http://www.w3.org/2000/01/rdf-schema#">
  <subClassOf rdf:resource="&rdfs;Precipitation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Agent -->
<owl:Class rdf:about="&rdfs;Agent">
  <subClassOf rdf:resource="&owlim;Object"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Airline -->
<owl:Class rdf:about="&rdfs;Airline">
  <subClassOf rdf:resource="&rdfs;Company"/>
</owl:Class>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#Airplane -->
<owl:Class rdf:about="&rdfs;Airplane">
  <subClassOf rdf:resource="&rdfs;Vehicle"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Airport -->
<owl:Class rdf:about="&rdfs;Airport">
  <subClassOf rdf:resource="&rdfs;Airline"/>
  <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#AirportClosure -->
<owl:Class rdf:about="&rdfs;AirportClosure">
  <subClassOf rdf:resource="&rdfs;Closure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#AirportReopen -->
<owl:Class rdf:about="&rdfs;AirportReopen">
  <subClassOf rdf:resource="&rdfs;TransportationRecovery"/>
  <seeAlso>airports reopen</seeAlso>
  <seeAlso>airport reopens</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#AmusementPark -->
<owl:Class rdf:about="&rdfs;AmusementPark">
  <subClassOf rdf:resource="&rdfs;RecreationalFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#AnabaticWind -->
<owl:Class rdf:about="&rdfs;AnabaticWind">
  <label>anabatic_wind</label>
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Apartment -->
<owl:Class rdf:about="&rdfs;Apartment">
  <subClassOf rdf:resource="&rdfs;Building"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Bank -->
<owl:Class rdf:about="&rdfs;Bank">
  <subClassOf rdf:resource="&rdfs;Company"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BankDamage -->
<owl:Class rdf:about="&rdfs;BankDamage">
  <subClassOf rdf:resource="&rdfs;CommercialDamage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Bay -->
<owl:Class rdf:about="&rdfs;Bay">
  <subClassOf rdf:resource="&rdfs;WaterRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Beach -->
<owl:Class rdf:about="&rdfs;Beach">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Blizzard -->
<owl:Class rdf:about="&rdfs;Blizzard">
  <subClassOf rdf:resource="&rdfs;Snow"/>

```

