

Article

A Geo-Event-Based Geospatial Information Service: A Case Study of Typhoon Hazard

Yu Zhang ^{1,2,3}, Wenzhou Wu ¹, Qi Wang ^{1,2,3} and Fenzhen Su ^{1,*}

¹ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; zhangyu@reis.ac.cn (Y.Z.); wuwz@reis.ac.cn (W.W.); wangq.15b@igsrr.ac.cn (Q.W.)

² College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

³ School of Geography, Beijing Normal University, Beijing 100875, China

* Correspondence: sufz@reis.ac.cn; Tel.: +86-10-6488-8956

Academic Editor: Jesus Martinez-Frias

Received: 4 January 2017; Accepted: 28 March 2017; Published: 31 March 2017

Abstract: Social media is valuable in propagating information during disasters for its timely and available characteristics nowadays, and assists in making decisions when tagged with locations. Considering the ambiguity and inaccuracy in some social data, additional authoritative data are needed for important verification. However, current works often fail to leverage both social and authoritative data and, on most occasions, the data are used in disaster analysis after the fact. Moreover, current works organize the data from the perspective of the spatial location, but not from the perspective of the disaster, making it difficult to dynamically analyze the disaster. All of the disaster-related data around the affected locations need to be retrieved. To solve these limitations, this study develops a geo-event-based geospatial information service (GEGIS) framework and proceeded as follows: (1) a geo-event-related ontology was constructed to provide a uniform semantic basis for the system; (2) geo-events and attributes were extracted from the web using a natural language process (NLP) and used in the semantic similarity match of the geospatial resources; and (3) a geospatial information service prototype system was designed and implemented for automatically retrieving and organizing geo-event-related geospatial resources. A case study of a typhoon hazard is analyzed here within the GEGIS and shows that the system would be effective when typhoons occur.

Keywords: geo-event; social media; ontology; geospatial information service

1. Introduction

Recent disasters have drawn attention to the vulnerability of human populations and infrastructure, and the extremely high cost of recovering from the damage they have caused [1]. Information needs to be immediately propagated to inform people of the severe impacts of the disaster, such as the damage, injury, and loss of life. Social media, such as Twitter or Flickr, with their timely and available information-delivering capabilities, can be useful for exchanging messages in such situations. Millions of social data are generated to support diverse activities during all stages of disaster management [2]. In order to achieve the benefits of social data, some projects and systems have been designed and implemented on disaster awareness and assessment [3–6]. Although progress is satisfying, some challenges still exist: the lack of a capability to leverage the authoritative (derived from the authority agencies) and social data to support timely and accurate information delivery during an emergency is still a problem; moreover, the geospatial data are generally organized from the perspective of the spatial location, making it difficult to dynamically analyze the actual disaster. All disaster-related data around the affected locations need to be retrieved.

A disaster can generally be considered a geo-event, which is a natural or social phenomenon taking place on the Earth's surface with definite spatiotemporal attributes (e.g., a hazard event, a military exercise, or a terrorist attack happening at some location and lasting for some time). However, most people are concerned with natural disasters, e.g., typhoons, floods, and earthquakes. Once a geo-event is detected, a mechanism should be triggered to retrieve and organize both authoritative and social data; thus, users, such as scientists and the general public, could use the results for their own purposes.

Applying social data to situational awareness and disaster management when the data are geo-referenced (e.g., location, place names) has become a growing research area [7–9], which is a must for aiding in the understanding and interpretation of the collected data [2]. However, finding meaningful information from unstructured social data for analysis is still challenging [7]. Assisted by the NLP and semantic web technologies, social data can be interpreted and used in disaster mapping to make disaster assessments [10–12].

Detecting disasters implied in the social data is the primary task for further analysis and management of disasters. Social data generally contains meaningless and event-irrelevant messages that weaken the detection performance to a certain degree [13]. In order to discover valuable information from the social data, machine learning algorithms are mostly adopted. Thus, detection approaches can be divided into supervised and unsupervised approaches. Common supervised approaches, such as support vector machine (SVM) [14] and naïve Bayes [15], are often based on the established corpus, which contains samples of specific event types [16]. Meanwhile, unsupervised approaches mostly rely on clustering approaches to discover topics from messages, such as latent Dirichlet allocation (LDA) [17,18] and the graph-based approach [19]. Considering the fact that disaster types are defined in advance, disaster detection involves determining the event type of each message. In this case, supervised approaches are more suitable. However, a mechanism is still needed to retrieve and organize the corresponding data for disaster analysis and management after detection.

Academic research into disaster management has largely focused on the application of social data to disaster response in recent years [20]. One important factor of the social data is the ability to provide more timely observations; the public can act as an early warning system for disasters [5,21–23]. Meanwhile, for the characteristics in each stage of a disaster (e.g., before, during, and after disasters), different standards should be proposed to improve situational awareness and damage assessment. For instance, Chae et al. [7] analyzed public behavior responses from social media. Deng et al. [9] proposed an index model to improve situational awareness and damage assessments.

Taking advantages of data mining technologies, some implicit relations can be achieved and analyzed to reduce the influence of a disaster. Some secondary disasters could thereby be avoided. For instance, Albuquerque et al. [24] presented an approach to enhance the identification of relevant messages from social media, relating relations between social data and the geographic features of disasters. Bakillah [25] used a cluster algorithm with semantic similarity to process complex social graphs extracted from Twitter, and discussed correlations between clustering patterns of people and hazard regions.

Moreover, some situational awareness and management systems have been developed to collect useful geo-referenced social data to provide information to the public and support crisis management in recent years [26], such as MicroMappers [3], CrisisTracker [27], LITMUS [28], Petajakarta [5], and EAIMS [29]. However, these systems are generally focused on one kind of disaster and trigger the corresponding data organizing from the perspective of spatial location, but they have failed to leverage authoritative data and dynamically respond to other disaster types, thus limiting their usage.

On the other hand, more work could be done when combined with the authoritative data to enhance the disaster situational awareness and management [4,30–32], such as forest fires [33], flood risks [32], and air pollution [34]. Generally, most authoritative data are derived from authority agencies, e.g., NOAA (National Oceanic and Atmospheric Administration) and USGS (United States Geological Survey), and contain many categories, such as hydrology, meteorology, and oceanography. Therefore, they are often regarded as credible resources for scientific research, but challenges still exist

in providing timely information and analysis results of the occurring disaster. However, for the timely and available characteristics of the social data, the weakness can be reduced by the supplement of social data in disaster management [32,35].

Dorn et al. [36] analyzed the effectiveness of orthophotos, LiDAR data, official land use data, OpenStreetMap data, and CORINE Land Cover data in flood simulations, and confirmed the usability of crowdsourced data. Jongman et al. [6] utilized satellite observations of water coverage and flood-related social media to support rapid disaster response, and suggested their usage in early disaster warning.

However, these studies are mainly focused on the verification of social data assisted by authoritative data, and the social data are almost collected in a static form that cannot be fitted to the dynamic combination of the data to deliver timely and accurate messages to the public when a disaster occurs.

To solve the limitations mentioned in the previous paragraphs, this study presents a geo-event-based geospatial information service (GEGIS) framework and constructs an ontology to support geo-event recognition and geospatial resource retrieval. This ontology implements almost automatic processing and thematic organization of geospatial resources. An SVM algorithm is utilized to calculate the thematic probability, which is the likelihood of a message crawled from social media to the specific topic, and determines its category. Then, named entity recognition (NER) technology, which is a subtask of information extraction that seeks to locate and classify named entities in text into predefined categories, is utilized to recognize the geo-event and its attributes in the message. Moreover, an ontology-based semantic match algorithm is presented to implement geospatial resource retrieval and organization. To verify the feasibility and effectiveness of the framework, a geospatial information service prototype system was designed and implemented. The system can automatically retrieve and organize the geo-event-related geospatial resources, depending on attributes of the geo-event. The main contribution of the study involves the following aspects:

- (1) We leverage the authoritative and social data to provide timely and accurate geo-event-related resources.
- (2) We present an approach from the perspective of the geo-event, which means each geo-event will contain a series of related, topical, classified data based on the established ontology, and the spatial extent of the event will cover all of the impact regions. Therefore, the dynamic analysis of the event and effective response to the event change can be easily achieved.

