Research on Crowdsourcing Emergency Information Extraction of Based on Events' Frame

Bo Yang\textsuperscript{1}, Jizhou Wang\textsuperscript{2}, Weijun Ma\textsuperscript{2}, Xi Mao\textsuperscript{3}

\textsuperscript{1}School of Mapping and Geographical Science, Liaoning Technical University, Fuxin, Liaoning, China, 123000.
\textsuperscript{2}Chinese Academy of Surveying & Mapping, Beijing, China, 100830.

Corresponding author. Tel.: +86 15810839035; email: michaelyangbo@outlook.com

Abstract. At present, the common information extraction method cannot extract the structured emergency event information accurately; the general information retrieval tool cannot completely identify the emergency geographic information; these ways also do not have an accurate assessment of these results of distilling. So, this paper proposes an emergency information collection technology based on event framework. This technique is to solve the problem of emergency information picking. It mainly includes emergency information extraction model (EIEM), complete address recognition method (CARM) and the accuracy evaluation model of emergency information (AEMEI). EIEM can be structured to extract emergency information and complements the lack of network data acquisition in emergency mapping. CARM uses a hierarchical model and the shortest path algorithm and allows the toponymy pieces to be joined as a full address. AEMEI analyzes the results of the emergency event and summarizes the advantages and disadvantages of the event framework. Experiments show that event frame technology can solve the problem of emergency information drawing and provides reference cases for other applications. When the emergency disaster is about to occur, the relevant departments query emergency's data that has occurred in the past. They can make arrangements ahead of schedule which defense and reducing disaster. The technology decreases the number of casualties and property damage in the country and world. This is of great significance to the state and society.

1. Introduction
Nowadays, there are various natural and social emergencies in the world, which has caused serious losses to the people and the country. When an emergency occurs, the aid department need the mapping department to provide quickly emergency mapping data \cite{1}. With the development of the society, the traditional mapping which focuses on the acquisition of basic geographic information cannot meet the needs of emergency mapping timely, efficient and low cost. However, we found that the multisource data contain a large number of thematic geographic information and can solve the problem which emergency mapping can promptly access information. If you can extract information from the massive source of data to extract the feature geographic information about emergency mapping, this will extremely enrich the content of emergency geographic information and improve an emergency support capability.

Currently, these main methods of information extraction include semi-structured approach, plain text way, and common tools which provide a stable technical and theoretical basis for general
information abstract [2]. Yet, these methods are only appropriate for it which includes a generic scene, a specific language, and the specified text type. Meanwhile, these measures cannot extract emergency geographic information, recognize the full address, and analyze the extractability results. Emergency geographic information has become the research hot spot of geographic information science, which different experts have different research aspects. Reference [3] presents a model and application of information extractability based on the ontology of emergency cases. First, the model uses the ontology to describe the case in the field of emergency knowledge. Then, the model is applied to draw cases under the guidance of domain knowledge and build case libraries. Finally, the case library is provided to emergency rescue decision makers. Reference [4] proposes a method of information extraction based on rule matching for emergency events which match the text according to the rules that match the incident and is used to normalize the extraction results. Experiments show that the method is very effective. Reference [5] raises a technique based on the extraction and rapid release of critical information from earthquakes, analyzes the needs of government at all levels and emergency teams after a devastating earthquake, and studies the extraction of emergency core information and the rapid release of emergency information during the fast response period. Reference [6] has developed a geographic information extraction system based on spatial relationship query. The system uses a geographic information extraction model and can efficiently handle the query of emergency keywords. The above research focuses on fragments of emergency geographic information which include extraction, matching, and querying [7]. Some experts also study the query and index of a class of emergency disaster events [8]-[11].

However, we note that the emergency information in the network is mainly in the form of events as a unit. Events’ frame technology is a formal expression of the integrity of the event. Thus, this paper puts forward a technique based on the events’ frame of the multisource of emergency geographic information extraction. The events’ frame technology is used to construct the emergency event extraction model, the complete method of toponymic recognition and spatialization. Finally, the accuracy evaluation model was used to evaluate the results.

2. System Framework
First, the emergency information extraction model is used to extract the original emergency event information. And, the extracted results are subjected to structured processing. Second, the address recognition method is used to identify the geographical segment. And, the complete address gets the spatial information [12]. Finally, the results are evaluated using the accuracy assessment. The user can query by entering a query statement [13]. Based on the events’ frame of the multisource of emergency geographic information extraction shown in Figure 1. There are three levels from bottom to top, including the data layer, logical processing layer, and front display layer.
Emergency information extraction modeling and structured expression

The query results are returned
Extract the result query
Display the results query

Figure 1. Crowdsourcing information extraction based on events' frame.