```

    <subClassOf rdf:resource="&rdfs;SnowStorm"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Breeze -->
<owl:Class rdf:about="&rdfs;Breeze">
    <label>breeze</label>
    <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Bridge -->
<owl:Class rdf:about="&rdfs;Bridge">
    <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BridgeClosure -->
<owl:Class rdf:about="&rdfs;BridgeClosure">
    <subClassOf rdf:resource="&rdfs;Closure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Building -->
<owl:Class rdf:about="&rdfs;Building">
    <subClassOf rdf:resource="&rdfs;Facility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BuildingReconstruction -->
<owl:Class rdf:about="&rdfs;BuildingReconstruction">
    <subClassOf rdf:resource="&rdfs;ResidentialRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BuredHome -->
<owl:Class rdf:about="&rdfs;BuredHome">
    <subClassOf rdf:resource="&rdfs;ResidentialDamage"/>
    <seeAlso>burned homes</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Bus -->
<owl:Class rdf:about="&rdfs;Bus">
    <subClassOf rdf:resource="&rdfs;Vehicle"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BusCancellation -->
<owl:Class rdf:about="&rdfs;BusCancellation">
    <subClassOf rdf:resource="&rdfs;Cancellation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BusDelay -->
<owl:Class rdf:about="&rdfs;BusDelay">
    <subClassOf rdf:resource="&rdfs;Delay"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#BusStation -->
<owl:Class rdf:about="&rdfs;BusStation">
    <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Camp -->
<owl:Class rdf:about="&rdfs;Camp">
    <subClassOf rdf:resource="&rdfs;RecreationalFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Canal -->
<owl:Class rdf:about="&rdfs;Canal">
    <subClassOf rdf:resource="&rdfs;HydrographicStructure"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Cancellation -->
<owl:Class rdf:about="&rdfs;Cancellation">
  <subClassOf rdf:resource="&rdfs;TransportationImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Canyon -->
<owl:Class rdf:about="&rdfs;Canyon">
  <subClassOf rdf:resource="&rdfs;Valley"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Cape -->
<owl:Class rdf:about="&rdfs;Cape">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Cave -->
<owl:Class rdf:about="&rdfs;Cave">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Channel -->
<owl:Class rdf:about="&rdfs;Channel">
  <subClassOf rdf:resource="&rdfs;WaterRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Charity -->
<owl:Class rdf:about="&rdfs;Charity">
  <subClassOf rdf:resource="&rdfs;Organization"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#City -->
<owl:Class rdf:about="&rdfs;City">
  <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#CleanUp -->
<owl:Class rdf:about="&rdfs;CleanUp">
  <subClassOf rdf:resource="&rdfs;ResidentialRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ClimateHazard -->
<owl:Class rdf:about="&rdfs;ClimateHazard">
  <label>Meteorological_event</label>
  <subClassOf rdf:resource="&rdfs;NaturalHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Closure -->
<owl:Class rdf:about="&rdfs;Closure">
  <subClassOf rdf:resource="&rdfs;FacilityImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Coast -->
<owl:Class rdf:about="&rdfs;Coast">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#CommercialDamage -->
<owl:Class rdf:about="&rdfs;CommercialDamage">
  <subClassOf rdf:resource="&rdfs;CommercialImpact"/>
</owl:Class>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#CommercialImpact -->
<owl:Class rdf:about="&rdfs;CommercialImpact">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#CommercialRecovery -->
<owl:Class rdf:about="&rdfs;CommercialRecovery">
  <subClassOf rdf:resource="&rdfs;HazardRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#CommunityResponse -->
<owl:Class rdf:about="&rdfs;CommunityResponse">
  <subClassOf rdf:resource="&rdfs;HazardResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Company -->
<owl:Class rdf:about="&rdfs;Company">
  <subClassOf rdf:resource="&rdfs;Organization"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Continent -->
<owl:Class rdf:about="&rdfs;Continent">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Country -->
<owl:Class rdf:about="&rdfs;Country">
  <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#County -->
<owl:Class rdf:about="&rdfs;County">
  <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Creek -->
<owl:Class rdf:about="&rdfs;Creek">
  <subClassOf rdf:resource="&rdfs;WaterRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Dealer -->
<owl:Class rdf:about="&rdfs;Dealer">
  <subClassOf rdf:resource="&rdfs;Building"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Delay -->
<owl:Class rdf:about="&rdfs;Delay">
  <subClassOf rdf:resource="&rdfs;TransportationImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Delta -->
<owl:Class rdf:about="&rdfs;Delta">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Desert -->
<owl:Class rdf:about="&rdfs;Desert">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Donation -->
<owl:Class rdf:about="&rdfs;Donation">
  <subClassOf rdf:resource="&rdfs;HazardRecovery"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Drizzle -->
<owl:Class rdf:about="&rdfs;Drizzle">
  <subClassOf rdf:resource="&rdfs;Rain"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#DustStorm -->
<owl:Class rdf:about="&rdfs;DustStorm">
  <label>dust_storm</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#EastWind -->
<owl:Class rdf:about="&rdfs;EastWind">
  <label>east_wind</label>
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#EducationalOrganization -->
<owl:Class rdf:about="&rdfs;EducationalOrganization">
  <subClassOf rdf:resource="&rdfs;Organization"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ElementaryClosure -->
<owl:Class rdf:about="&rdfs;ElementaryClosure">
  <subClassOf rdf:resource="&rdfs;SchoolClosure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#EmergencyResponse -->
<owl:Class rdf:about="&rdfs;EmergencyResponse">
  <subClassOf rdf:resource="&rdfs;HazardResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Evacuation -->
<owl:Class rdf:about="&rdfs;Evacuation">
  <subClassOf rdf:resource="&rdfs;CommunityResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Facility -->
<owl:Class rdf:about="&rdfs;Facility">
  <subClassOf rdf:resource="&rdfs;Place"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FacilityImpact -->
<owl:Class rdf:about="&rdfs;FacilityImpact">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FacilityRecovery -->
<owl:Class rdf:about="&rdfs;FacilityRecovery">
  <subClassOf rdf:resource="&rdfs;HazardRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Fatality -->
<owl:Class rdf:about="&rdfs;Fatality">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>
  <seeAlso>die</seeAlso>
  <seeAlso>kill</seeAlso>
  <seeAlso>dead</seeAlso>
  <seeAlso>killed</seeAlso>
  <seeAlso>death</seeAlso>

```