The remainder of the study is organized as follows: Section 2 presents the GEGIS and addresses the approach to implement the framework based on the ontology; implementation of the approach and result analysis are described in Section 3, and a typhoon hazard case study is illustrated; Section 4 provides a discussion of the work; and, finally, Section 5 presents conclusions and suggestions for future work.

2. GEGIS Framework

To implement the GEGIS, we regard the ontology as a key component and integrate the technologies of NLP, NER, and the semantic web into a unified framework (Figure 1). The GEGIS mainly contains three modules: the social data service module, which can monitor, crawl, and classify social data, from which structured information is achieved; the geospatial service module, which can monitor and crawl geographical information services from the web; and the service manager module, which can retrieve and organize geospatial resources.

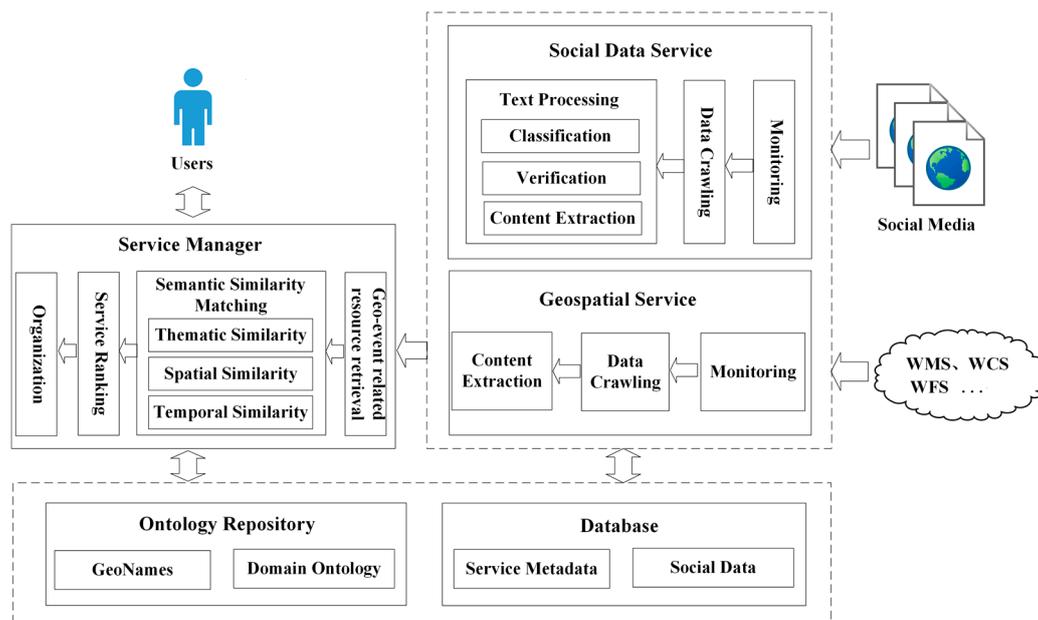


Figure 1. The geo-event-based geospatial information service (GEGIS) framework.

GEGIS processes are briefly described as follows:

- (1) Social media is monitored and crawled by a web crawler every five minutes from Sina Weibo [37], which is a popular social media site similar to Twitter in China, and records normal life in words or pictures. Each message with its published time and location (if they exist) is recorded in a text file.
- (2) Assisted by a predefined classification corpus, the thematic probability distribution of the message is calculated using an SVM to determine its corresponding category. Then, NPL tools and regexes are used to extract the geo-event with its attributes from the message. Finally, approaches are adopted to exclude unreliable and redundant data, as is discussed in Section 2.2.
- (3) The geographical information services are monitored, and their metadata are crawled and stored, mainly using the GetCapabilities function. The metadata contains some descriptions and parameters of the service, such as the description and the spatial extent of the layer.
- (4) Based on the ontology, geo-event-related resources are retrieved by calculating the semantic similarity of extracted attributes from messages, as is discussed in Section 2.3.
- (5) The latest geo-event, with its corresponding geospatial resources, is delivered to users and shown at the top rank.

There are, mainly, three aspects in the GEGIS framework: ontology construction, text processing, and semantic matching. These are discussed, respectively, in the following sections.

2.1. Constructing Ontology

The ontology is the semantic foundation supporting retrieval and organization of related geospatial resources, and a key part of the GEGIS. The ontology is constructed using Protégé Editor, which was designed and implemented by Stanford University. To maximize reuse of domain terminology, a unified domain concept set is constructed with the aid of maritime domain terminology from the Semantic Web For Earth and Environmental Terminology [38]. Natural disaster classification terminology provided by the national government [39] is used as the geo-event-related terminology (Figure 2a). Meanwhile, a Chinese place name database is constructed to provide basic place names. A geo-event related ontology is thereby constructed.

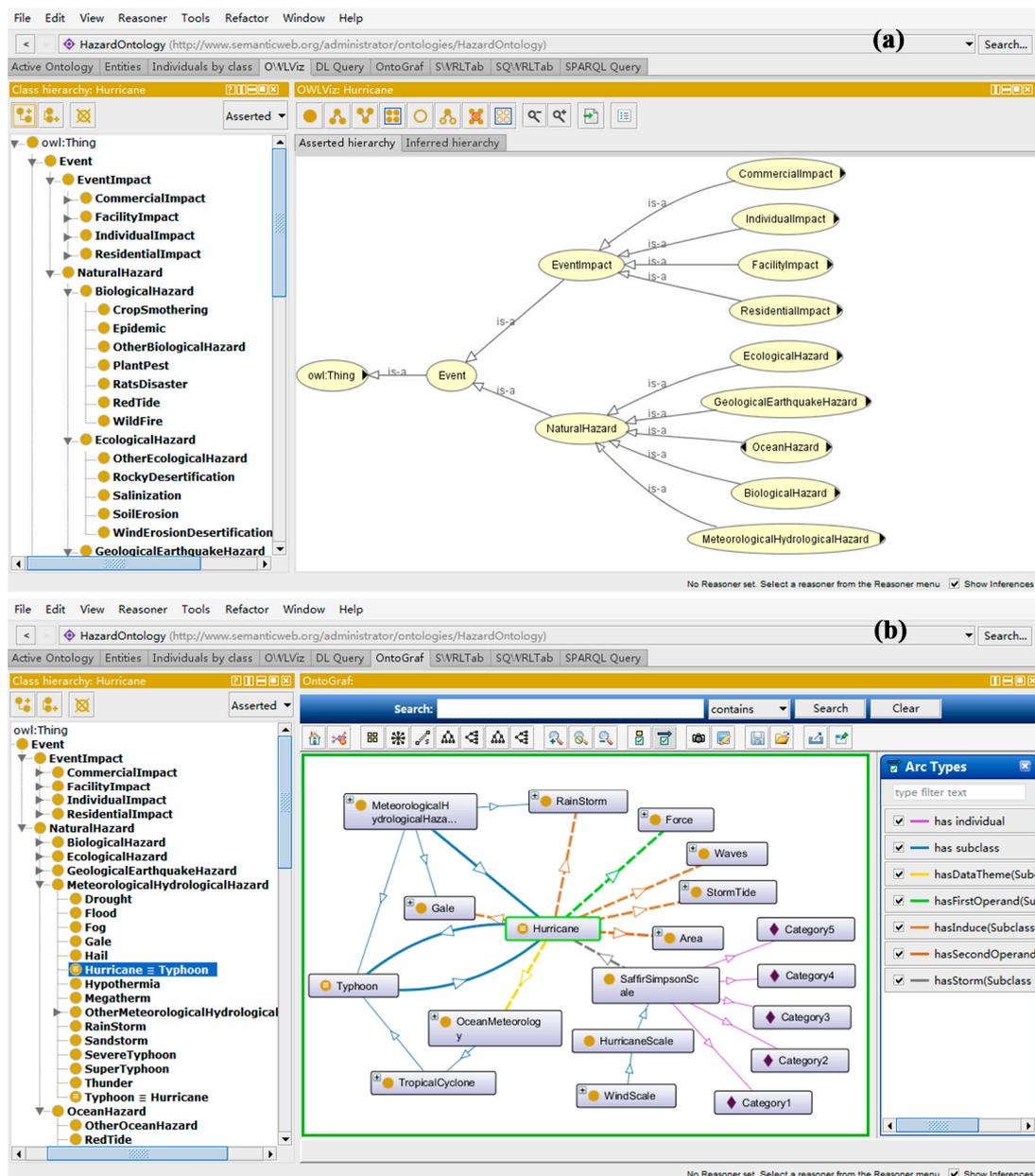


Figure 2. Ontology construction. (a) Fragment of ontology; (b) sublevel ontology on the typhoon.