```

    <seeAlso>death roll</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FireFighter -->
<owl:Class rdf:about="&rdfs;FireFighter">
    <subClassOf rdf:resource="&rdfs;EmergencyResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FlightCancellation -->
<owl:Class rdf:about="&rdfs;FlightCancellation">
    <subClassOf rdf:resource="&rdfs;Cancellation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FlightDelay -->
<owl:Class rdf:about="&rdfs;FlightDelay">
    <subClassOf rdf:resource="&rdfs;Delay"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FloodedArea -->
<owl:Class rdf:about="&rdfs;FloodedArea">
    <subClassOf rdf:resource="&rdfs;UtilityDamage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FoodSupply -->
<owl:Class rdf:about="&rdfs;FoodSupply">
    <subClassOf rdf:resource="&rdfs;CommunityResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FoodtruckDelivery -->
<owl:Class rdf:about="&rdfs;FoodtruckDelivery">
    <subClassOf rdf:resource="&rdfs;TransportationResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Forecast -->
<owl:Class rdf:about="&rdfs;Forecast">
    <subClassOf rdf:resource="&rdfs;HazardPrediction"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Forest -->
<owl:Class rdf:about="&rdfs;Forest">
    <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ForestWildfire -->
<owl:Class rdf:about="&rdfs;ForestWildfire">
    <label>alert</label>
    <subClassOf rdf:resource="&rdfs;WildfireHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#FuelShortage -->
<owl:Class rdf:about="&rdfs;FuelShortage">
    <subClassOf rdf:resource="&rdfs;UtilityShortage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#GasLeak -->
<owl:Class rdf:about="&rdfs;GasLeak">
    <subClassOf rdf:resource="&rdfs;UtilityDamage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#GeologicHazard -->
<owl:Class rdf:about="&rdfs;GeologicHazard">
    <label>Geological_hazard</label>
    <subClassOf rdf:resource="&rdfs;NaturalHazard"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Government -->
<owl:Class rdf:about="&rdfs;Government">
  <subClassOf rdf:resource="&rdfs;Organization"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Group -->
<owl:Class rdf:about="&rdfs;Group">
  <subClassOf rdf:resource="&rdfs;Agent"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Gulf -->
<owl:Class rdf:about="&rdfs;Gulf">
  <subClassOf rdf:resource="&rdfs;WaterRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Gust -->
<owl:Class rdf:about="&rdfs;Gust">
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Hail -->
<owl:Class rdf:about="&rdfs;Hail">
  <subClassOf rdf:resource="&rdfs;Ice"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HailStorm -->
<owl:Class rdf:about="&rdfs;HailStorm">
  <label>hailstorm</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Harbor -->
<owl:Class rdf:about="&rdfs;Harbor">
  <subClassOf rdf:resource="&rdfs;HydrographicStructure"/>
  <subClassOf rdf:resource="&rdfs;WaterRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HazardImpact -->
<owl:Class rdf:about="&rdfs;HazardImpact">
  <subClassOf rdf:resource="&rdfs;HazardManagement"/>
  <subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="rdfs:Partof"/>
      <owl:someValuesFrom rdf:resource="&owl;Thing"/>
    </owl:Restriction>
  </subClassOf>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HazardManagement -->
<owl:Class rdf:about="&rdfs;HazardManagement">
  <subClassOf rdf:resource="&owlim;Happening"/>
  <subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="rdfs:Partof"/>
      <owl:someValuesFrom rdf:resource="&rdfs;NaturalHazard"/>
    </owl:Restriction>
  </subClassOf>
</owl:Class>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#HazardPrediction -->
<owl:Class rdf:about="&rdfs;HazardPrediction">
  <subClassOf rdf:resource="&rdfs;HazardManagement"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HazardRecovery -->
<owl:Class rdf:about="&rdfs;HazardRecovery">
  <subClassOf rdf:resource="&rdfs;HazardManagement"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HazardResponse -->
<owl:Class rdf:about="&rdfs;HazardResponse">
  <subClassOf rdf:resource="&rdfs;HazardManagement"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HeavyRain -->
<owl:Class rdf:about="&rdfs;HeavyRain">
  <subClassOf rdf:resource="&rdfs;Rain"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HeavySnow -->
<owl:Class rdf:about="&rdfs;HeavySnow">
  <subClassOf rdf:resource="&rdfs;Snow"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HeavyWind -->
<owl:Class rdf:about="&rdfs;HeavyWind">
  <label>heavy_wind</label>
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HighSchool -->
<owl:Class rdf:about="&rdfs;HighSchool">
  <subClassOf rdf:resource="&rdfs;School"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HighSchoolClosure -->
<owl:Class rdf:about="&rdfs;HighSchoolClosure">
  <subClassOf rdf:resource="&rdfs;SchoolClosure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HighwayClosure -->
<owl:Class rdf:about="&rdfs;HighwayClosure">
  <subClassOf rdf:resource="&rdfs;RoadClosure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HomeReconstruction -->
<owl:Class rdf:about="&rdfs;HomeReconstruction">
  <subClassOf rdf:resource="&rdfs;ResidentialRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Hospital -->
<owl:Class rdf:about="&rdfs;Hospital">
  <subClassOf rdf:resource="&rdfs;Building"/>
  <subClassOf rdf:resource="&rdfs;Organization"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Hotel -->
<owl:Class rdf:about="&rdfs;Hotel">
  <subClassOf rdf:resource="&rdfs;Building"/>
</owl:Class>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#Hour -->
<owl:Class rdf:about="&rdfs;Hour">
  <subClassOf rdf:resource="&rdfs;TemporalUnit"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Hurricane -->
<owl:Class rdf:about="&rdfs;Hurricane">
  <label>hurricane</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HydrographicStructure -->
<owl:Class rdf:about="&rdfs;HydrographicStructure">
  <subClassOf rdf:resource="&rdfs;Facility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#HydrologicHazard -->
<owl:Class rdf:about="&rdfs;HydrologicHazard">
  <label>Hydrological_hazard</label>
  <subClassOf rdf:resource="&rdfs;NaturalHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Ice -->
<owl:Class rdf:about="&rdfs;Ice">
  <subClassOf rdf:resource="&rdfs;Precipitation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Injury -->
<owl:Class rdf:about="&rdfs;Injury">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#InsuranceCompany -->
<owl:Class rdf:about="&rdfs;InsuranceCompany">
  <subClassOf rdf:resource="&rdfs;Company"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Island -->
<owl:Class rdf:about="&rdfs;Island">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Jungle -->
<owl:Class rdf:about="&rdfs;Jungle">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#KatabaticWind -->
<owl:Class rdf:about="&rdfs;KatabaticWind">
  <label>Katabatic_wind</label>
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#KindergartenClosure -->
<owl:Class rdf:about="&rdfs;KindergartenClosure">
  <subClassOf rdf:resource="&rdfs;SchoolClosure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Lake -->
<owl:Class rdf:about="&rdfs;Lake">
  <subClassOf rdf:resource="&rdfs;WaterRegion"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#LandRegion -->
<owl:Class rdf:about="&rdfs;LandRegion">
  <subClassOf rdf:resource="&rdfs;Place"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#LightSnow -->
<owl:Class rdf:about="&rdfs;LightSnow">
  <subClassOf rdf:resource="&rdfs;Snow"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Lightning -->
<owl:Class rdf:about="&rdfs;Lightning">
  <subClassOf rdf:resource="&rdfs;ThurnderStorm"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Lock -->
<owl:Class rdf:about="&rdfs;Lock">
  <subClassOf rdf:resource="&rdfs;HydrographicStructure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#MedicalCare -->
<owl:Class rdf:about="&rdfs;MedicalCare">
  <subClassOf rdf:resource="&rdfs;EmergencyResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#MedicalCompany -->
<owl:Class rdf:about="&rdfs;MedicalCompany">
  <subClassOf rdf:resource="&rdfs;Company"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#MiddleSchool -->
<owl:Class rdf:about="&rdfs;MiddleSchool">
  <subClassOf rdf:resource="&rdfs;School"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#MiddleSchoolClosure -->
<owl:Class rdf:about="&rdfs;MiddleSchoolClosure">
  <subClassOf rdf:resource="&rdfs;SchoolClosure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#MilitaryArea -->
<owl:Class rdf:about="&rdfs;MilitaryArea">
  <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Minute -->
<owl:Class rdf:about="&rdfs;Minute">
  <subClassOf rdf:resource="&rdfs;TemporalUnite"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Montain -->
<owl:Class rdf:about="&rdfs;Montain">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#NaturalHazard -->
<owl:Class rdf:about="&rdfs;NaturalHazard">
  <label>Natural_hazard</label>
  <subClassOf rdf:resource="&owlim;Happening"/>
  <subClassOf>
    <owl:Restriction>

```