Figure 2a shows a fragment of the ontology, mainly referring to natural hazards, and Figure 2b shows a sublevel ontology on the typhoon (hurricane). The top ontology is the abstraction of all concepts and contains two parts: one is the natural hazard event, and the other is the event impact. The natural hazard events can be divided into five subclasses: MeteorologicalHydrologicalHazard, OceanicHazard, GeologicalEarthquakeHazard, BiologicalHazard, and EcologicalHazard. Each subclass can be further divided, such as into typhoon, landslide, and flood. Each event has several attributes: the hasInduce property is used to establish relationships between different events, such as landslides and rainstorms caused by a typhoon; the hasDataTheme property is used to establish relationships between events and resource themes, such as the correlation between a typhoon hazard and wind field data. The event impacts can be divided into four subclasses: FacilityImpact, IndividualImpact, CommercialImpact, and ResidentialImpact. For example, typhoons can induce economic loss. Assisted by the ontology, the retrieval and thematic organization of geospatial resources can be achieved.

2.2. Processing Text

Text processing uses NLP technology to convert unstructured text into structured text, which can be recognized and processed by a computer. The procedure contains three stages: text classification, extraction, and verification.

2.2.1. Text Classification

For messages crawled from the web, identifying a geo-event implicit in the message is an essential premise for implementing content extraction, resource retrieval, and organization. Generally, a message cannot exceed 140 words; thus, sentences in the message often concisely express the specific purpose. However, a large proportion of messages are irrelevant to geo-events, although some of them are geo-referenced. In order to identify useful information from crawled messages, a text classification algorithm should be adopted. We used an SVM, which is a popular supervised learning algorithm that can be utilized to classify the text using a training dataset in the corpus. An SVM generally uses a kernel function to map the original data to a high-dimensional space where linear classification of two types of data can be conducted. Moreover, a corpus containing classified texts are established to support the classification. Assisted by LibSVM [40], classification of a geo-event can be achieved and is discussed in Section 3.

2.2.2. Text Extraction

After text classifying, a further process is necessary to extract useful information from a message, e.g., time, location, and geo-event type. In most cases, a geo-event is just simply mentioned in the message without any other words to describe the disaster. Therefore, important information, such as impact scope and level, could not be obtained from these messages, and a useful message about the disaster should contain words that are helpful for situational awareness and assessment, e.g., power failure and road flooding.

Generally, words in a message are partitioned by using a natural language processing and information retrieval (NLPIR) [41] package, and the geo-event and locations are then extracted by using regexes, which are mostly dependent on sentence structure after using NLPIR. Therefore, the geo-event with its name, time, location, and spatial relation is extracted. For instance, an original typhoon-related message after text partitioning is shown as follows:

“**今 年**/*t***第1**/*m***号**/*q***台 风**/*n*“/*wyz***尼 伯 特**/*nrf*“/*wyy* (*wkz***台 风**/*n***级**/*q*) *wky*, *wd***今 天**/*t***00***时*/*t***中 心**/*n***位 于**/*v***北 纬**/*b***23.3**/*m***度**/*qv*, *wd***东 经**/*b***119.7**/*m***度**/*qv*, *wd*...” (Translation: The centre of No.1 Nepal Typhoon (typhoon level) is 23.3°N, 119.7°E at zero o'clock . . .),

where the original words are shown in bold, and the italic words are the identifications of parts of speeches predefined in NLPIR, i.e., noun (/n), verb (/v), time word (/t), numeral word (/m), punctuation (/wyz, /wyy, /wkz, /wky, and /wd), quantifier (/q and /qv), and distinguishing words (/b). In this case, for an identified geo-event type, its name can be extracted by using regex “**台 风**/*n*([4e00-9fa5201c201d]+)/*wyz*[4E00-9FA5]+/*nrf*([4e00-9fa5201c201d]+)/*wyy*,” and the location can be extracted by using “[4e00-9fa5]+/*b*([0-9]+.[0-9]*)/*m*[4e00-9fa5]/*qv*”. Since a simple gazetteer in the NLPIR package can be modified by the developer, the place name will be extracted directly if a word is identified as a location word. For a message containing a geo-event without any location, if a published location exists, it will be used as a place name, or it will be excluded.

Other factors, such as time and spatial relations, are also extracted by using regexes. For the near real-time requirement of the system, time words are required to be in the same time interval. Therefore, time words, including the published time and the extracted time, are used. The extracted time from the message is identified first, and the published time will only be used if the extracted time does not exist. Then, the identified time is judged to exclude the messages outside the required time. As for the spatial relation, for instance, original words expressing spatial relations after text partitioning are as follows:

“东南/f方向/n40/m公里/q” (Translation: 40 km southeast)

Thus, the regex can be expressed as “[\u4e00-\u9fa5]+/f[\u4e00-\u9fa5]+/n[0-9]+/m[\u4e00-\u9fa5]+/q”. Moreover, the impact words of the geo-event are extracted by using a keyword matching algorithm based on the established ontology. Finally, the text is stored in the database with extracted attributes, such as location, time, event name, and impacts.

2.2.3. Text Verification

Reliability and authenticity issues are the single greatest challenge for the use of social media [42,43]. Since ambiguous, inaccurate, or confused expressions in the messages published by non-professional users about the disaster exist, poor information quality exists such that a message cannot be fully leveraged by scientists and decision-makers. Therefore, some approaches are adopted to reduce problematic messages.

For the data accumulated in a time interval, the geo-event with the maximum frequency will be selected, and the problematic data will then be processed.

Data redundancy, which is mostly caused by reposting messages, often exists in the accumulated data, reducing the usefulness of the data. Therefore, the study introduced an information fingerprint algorithm [44] to reduce the data redundancy. Another challenge lying in the verification is the ambiguous nature of the geo-event name. On some occasions, different names indicate the same geo-event, which is often caused by non-uniform naming between different government agencies, such as Tropical Storm Mekkhala, named Amang in the Philippines (2015). In this case, an approach combines the geographic approach and similarity approach to determine the similarity of the geo-event. Historical data stored in the database are retrieved using Lucene [45] under the condition of the geo-event names, and the similarity of historical traces and impact regions in the same time interval are calculated. The similarity mainly depends on the overlapping area of the two regions. If the calculation scores are similar, the names will be considered as the same geo-event.

However, two types of errors should be further noticed, i.e., false positives and false negatives, which, respectively, express a rumor of an event and an absence of information about the existence of an event [1]. For a false positive, the location of the geo-event is used as an important factor to exclude some errors, based on the hypothesis that an event location should not exceed the impact region of the event in the specific time interval. Then, the authoritative social data around the suspicious location, mainly from government agencies and credible users, are used to verify the data quality based on the similarity match of the keywords in the messages using an ontology. The false positive error will, thereby, be reduced. Nonetheless, for a false negative, effective non-artificial approaches are still lacking in the current work, and an interactive interface will be added in the future to improve the credibility of the system.

After those processes, words representing the time, location, and geo-event are used to build triples, i.e., query = (geo-event, time, location). In order to simplify the semantic similarity matching, location is set to the minimum bounding box, which is a minimum rectangle containing all of the places extracted from related documents in the time interval based on the ontology, and time just means the time interval, which contains the start time and end time. Moreover, spatial relation is only reserved for the authoritative document, and is assisted in the location expression of the geo-event. The usage of triples in the semantic similarity matching to retrieve related geospatial resources will be presented in the next section.