```

        <owl:onProperty rdf:resource="rdfs:Partof"/>
        <owl:someValuesFrom rdf:resource="&owl;Thing"/>
    </owl:Restriction>
</subClassOf>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#NorthWind -->
<owl:Class rdf:about="&rdfs;NorthWind">
    <label>north_wind</label>
    <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Oasis -->
<owl:Class rdf:about="&rdfs;Oasis">
    <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Ocean -->
<owl:Class rdf:about="&rdfs;Ocean">
    <subClassOf rdf:resource="&rdfs;Sea"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#OffShoreWind -->
<owl:Class rdf:about="&rdfs;OffShoreWind">
    <label>off-shore_wind</label>
    <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#OnShoreWind -->
<owl:Class rdf:about="&rdfs;OnShoreWind">
    <label>on-shore_wind</label>
    <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Organization -->
<owl:Class rdf:about="&rdfs;Organization">
    <subClassOf rdf:resource="&rdfs;Group"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Outage -->
<owl:Class rdf:about="&rdfs;Outage">
    <subClassOf rdf:resource="&rdfs;UtilityDamage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Peninsula -->
<owl:Class rdf:about="&rdfs;Peninsula">
    <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Person -->
<owl:Class rdf:about="&rdfs;Person">
    <subClassOf rdf:resource="&rdfs;Agent"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Pipeline -->
<owl:Class rdf:about="&rdfs;Pipeline">
    <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Place -->
<owl:Class rdf:about="&rdfs;Place">
    <subClassOf rdf:resource="&owlim;Object"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Plain -->
<owl:Class rdf:about="&rdfs;Plain">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Plane -->
<owl:Class rdf:about="&rdfs;Plane">
  <subClassOf rdf:resource="&rdfs;TransportationResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Plateau -->
<owl:Class rdf:about="&rdfs;Plateau">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Police -->
<owl:Class rdf:about="&rdfs;Police">
  <subClassOf rdf:resource="&rdfs;EmergencyResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#PoliticalRegion -->
<owl:Class rdf:about="&rdfs;PoliticalRegion">
  <subClassOf rdf:resource="&rdfs;Place"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#PowerRestore -->
<owl:Class rdf:about="&rdfs;PowerRestore">
  <subClassOf rdf:resource="&rdfs;UtilityRecovery"/>
  <seeAlso>restore power</seeAlso>
  <seeAlso>backup power</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#PowerShortage -->
<owl:Class rdf:about="&rdfs;PowerShortage">
  <subClassOf rdf:resource="&rdfs;UtilityShortage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Precipitation -->
<owl:Class rdf:about="&rdfs;Precipitation">
  <label>precipitation_process</label>
  <subClassOf rdf:resource="&owlim;Weather"/>
  <seeAlso></seeAlso>
  <seeAlso>precipitation</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Province -->
<owl:Class rdf:about="&rdfs;Province">
  <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RailCancellation -->
<owl:Class rdf:about="&rdfs;RailCancellation">
  <subClassOf rdf:resource="&rdfs;Cancellation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RailDamage -->
<owl:Class rdf:about="&rdfs;RailDamage">
  <subClassOf rdf:resource="&rdfs;TransportationDamage"/>
</owl:Class>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#Railway -->
<owl:Class rdf:about="&rdfs;Railway">
  <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Rain -->
<owl:Class rdf:about="&rdfs;Rain">
  <subClassOf rdf:resource="&rdfs;Precipitation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RainStorm -->
<owl:Class rdf:about="&rdfs;RainStorm">
  <label>rain_storm</label>
  <subClassOf rdf:resource="&rdfs;Storm"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RecreationalFacility -->
<owl:Class rdf:about="&rdfs;RecreationalFacility">
  <subClassOf rdf:resource="&rdfs;Facility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Relief -->
<owl:Class rdf:about="&rdfs;Relief">
  <subClassOf rdf:resource="&rdfs;EmergencyResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RemoveDebris -->
<owl:Class rdf:about="&rdfs;RemoveDebris">
  <subClassOf rdf:resource="&rdfs;CommunityResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Rescue -->
<owl:Class rdf:about="&rdfs;Rescue">
  <subClassOf rdf:resource="&rdfs;EmergencyResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Reservoir -->
<owl:Class rdf:about="&rdfs;Reservoir">
  <subClassOf rdf:resource="&rdfs;HydrographicStructure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ResidentialDamage -->
<owl:Class rdf:about="&rdfs;ResidentialDamage">
  <subClassOf rdf:resource="&rdfs;ResidentialImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ResidentialImpact -->
<owl:Class rdf:about="&rdfs;ResidentialImpact">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ResidentialRecovery -->
<owl:Class rdf:about="&rdfs;ResidentialRecovery">
  <subClassOf rdf:resource="&rdfs;HazardRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RestaurantDamage -->
<owl:Class rdf:about="&rdfs;RestaurantDamage">
  <subClassOf rdf:resource="&rdfs;CommercialDamage"/>
</owl:Class>

<!-- http://www.w3.org/2000/01/rdf-schema#RestoreBus -->

```



```

<owl:Class rdf:about="&rdfs;RestoreBus">
  <subClassOf rdf:resource="&rdfs;TransportationRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RestoreRail -->
<owl:Class rdf:about="&rdfs;RestoreRail">
  <subClassOf rdf:resource="&rdfs;TransportationRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#River -->
<owl:Class rdf:about="&rdfs;River">
  <subClassOf rdf:resource="&rdfs;Stream"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RoadClosure -->
<owl:Class rdf:about="&rdfs;RoadClosure">
  <subClassOf rdf:resource="&rdfs;Closure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RoadDamage -->
<owl:Class rdf:about="&rdfs;RoadDamage">
  <subClassOf rdf:resource="&rdfs;TransportationDamage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RoadReconstruction -->
<owl:Class rdf:about="&rdfs;RoadReconstruction">
  <subClassOf rdf:resource="&rdfs;FacilityRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#RoofDamage -->
<owl:Class rdf:about="&rdfs;RoofDamage">
  <subClassOf rdf:resource="&rdfs;ResidentialDamage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SandStorm -->
<owl:Class rdf:about="&rdfs;SandStorm">
  <label>sandstorm</label>
  <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#School -->
<owl:Class rdf:about="&rdfs;School">
  <subClassOf rdf:resource="&rdfs;Building"/>
  <subClassOf rdf:resource="&rdfs;EducationalOrganization"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SchoolClosure -->
<owl:Class rdf:about="&rdfs;SchoolClosure">
  <subClassOf rdf:resource="&rdfs;Closure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SchoolReopen -->
<owl:Class rdf:about="&rdfs;SchoolReopen">
  <subClassOf rdf:resource="&rdfs;FacilityRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Sea -->
<owl:Class rdf:about="&rdfs;Sea">
  <subClassOf rdf:resource="&rdfs;HydrographicStructure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Second -->
<owl:Class rdf:about="&rdfs;Second">

```