2.3. Semantic Similarity Matching

After the extraction of the geo-event with its spatiotemporal attributes, related geospatial resources (both authoritative and social data) will be retrieved by using a semantic similarity matching algorithm based on the ontology to measure the degree of semantic similarity between them. Semantic similarity matching is generally composed of thematic similarity, spatial similarity, and temporal similarity

matching, each of which is calculated via the ontology and is given a certain weight in the result. Finally, the semantic similarity is the weighted sum of results.

2.3.1. Thematic Similarity

Thematic similarity is for matching the data topic words by querying hasDataTheme property of the geo-event (query condition) to that mentioned in the related description of service metadata (target) based on the ontology, and is mainly determined by the distance between two concepts in the ontology. Many research works have treated semantic similarity and discussed it within a perspective of ontology concepts and their relationships [46–49]. The ontology is usually represented by a directed graph. Each concept is represented as a node in the graph, and the hierarchical relationship between two adjacent concepts is represented as an edge connecting two adjacent nodes in the graph. The thematic similarity calculation can be considered the shortest path of edges between two nodes, i.e., semantic distance.

Considering the directional property of the graph, the positive and negative semantic distances of two adjacent concepts with a parent–child relationship should be different. For example, Figure 3 shows a fragment of a data theme in the ontology. The distance from “SST” to “OceanTemperature” should be shorter than the distance from “OceanTemperature” to “SST”. Consequently, when the retrieval condition is “OceanTemperature,” the node labeled “OceanTemperature,” with its child nodes and direct parent node, should both be used to match the condition. However, for the service matched with “OceanHydrology,” its similarity score should be less than that of the service matched with “SST,” because a user is more inclined to accept the retrieval of the SST-related service.

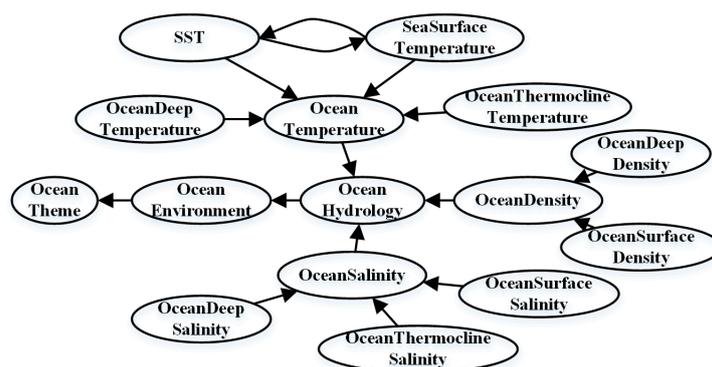


Figure 3. Fragment of data theme in the ontology.

To address the difference between two concepts in the ontology, weights are assigned according to their relationship. Then, a semantic distance weight table (Table 1), proposed by Ge [50], is utilized.

Table 1. Weight table.

	<i>p</i>	<i>g</i>	<i>s</i>	<i>b</i>
Φ	0	2	3	7
<i>p</i>	0	2	3	7
<i>g</i>	2	4	3.5	9
<i>s</i>	3	3.5	6	10
<i>b</i>	7	9	10	14

In the table, *g* means the edge from the child node to the parent node, *s* means the edge from the parent node to the child node, *p* means the edge connects two synonymous nodes, *b* means the binary relation to express the complicated semantic relationship between the two concepts, and Φ means no-operation and is only used with other single variables to determine the weight in a single operation. Moreover, the row calculation is first, followed by that of the column. A multi-operation is composed

of several single operations: the first two operations are regarded as continuous operations and the weight is chosen in the corresponding cell, while the rest are simply regarded as single operations. To address this calculation, the concepts in Figure 3 are used as an example: for the distance from “SST” to “OceanSalinity,” the operation is like $g > g > s$ (SST > OceanTemperature > OceanHydrology > OceanSalinity), and the weight is 7. However, the weights shown in the table are mainly empirical values according to the paths and directions along the edges between two nodes [50], so further consideration is required. The equation of semantic distance can be expressed as follows:

$$Dist(v_i, v_j) = \sum_{k=1}^n w_k \times \left(\frac{Nv_i + Nv_j}{Nv_i + Nv_j + 2 \times N_{LCA}} \right) \quad (1)$$

where $\sum_{k=1}^n w_k$ indicates the weighted sum of the shortest distance between nodes v_i and v_j , as is referred in the previous paragraph, and Nv_i and Nv_j , respectively, express the weighted distance from nodes v_i or v_j to their lowest joint ancestor node. N_{LCA} is the weighted distance from the lowest joint ancestor node to the root node.

Semantic distance is the foundation of the thematic similarity calculation, and is inversely proportional to the semantic similarity, and can be expressed as follows:

$$sim(v_i, v_j) = \frac{1}{Dist(v_i, v_j) + 1} \quad (2)$$

where $Dist(v_i, v_j)$ is obtained from Equation (1).

The final thematic similarity equation is

$$themeSim = \frac{\sum_{i=1}^N sim_i}{N} \quad (N > 0) \quad (3)$$

where N is the number of matched items, and sim_i is the thematic similarity of each item.

2.3.2. Spatial Similarity

The spatial property of the geo-event can be expressed as a spatial extent or a place name. A place name is generally associated with its spatial location. By relating the spatial property of the geo-event (query condition) with the spatial extent of the service (target), the service meeting the query condition will be retrieved [51]. To simplify calculations of spatial similarity between query conditions and the target, a bounding box is utilized. Consequently, a transformation may be needed to convert the place name to its spatial extent, i.e., a minimum bounding box. Therefore, the spatial similarity can be achieved by using spatial extents of query condition and target, and is expressed as follows:

$$spatialSim(A, B) = \frac{area(A \cap B)}{area(A)} \quad (4)$$

where A is the query condition, B is the target, and $A \cap B$ is the overlap area of A and B . The spatial similarity is obtained from the ratio of overlap area to the area of A .

2.3.3. Temporal Similarity

The temporal property of the service determines the timeliness of the geographical information service. For the same geo-event, the dynamic change can be observed according to the time sequence of the related services, and knowledge of the geo-event can be analyzed.

The temporal similarity calculation can be classified into time point and time period calculations. The former can be converted to a period calculation by setting the start time as the same as the end time. To unify time granularity, all time formats are converted to a standard time format, i.e., “yyyy-MM-dd HH:mm:ss,” where y indicates the year, M the month, d the day, H the hour, m the minute, and s

the second. Then, the time period calculation is determined by using the temporal property of the geo-event (query condition) and the service (target), and can be expressed as

$$timeSim(t, T) = \frac{duration(t \cap T)}{duration(t)} \quad (5)$$

where t is the query condition, T the target, and $t \cap T$ is the overlap time between the query condition and target. The temporal similarity is obtained from the ratio of temporal overlap to the query condition.

2.3.4. Final Semantic Similarity

The final semantic similarity calculation is the weighted sum of the thematic, spatial, and temporal similarities, and can be expressed as follows:

$$sum = w_1 \times themeSim + w_2 \times spatialSim + w_3 \times timeSim \quad (6)$$

where w_1 , w_2 , and w_3 , respectively, express weights of the thematic, spatial, and temporal similarities, and their sum equals 1. $themeSim$ is the result of the thematic similarity, $spatialSim$, that of the spatial similarity, and $timeSim$, that of the temporal similarity. In order to assign a proper weight to each part, an approach proposed by Andrade [52] is adopted. Each weight is calculated using the Pearson correlation coefficient based on the training set, which contains result samples of weights from several spatial queries. We set $w_1 = 0.4$, $w_2 = 0.3$, and $w_3 = 0.3$, which indicate that the result is focused on the thematic similarity. Therefore, results delivered to users are ranked according to the score each service obtains.