```

    <subClassOf rdf:resource="&rdfs;TemporalUnit"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Shelter -->
<owl:Class rdf:about="&rdfs;Shelter">
    <subClassOf rdf:resource="&rdfs;CommunityResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Sleet -->
<owl:Class rdf:about="&rdfs;Sleet">
    <subClassOf rdf:resource="&rdfs;Ice"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Snow -->
<owl:Class rdf:about="&rdfs;Snow">
    <subClassOf rdf:resource="&rdfs;Precipitation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SnowStorm -->
<owl:Class rdf:about="&rdfs;SnowStorm">
    <subClassOf rdf:resource="&rdfs;ClimateHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SouthWind -->
<owl:Class rdf:about="&rdfs;SouthWind">
    <label>south_wind</label>
    <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Spring -->
<owl:Class rdf:about="&rdfs;Spring">
    <subClassOf rdf:resource="&rdfs;Stream"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Stadium -->
<owl:Class rdf:about="&rdfs;Stadium">
    <subClassOf rdf:resource="&rdfs;Building"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#State -->
<owl:Class rdf:about="&rdfs;State">
    <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Storm -->
<owl:Class rdf:about="&rdfs;Storm">
    <label>storm</label>
    <subClassOf rdf:resource="&owlim;Weather"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Stream -->
<owl:Class rdf:about="&rdfs;Stream">
    <subClassOf rdf:resource="&rdfs;HydrographicStructure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Street -->
<owl:Class rdf:about="&rdfs;Street">
    <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#StreetClosure -->
<owl:Class rdf:about="&rdfs;StreetClosure">
    <subClassOf rdf:resource="&rdfs;RoadClosure"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#StrongWind -->
<owl:Class rdf:about="&rdfs;StrongWind">
  <label>strong_wind</label>
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SubwayCancellation -->
<owl:Class rdf:about="&rdfs;SubwayCancellation">
  <subClassOf rdf:resource="&rdfs;Cancellation"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#SubwayRunning -->
<owl:Class rdf:about="&rdfs;SubwayRunning">
  <subClassOf rdf:resource="&rdfs;TranportationRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Supercell -->
<owl:Class rdf:about="&rdfs;Supercell">
  <subClassOf rdf:resource="&rdfs;ThurnderStorm"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TemporalUnite -->
<owl:Class rdf:about="&rdfs;TemporalUnite">
  <subClassOf rdf:resource="&rdfs;Time"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#ThurnderStorm -->
<owl:Class rdf:about="&rdfs;ThurnderStorm">
  <subClassOf rdf:resource="&rdfs;Storm"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Time -->
<owl:Class rdf:about="&rdfs;Time">
  <subClassOf rdf:resource="&owlim;Object"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TimeZone -->
<owl:Class rdf:about="&rdfs;TimeZone">
  <subClassOf rdf:resource="&rdfs;Time"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Tower -->
<owl:Class rdf:about="&rdfs;Tower">
  <subClassOf rdf:resource="&rdfs;Building"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Train -->
<owl:Class rdf:about="&rdfs;Train">
  <subClassOf rdf:resource="&rdfs;Vehicle"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TranportationDamage -->
<owl:Class rdf:about="&rdfs;TranportationDamage">
  <subClassOf rdf:resource="&rdfs;TranportationImpact"/>
  <seeAlso>destroy</seeAlso>
  <seeAlso>Destroyed</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TranportationImpact -->
<owl:Class rdf:about="&rdfs;TranportationImpact">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>

```

```

</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TransportationRecovery -->
<owl:Class rdf:about="&rdfs;TransportationRecovery">
  <subClassOf rdf:resource="&rdfs;HazardRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TransportFacility -->
<owl:Class rdf:about="&rdfs;TransportFacility">
  <subClassOf rdf:resource="&rdfs;Facility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#TransportationResponse -->
<owl:Class rdf:about="&rdfs;TransportationResponse">
  <subClassOf rdf:resource="&rdfs;HazardResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Transtation -->
<owl:Class rdf:about="&rdfs;Transtation">
  <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Truck -->
<owl:Class rdf:about="&rdfs;Truck">
  <subClassOf rdf:resource="&rdfs;Vehicle"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Tundra -->
<owl:Class rdf:about="&rdfs;Tundra">
  <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Tunnel -->
<owl:Class rdf:about="&rdfs;Tunnel">
  <subClassOf rdf:resource="&rdfs;TransportFacility"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#University -->
<owl:Class rdf:about="&rdfs;University">
  <subClassOf rdf:resource="&rdfs;School"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#UniversityClosure -->
<owl:Class rdf:about="&rdfs;UniversityClosure">
  <subClassOf rdf:resource="&rdfs;SchoolClosure"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#UrbanDistrict -->
<owl:Class rdf:about="&rdfs;UrbanDistrict">
  <subClassOf rdf:resource="&rdfs;PoliticalRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#UtilityDamage -->
<owl:Class rdf:about="&rdfs;UtilityDamage">
  <subClassOf rdf:resource="&rdfs;UtilityImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#UtilityImpact -->
<owl:Class rdf:about="&rdfs;UtilityImpact">
  <subClassOf rdf:resource="&rdfs;HazardImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#UtilityRecovery -->
<owl:Class rdf:about="&rdfs;UtilityRecovery">

```

```

    <subClassOf rdf:resource="&rdfs;HazardRecovery"/>
    <seeAlso>Restoration</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#UtilityShortage -->
<owl:Class rdf:about="&rdfs;UtilityShortage">
    <subClassOf rdf:resource="&rdfs;UtilityImpact"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Valley -->
<owl:Class rdf:about="&rdfs;Valley">
    <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Vehicle -->
<owl:Class rdf:about="&rdfs;Vehicle">
    <subClassOf rdf:resource="&owlim;Object"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Victim -->
<owl:Class rdf:about="&rdfs;Victim">
    <subClassOf rdf:resource="&rdfs;Person"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Volcano -->
<owl:Class rdf:about="&rdfs;Volcano">
    <subClassOf rdf:resource="&rdfs;Mountain"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Volunteer -->
<owl:Class rdf:about="&rdfs;Volunteer">
    <subClassOf rdf:resource="&rdfs;CommunityResponse"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Volunteer -->
<owl:Class rdf:about="&rdfs;Volunteer">
    <subClassOf rdf:resource="&rdfs;Person"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WanaterFall -->
<owl:Class rdf:about="&rdfs;WanaterFall">
    <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Warning -->
<owl:Class rdf:about="&rdfs;Warning">
    <subClassOf rdf:resource="&rdfs;HazardPrediction"/>
    <seeAlso>alarms</seeAlso>
    <seeAlso>alert</seeAlso>
    <seeAlso>alarm</seeAlso>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WaterBank -->
<owl:Class rdf:about="&rdfs;WaterBank">
    <subClassOf rdf:resource="&rdfs;LandRegion"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WaterPump -->
<owl:Class rdf:about="&rdfs;WaterPump">
    <subClassOf rdf:resource="&rdfs;UtilityDamage"/>
</owl:Class>

```

```

<!-- http://www.w3.org/2000/01/rdf-schema#WaterRegion -->
<owl:Class rdf:about="&rdfs;WaterRegion">
  <subClassOf rdf:resource="&rdfs;Place"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WaterRestore -->
<owl:Class rdf:about="&rdfs;WaterRestore">
  <subClassOf rdf:resource="&rdfs;UtilityRecovery"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WaterShortage -->
<owl:Class rdf:about="&rdfs;WaterShortage">
  <subClassOf rdf:resource="&rdfs;UtilityShortage"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WestWind -->
<owl:Class rdf:about="&rdfs;WestWind">
  <label>west_wind</label>
  <subClassOf rdf:resource="&rdfs;Wind"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#WildfireHazard -->
<owl:Class rdf:about="&rdfs;WildfireHazard">
  <label>Wildfire_hazard</label>
  <subClassOf rdf:resource="&rdfs;NaturalHazard"/>
</owl:Class>
<!-- http://www.w3.org/2000/01/rdf-schema#Wind -->
<owl:Class rdf:about="&rdfs;Wind">
  <label>wind</label>
  <subClassOf rdf:resource="&owlim;Weather"/>
</owl:Class>
<!--
////////////////////////////////////
//
// Annotations
//
////////////////////////////////////
-->
<rdf:Description rdf:about="&owlim;Time">
  <label>Time</label>
</rdf:Description>
<rdf:Description rdf:about="&owlim;Hazard_level">
  <label>Hazard_level</label>
</rdf:Description>
<rdf:Description rdf:about="&owlim;Area">
  <label>Area</label>
</rdf:Description>
</rdf:RDF>
<!-- Generated by the OWL API (version 3.2.0.1502) http://owlapi.sourceforge.net -->

```