2.3.5. Case Study for Semantic Similarity Matching

In order to further elaborate the procedures in the semantic similarity matching, an example is illustrated. For a typhoon-related message accumulated in a time interval, its content is as follows:

“我不说话，今天那么大风，然后热搜我只看见广州停电，我大茂名无话可说！#彩虹台风# 2015-10-04 23:53”(Translation: I would say nothing for the strong wind. But when I searched on the web, I only saw power off in Guangzhou, what about Maoming? #Rainbow Typhoon# published time: 2015-10-04 23:53).

After processing the sentence, locations (“Guangzhou” and “Maoming”) and time (“2015-10-04 23:53”) were extracted. Meanwhile, other messages in the time interval were also processed. The spatial extent in this time interval could be expressed as a minimum bounding box containing all of the related locations, i.e., ($xmin = 104.50$, $ymin = 19.92$, $xmax = 116.70$, $ymin = 26.53$), and the time interval was (2015-10-04 23:50–2015-10-04 23:55).

For the typhoon event, its related thematic data concepts were achieved by retrieving the `hasDataTheme` property based on the ontology, e.g., Wind, Rainfall and Temperature. Meanwhile, a geospatial service on the rainfall was just used as a candidate service for semantic similarity matching. The metadata of the service was obtained by using the `GetCapabilities` function. In the metadata, the layer’s description was “2015年10月4日由台风造成的累积降雨量” (Translation: Accumulated rainfall on 4 October 2015 caused by typhoon), and the spatial extent of the layer was ($xmin = 102.38$, $ymin = 17.85$, $xmax = 122.00$, $ymin = 26.98$). The layer’s description was also processed using an NLP package, and the noun word “rainfall” extracted was used in the thematic similarity matching. The time words extracted were used as the temporal attributes of the geospatial service. However, if time words did not exist in the description, the published time of the service would be used, or the temporal similarity was just zero.

Obviously, the thematic similarity between the geospatial service and Rainfall concept related to the geo-event was 1.0. However, for the similarity between the geospatial service and Wind

concept, the semantic distance weight was calculated first according to Table 1. In Figure 4, the weight calculation could be regarded as a multi-operation, i.e., $g > g > s$ (Rainfall > Precipitation > OceanMeteorology > Wind); thus, the weight value was 7. Then, the distance between Rainfall and Wind concepts was 2.61 based on Equation (1), and the final thematic similarity was 0.38 based on Equations (2) and (3). For the spatial similarity, the spatial extents of the geo-event and geospatial service were used, and the spatial similarity was 1 based on Equation (4). For the temporal similarity, the temporal attributes of the geo-event and geospatial service were used, and the temporal similarity was 1 based on Equation (5). Finally, for the rainfall topic, the final similarity was 1 and, for the wind topic, the final similarity was 0.75 based on Equation (6). Thereafter, the results were ranked and organized according to data theme.

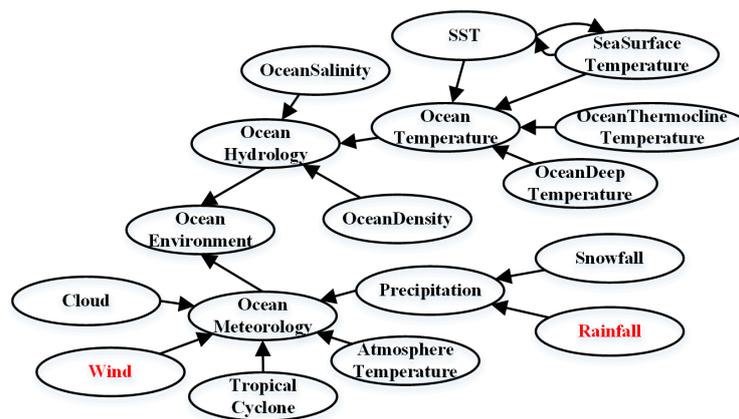


Figure 4. Thematic similarity calculation between Rainfall and Wind.

Ultimately, when a geo-event occurs, related services are automatically retrieved and organized according to the aforementioned processes, and the candidates are delivered to users.

3. Results

To verify the GEGIS approach, a geospatial information service prototype system was designed and implemented. Geo-events are obtained and processed from the web, and related geospatial resources are provided on the fly. We consider here a typhoon hazard case study of the South China Sea.

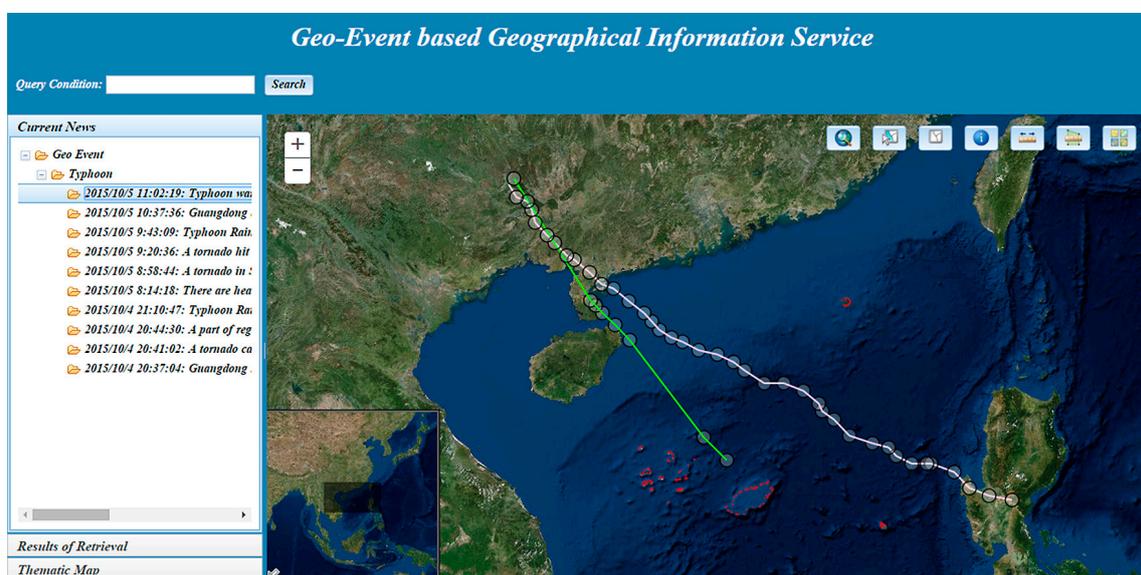
In order to collect messages from Sina Weibo in a timely manner, a web crawler, based on the Sina Weibo API [53], is implemented. As for a message crawled from Sina Weibo, its category should be identified first, as previously discussed. In order to reduce the interference of other messages to the typhoon-related messages in the situational awareness and assessment, a supervised classification algorithm is adopted. A corpus, containing 22 categories (e.g., travel, food, hazard (typhoon), and society), is constructed with the crawled messages to provide support for classification, and the total document number is 38,899. The document number of “typhoon” is 612 and is about 1.6% of the total number. The classified documents are divided into two sets: a training set and a test set, and the ratio of them is set to be 1:1. Based on the corpus, an SVM algorithm using LibSVM is implemented. For the purpose of the study, the test set labeled “typhoon” is illustrated in the experiment. Moreover, the SVM algorithm utilized demonstrates its ability to classify documents, so not all of the conditions are tested in the experiment: a linear kernel and a C-SVC method are set to be invariants, under the condition of which the cost factor reflecting the importance of outliers is tested (Table 2).

Table 2. Classification precision of typhoon.

Condition	Retrieved Number	Right Number	Precision
c = 0.1	207	182	60.5%
c = 1	349	247	82.0%
c = 2	367	247	82.0%
c = 5	362	246	81.7%
c = 10	370	248	82.4%

In Table 2, obvious changes of precision exist from $c = 0.1$ to $c = 10$. When $c = 0.1$, the precision is just 60.5%. However, precision increases and remains stable status afterward once $c = 1$. Under all conditions, the precision cannot reach 90%, which means that some noise and errors exist in each category. Some possible reasons include the heterogeneous distribution of samples in each category and word frequencies in each document. Taking everything into consideration, the condition of $c = 1$ is adopted to identify the geo-event.

The Rainbow typhoon first occurred in the Philippine Sea and then moved towards China. More messages occurred on Sina Weibo after 3 October 2015 and were crawled and processed in the system. The typhoon trajectory extracted from the message is shown as a green line in Figure 5, and the pink line is the actual trajectory. Round dots are typhoon locations at a given time. As the figure shows, the difference between the aforementioned two trajectories is smaller near land, but is otherwise larger. The reason for this is that, when the typhoon was near land, its location and spatial relationship were expressed in greater detail, and the simulation was more accurate. Another reason is the ambiguity caused by the vague description of the spatial location. For example, from a description of the actual typhoon location—“Typhoon is located in 340 km east by south of Wenchang City, Hainan Province”—locations and their spatial relationships can be extracted. However, the spatial relationship “east by south” indicates any location from east to southeast and cannot be accurately expressed.

**Figure 5.** Prototype system of the GEGIS.

The retrieval and organization results of the services are shown in Figure 6. Depending on the weighted sum of thematic, spatial, and temporal similarities, the final similarity was obtained and ordered according to the scores. When a message crawled and processed from the web contained a typhoon event, related services were retrieved and organized for immediate response. Figure 6 presents a rainstorm produced by a typhoon. The data service shown is accumulated rainfall on 4

October 2015 (the data are interpolated results from related observation sites at [54] and published on the web manually). Darker colors indicate greater rainfall amounts.

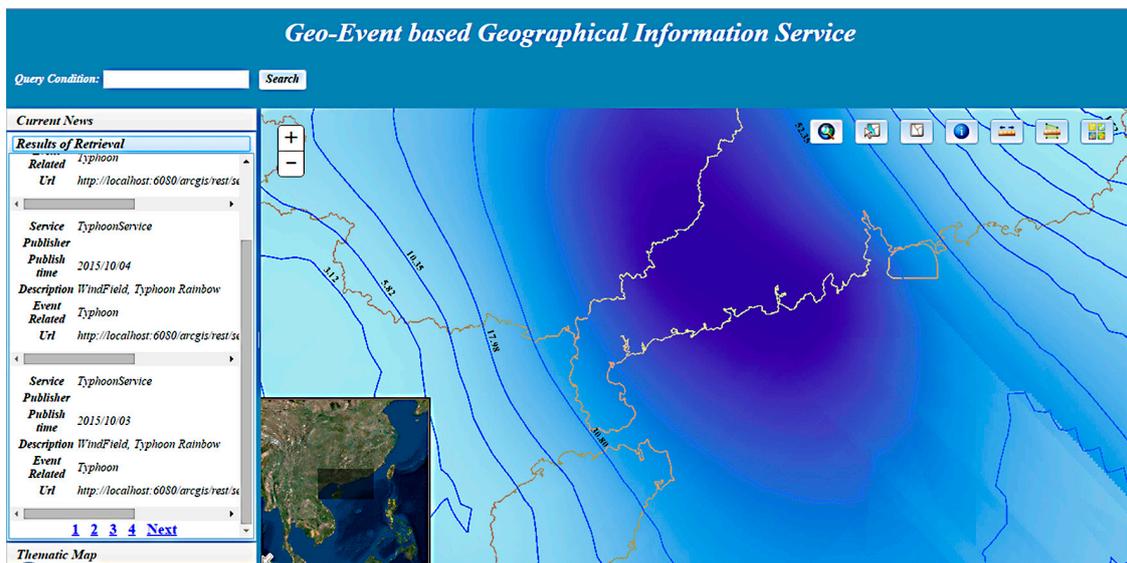


Figure 6. Retrieval and organization results.

Words about the 4 October typhoon impacts are extracted using a regex based on attribute extraction templates. Taking advantage of the kernel density function in calculating a magnitude-per-unit area from point features to fit a smoothly tapered surface to each point, a tool implementing the function from the ArcGIS toolbox was adopted to estimate the impact regions. Finally, the thematic map of hazard impacts was made using ArcGIS and published as a map service. The result is shown in Figure 7. The worst situation was in Zhanjiang City, Guangdong Province, because its location was in the typhoon path where more hazard-related data were released. Hazard impacts extracted from the messages included traffic gridlock, power outages, and passengers stranded. Surrounding regions, such as Beihai, Haikou, and Maoming, were also seriously impacted.

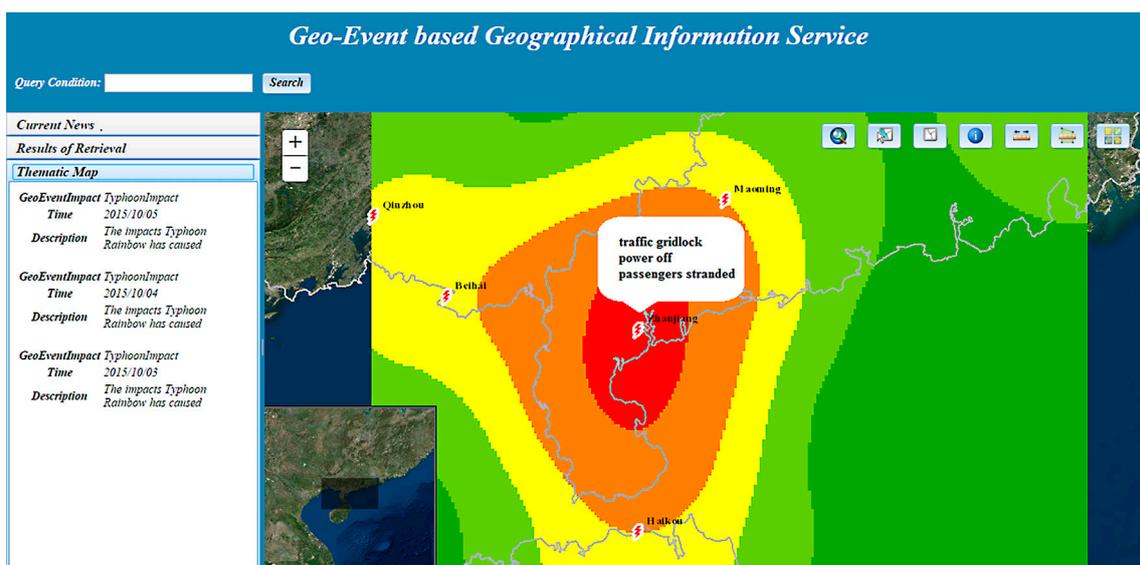


Figure 7. Statistics of hazard impacts.

4. Discussion

Considering the real-time impacts of a disaster, immediate information is needed to evaluate the situation and the damage so that corresponding relief measures can be undertaken. Most importantly, data quality is the primary consideration and is normally guaranteed by following rigorously defined procedures from authorities. There is no doubt that the authoritative data is slower than the social data. Nonetheless, for abundant social data provided by different users or departments without verification, extra work is still needed to enhance data quality, and manual verification is widely used on most occasions.

Trust and credibility are crucial for a disaster system to provide useful information. However, due to the lack of limitations and standards, the arbitrary expression of social data would increase the difficulties in extracting useful information. Generally, social data from government agencies are regarded as primary trustworthy resources that have since been assessed and verified by certain people, e.g., a typhoon forecast from a weather station, but they cannot satisfy users' urgent needs for information due to latency. Moreover, some credible data can be obtained from people who are creditable for having certain types of information, such as police officers and local media personalities [55], which can complement a timely disaster service.

Other data without direct demonstration to prove their trust and credibility, however, should require extra work to be used properly. Although many approaches proposed by Palen et al. [55] have emphasized the verification of the data, an automatic procedure to verify the data quality and exclude all the non-credible data is still lacking. On most occasions, an interactive verification between users and computers is necessary, which is the weakness of the proposed framework. The study adopts an approach combining both geographic and similarity approaches based on the spatial location to reduce the unreliable data: places extracted from the social data are mapped into the spatial extent based on the gazetteer, and spatial similarity is then calculated using the spatial extent to determine spatial correlations of messages to places where the disaster occurs. Some unreliable data outside the impact region will thereby be excluded. To further reduce the unreliable data, extra work is still needed.

In order to retrieve related geospatial resources other than the social data, some similarity-matching algorithms are adopted based on the established ontology. The study compares the thematic, spatial, and temporal similarities between query conditions and the related description of geospatial resources, and the final similarity is calculated through the weighted sum of the previous similarities. The weight value assigned is used to highlight, relatively, the thematic importance over spatial and temporal factors, but is subjective for the manual intervention. In order to assign a credible weight to the similarity function, some quantitative and qualitative evaluations should be considered to determine the proper weight value in the future. Another issue concerning the effectiveness of the similarity matching is the semantic distance calculation in the thematic similarity. The approach adopted considers the relation between two nodes, so directions along the shortest path between them are emphasized and regarded as a primary part to determine the weight. However, the weight value assigned needs further consideration to properly measure their differences. Meanwhile, other semantic distance calculation approaches should also be verified to adopt a better one.

Finally, traditional works generally organize geospatial resources from the perspective of the spatial location, which indicates that the geo-event-related data are implied in the data categories and organized based on location. All disaster-related data around the impact regions need to be retrieved when a disaster occurs. However, the GEGIS organizes geospatial resources from the perspective of the geo-event, which indicates that each geo-event will contain a series of related topical, classified, geospatial resources based on the ontology, and the spatial extent of the event will cover all of the impact regions. Therefore, the dynamic analysis of the event and effective response to the event change can be easily achieved. The GEGIS is especially useful for people who are interested in the disaster but are geographically remote, and it provides valuable data for making decisions. For example, more vulnerable regions may be discovered by analyzing the historical traces of a given geo-event.

The GEGIS is also helpful for the use of existing resources to implement analysis and mining for a geo-event.

5. Conclusions

In this study, the integrated technologies of NLP, NER, and the semantic web were introduced to support the GEGIS, which can respond to the geo-event and trigger timely related geospatial resource retrieval and organization from the perspective of the event based on its spatiotemporal attributes extracted from the web by means of the established ontology. As verification for the GEGIS framework, a geospatial information service prototype system was implemented, and a typhoon hazard case study was examined.

The framework proposed herein provides a new method for an intelligent geospatial information service, but there are still some areas that require strengthening. Future work mainly includes the following aspects:

First, accuracy of geo-event classification is a precondition for the selection of appropriate services. The SVM algorithm was used to identify the geo-event crawled and processed from the web. The accuracy mainly depends on impact factors, such as the setting of SVM parameters and the thematic distribution in the training corpus. Poor choice typically reduces the accuracy such that returned services would fail to match the user's purpose. Although SVM shows good performance in the experiment, the training corpus built requires much effort to be correctly classified and updated regularly [56]. Thus, its usefulness, extendibility, and flexibility are limited. Other machine learning algorithms should be tested to determine the best one.

Second, ambiguous and uncertain locations related to the geo-event should be further considered and expressed. Location descriptions are sometimes ambiguous and generally expressed as "location + spatial relation". In this case, an ontology containing place names and spatial relationships is generally used to reason the actual meaning of the location. The present approach can only recognize simple spatial relationships (e.g., around Beijing) and was used in the simulation of a typhoon path. For complex spatial relationships, extra work is necessary to enhance their discovery. However, the representation of an ambiguous spatial extent simply uses the minimum bounding box containing all place names mentioned in the geo-event, without any consideration of their spatial relationships. This coarse representation of the spatial extent has poor accuracy and will be improved in the future.

Third, extra work should be made to improve the data quality; thus, users can directly utilize the data for their purposes. Social media can provide timely information for users when a disaster occurs, compared with the information obtained from authorities, which are generally slow, but with high trust and credibility. The approach used in the study has excluded some unreliable and redundant data, but is still ineffective in processing data in some occasions, where manual verification is still needed. Therefore, an interactive user interface should be designed to verify the data. Moreover, a user scoring mechanism can be adopted to assess the data as well as the data providers to evaluate the provider. Thus, the published data can be regarded as a credible source.

Acknowledgments: The National Key Research and Development Program of China (2016YFC1402900).

Author Contributions: All the authors contributed to the development of GEGIS and this manuscript. Fenzhen Su proposed the research and conceived the system framework. Yu Zhang and Qi Wang designed and implemented GEGIS and wrote the manuscript. Wenzhou Wu edited the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodchild, M.F.; Glennon, J.A. Crowdsourcing geographic information for disaster response: A research frontier. *Int. J. Dig. Earth* **2010**, *3*, 231–241. [[CrossRef](#)]
2. Granell, C.; Ostermann, F.O. Beyond data collection: Objectives and methods of research using vgi and geo-social media for disaster management. *Comput. Environ. Urban Syst.* **2016**, *59*, 231–243. [[CrossRef](#)]

3. Meier, P. Micromappers: Microtasking for Disaster Response. Available online: <https://irevolutions.org/2013/09/18/micromappers/> (accessed on 27 March 2017).
4. Wan, Z.; Hong, Y.; Khan, S.; Gourley, J.; Flamig, Z.; Kirschbaum, D.; Tang, G. A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration. *Environ. Model. Softw.* **2014**, *58*, 86–94. [[CrossRef](#)]
5. Turpin, E.; Holderness, T. Petajakarta.org Major Open Data Collection-Hydrological Infrastructure Network for Jakarta, Indonesia. Available online: <https://researchdata.ands.org.au/petajakartaorg-major-open-jakarta-indonesia/561003> (accessed on 27 March 2017).
6. Jongman, B.; Wagemaker, J.; Romero, B.R.; de Perez, E.C. Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and twitter signals. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2246–2266. [[CrossRef](#)]
7. Chae, J.; Thom, D.; Jang, Y.; Kim, S.; Ertl, T.; Ebert, D.S. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Comput. Graph.* **2014**, *38*, 51–60. [[CrossRef](#)]
8. Ghahremanlou, L.; Sherchan, W.; Thom, J.A. Geotagging twitter messages in crisis management. *Comput. J.* **2015**, *58*, 1937–1954. [[CrossRef](#)]
9. Deng, Q.; Liu, Y.; Zhang, H.; Deng, X.; Ma, Y. A new crowdsourcing model to assess disaster using microblog data in typhoon haiyan. *Nat. Hazards* **2016**, *84*, 1241–1256. [[CrossRef](#)]
10. Wang, Y.; Wang, T.; Ye, X.; Zhu, J.; Lee, J. Using social media for emergency response and urban sustainability: A case study of the 2012 beijing rainstorm. *Sustainability* **2016**, *8*, 25. [[CrossRef](#)]
11. Schnebele, E.; Cervone, G.; Waters, N. Road assessment after flood events using non-authoritative data. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 1007–1015. [[CrossRef](#)]
12. Middleton, S.E.; Middleton, L.; Modafferi, S. Real-time crisis mapping of natural disasters using social media. *IEEE Intell. Syst.* **2014**, *29*, 9–17. [[CrossRef](#)]
13. Atefeh, F.; Khreich, W. A survey of techniques for event detection in twitter. *Comput. Intell.* **2015**, *31*, 132–164. [[CrossRef](#)]
14. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In Proceedings of the International Conference on World Wide Web, WWW 2010, Raleigh, NC, USA, 21–25 April 2010; pp. 851–860.
15. Becker, H.; Naaman, M.; Gravano, L. Beyond Trending Topics: Real-World Event Identification on Twitter. Available online: <https://mice.cs.columbia.edu/getTechreport.php?techreportID=1462&format=pdf&> (accessed on 27 March 2017).
16. McMinn, A.J.; Moshfeghi, Y.; Jose, J.M. Building a large-scale corpus for evaluating event detection on twitter. In Proceedings of the ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 409–418.
17. Cheng, T.; Wicks, T. Event detection using twitter: A spatio-temporal approach. *PLoS ONE* **2014**, *9*, e97807. [[CrossRef](#)] [[PubMed](#)]
18. Zhao, W.X.; Jiang, J.; Weng, J.; He, J.; Lim, E.P.; Yan, H.; Li, X. Comparing twitter and traditional media using topic models. 2011. Available online: http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=2374&context=sis_research (accessed on 27 March 2017).
19. Hromic, H.; Prangnawarat, N.; Hulpuş, I.; Karnstedt, M.; Hayes, C. Graph-Based Methods for Clustering Topics of Interest in Twitter. Available online: http://link.springer.com/chapter/10.1007/978-3-319-19890-3_61 (accessed on 27 March 2017).
20. Haworth, B. Emergency management perspectives on volunteered geographic information: Opportunities, challenges and change. *Comput. Environ. Urban Syst.* **2016**, *57*, 189–198. [[CrossRef](#)]
21. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Ann. Am. Assoc. Geogr.* **2012**, *102*, 571–590. [[CrossRef](#)]
22. Crooks, A.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. Earthquake: Twitter as a distributed sensor system. *Trans. GIS* **2013**, *17*, 124–147. [[CrossRef](#)]
23. Barrington, L.; Ghosh, S.; Greene, M.; Har-Noy, S.; Berger, J.; Gill, S.; Lin, A.Y.-M.; Huyck, C. Crowdsourcing earthquake damage assessment using remote sensing imagery. *Ann. Geophys.* **2012**, *54*. [[CrossRef](#)]
24. De Albuquerque, J.P.; Herfort, B.; Brenning, A.; Zipf, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 667–689. [[CrossRef](#)]

25. Bakillah, M.; Li, R.-Y.; Liang, S.H. Geo-located community detection in twitter with enhanced fast-greedy optimization of modularity: The case study of typhoon haiyan. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 258–279. [[CrossRef](#)]
26. Roche, S.; Propeck-Zimmermann, E.; Mericskay, B. Geoweb and crisis management: Issues and perspectives of volunteered geographic information. *Geojournal* **2013**, *78*, 21–40. [[CrossRef](#)]
27. Rogstadius, J.; Vukovic, M.; Teixeira, C.; Kostakos, V.; Karapanos, E.; Laredo, J.A. Crisistracker: Crowdsourced social media curation for disaster awareness. *IBM J. Res. Dev.* **2013**, *57*, 4:1–4:13. [[CrossRef](#)]
28. Musaev, A.; Wang, D.; Cho, C.A.; Pu, C. Landslide detection service based on composition of physical and social information services. In Proceedings of the 2014 IEEE International Conference on Web Services (ICWS), Anchorage, AK, USA, 27 June–2 July 2014.
29. McCreadie, R.; Macdonald, C.; Ounis, I. Eaims: Emergency Analysis Identification and Management system. Available online: <http://eprints.gla.ac.uk/119286/> (accessed on 27 March 2017).
30. Koswatte, S.; McDougall, K.; Liu, X. Sdi and crowdsourced spatial information management automation for disaster management. *Surv. Rev.* **2015**, *47*, 307–315. [[CrossRef](#)]
31. Lwin, K.K.; Zettsu, K.; Sugiura, K. Geovisualization and correlation analysis between geotagged twitter and jma rainfall data: Case of heavy rain disaster in hiroshima. In Proceedings of the 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICS DM), Fuzhou, China, 8–10 July 2015.
32. Horita, F.E.; de Albuquerque, J.P.; Degrossi, L.C.; Mendiondo, E.M.; Ueyama, J. Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. *Comput. Geosci.* **2015**, *80*, 84–94. [[CrossRef](#)]
33. De Longueville, B.; Annoni, A.; Schade, S.; Ostlaender, N.; Whitmore, C. Digital earth’s nervous system for crisis events: Real-time sensor web enablement of volunteered geographic information. *Int. J. Dig. Earth* **2010**, *3*, 242–259. [[CrossRef](#)]
34. Yue, P.; Zhang, C.; Zhang, M.; Zhai, X.; Jiang, L. An sdi approach for big data analytics: The case on sensor web event detection and geoprocessing workflow. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4720–4728. [[CrossRef](#)]
35. Niko, D.L.; Hwang, H.; Lee, Y.; Kim, C. *Integrating User-Generated Content and Spatial Data into Web Gis for Disaster History*; Springer: Berlin, Germany, 2011; pp. 245–255.
36. Dorn, H.; Vetter, M.; Höfle, B. Gis-based roughness derivation for flood simulations: A comparison of orthophotos, lidar and crowdsourced geodata. *Remote Sens.* **2014**, *6*, 1739–1759. [[CrossRef](#)]
37. Sina Weibo. Available online: <http://weibo.com/> (accessed on 29 March 2017).
38. SWEET. Available online: <https://sweet.jpl.nasa.gov/> (accessed on 29 March 2017).
39. Zhang, B.; Ma, Y.; Li, Y. Standardization of natural disaster classification in china. *J. Nat. Dis.* **2013**, *5*, 2.
40. LibSVM. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed on 29 March 2017).
41. NLPiR. Available online: <http://ictclas.nlpir.org/> (accessed on 29 March 2017).
42. Verma, S.; Vieweg, S.; Corvey, W.; Palen, L.; Martin, J.H.; Palmer, M.; Schram, A.; Anderson, K.M. Natural language processing to the rescue? Extracting “situational awareness” tweets during mass emergency. In Proceedings of the International Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
43. Popoola, A.; Krasnoshtan, D.; Toth, A.P.; Naroditskiy, V.; Castillo, C.; Meier, P.; Rahwan, I. Information verification during natural disasters. In Proceedings of the International Conference on World Wide Web, San Francisco, CA, USA, 27 October–1 November 2013; pp. 1029–1032.
44. Kent, C.K.; Salim, N. Features based text similarity detection. *arXiv* **2010**, arXiv:1001.3487.
45. Lucene. Available online: <https://lucene.apache.org> (accessed on 29 March 2017).
46. Al-Bakri, M.; Fairbairn, D. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *Int. J. Geogr. Inform. Sci.* **2012**, *26*, 1437–1456. [[CrossRef](#)]
47. Andrea Rodriguez, M.; Egenhofer, M.J. Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure. *Int. J. Geogr. Inform. Sci.* **2004**, *18*, 229–256. [[CrossRef](#)]
48. Chen, N.; He, J.; Yang, C.; Wang, C. A node semantic similarity schema-matching method for multi-version web coverage service retrieval. *Int. J. Geogr. Inform. Sci.* **2012**, *26*, 1051–1072. [[CrossRef](#)]
49. Li, W.; Raskin, R.; Goodchild, M.F. Semantic similarity measurement based on knowledge mining: An artificial neural net approach. *Int. J. Geogr. Inform. Sci.* **2012**, *26*, 1415–1435. [[CrossRef](#)]

50. Ge, W. Research on Geographical Information Service Discovery Methods. Ph.D. Thesis, PLA Information Engineering University, Zhenzhou, China, 2012.
51. Vasardani, M.; Winter, S.; Richter, K.-F. Locating place names from place descriptions. *Int. J. Geogr. Inform. Sci.* **2013**, *27*, 2509–2532. [[CrossRef](#)]
52. Andrade, F.G.D.; Baptista, C.D.S. Improving geographic information retrieval in spatial data infrastructures. *GeoInformatica* **2014**, *18*, 793–818. [[CrossRef](#)]
53. Sina Weibo API. Available online: <http://open.weibo.com/wiki/%E5%BE%AE%E5%8D%9AAPI> (accessed on 29 March 2017).
54. Rainfall Data from NOAA. Available online: <http://gis.ncdc.noaa.gov/> (accessed on 29 March 2017).
55. Palen, L.; Vieweg, S.; Anderson, K.M. Supporting “everyday analysts” in safety-and time-critical situations. *Inform. Soc.* **2010**, *27*, 52–62. [[CrossRef](#)]
56. Hua, T.; Chen, F.; Zhao, L.; Lu, C.-T.; Ramakrishnan, N. Automatic targeted-domain spatiotemporal event detection in twitter. *GeoInformatica* **2016**, *20*, 765–795. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).