3. Key Technology Implementation

3.1. Emergencies Information Extraction Model

The emergency event extraction model is the core module of the events' frame. Its main function is to analyze the extraction of various types of emergency events. Then, it extracts comprehensive information about the incident. Finally, it will extract the results of the structured expression and storage.

The specific process of the extraction method based on emergency events is as follows (refer to Figure 2).
3.1.1. **Preprocess** the contents of the emergency information extraction

The structured content is preserved which include title, type, level, longitude, latitude, time, source, detail, storage, etc.

3.1.2. List the attribute description of the extracted information

According to the emergency information extraction results, the establishment of emergency events and attributes of the relationship between the documents.

3.1.3. **Establish** a set of emergency events

According to the emergency response to the comprehensive extraction of structured organization and management, the information of the emergency event attribute which the discretization of the extracted result builds a structured set of texts indexes. Through the above three steps to form a contingency geographic information structure of the system.

3.1.4. **Establish** the link between the extracted attributes

The relationship between the extracted attributes is established according to the relationship in the index sets.
We conduct a case study of a series of emergency incident extraction processes [14] which include natural disaster, accident disaster, public health, safe society, etc. The following figure shows the extraction model for emergency events (refer to Figure 3).

![Figure 3 Emergency information extraction model.](image)

The model consists of four elements:

3.1.5. Temporal and Spatial Characteristics of Emergencies
The description of the basic information about the incident, which includes the basic information and professional attributes of emergencies, such as earthquake magnitude and intensity.

3.1.6. Emergency Department
These are the subjects of the various tasks at all levels of emergency response process. This is mainly at all levels of government agencies.

3.1.7. Emergency Task
The sum of the subtasks taken during the emergency response process. Each task is a time and space
within a range to achieve a certain purpose of a series of actions integrated.

3.1.8. Auxiliary Decision Model
These models provide supplementary information for task execution which includes disaster prediction model, disaster assessment, emergency evacuation and resource allocation.

3.2. Full-toponomy Recognition Technology Based on a Hierarchical Model
Toponomy recognition based on hierarchical model is the key technology of emergency information extraction model. The hierarchical model is constructed according to China address standard. China's address is decreasing and the scope is from large to small, including the state, province, city, county, district, etc. According to the results of the emergency event to those containing the toponom fragments of the data, the main function of the technology is to identify the complete address based on the emergency information. The address recognition process based on the hierarchical model is shown in the following figure (refer to Figure 4). This section will cover the main technical aspects of address recognition, including the hierarchical model and the N-shortest path algorithm.

![Fig.4 The technical process of place name recognition.](image)

On the basis of referring to the rules of toponomy in China, this paper puts forward the hierarchical toponomy model. The hierarchical model of the identification process is divided into three cases. The hierarchical model only identifies names in national standards. When the toponomy is outside the national standard, the operation of this article is the log output of the name. When the hierarchical model recognition event is partially recognized and another part is not recognized, this article stores the identified content and will be unrecognized content log output. Finally, this article analyzes all the logs. If the toponomy in the log is the missing content of the database, then this article will add the missing content to the corresponding part, otherwise, do not do statistics (refer to Figure 5).

The hierarchy model is defined as follows:
1) \(<\text{standard toponomy}>::=\langle\text{region}\rangle\langle\text{address}\rangle\)
2) <Region>::=[country][province][city][county]
3) <Address>::=road name|community name|place name|building name|floor number|room number

Fig. 5 Geographical address hierarchy model.

The hierarchical model splits the names into strings with four elements and does not strictly define the semantics of each layer. In practice, without affecting the efficiency of extraction and identification, the first four are regional geographical names. This is the national, provincial, city, county, district. And the same level of provincial and municipalities, autonomous regions, and special administrative regions. A large number of cases studies found that the four-tier model to meet the demand. In general, there is a spatial inclusion relationship between layers and layers and a space distance on the adjoining relationship. The hierarchical model needs to take full account of the spatial relationship between the toponyms [15]. Spatial relationship refers to the location relationship between the names of elements, including, adjacent, adjoining, azimuth and distance. The model takes the strategy of spatial relation matching. The experimental results show that the spatial constraint data can improve the accuracy of toponymic recognition. The abstraction of the hierarchical model ensures its versatility and strong ability to express the names.

Separate hierarchical models can only connect simple names and cannot achieve full address acquisition. This article uses the N-shortest path algorithm to complete the toponomic fragment real connection [16], [17]. For details, refer to the Dijkstra algorithm [18].

3.3. Evaluation Model of Emergency Information Extraction
On the basis of reference to the evaluation criteria of general information extraction, this paper presents an emergency information extraction and evaluation model. The model is another module of the events' frame. The function of the model is to evaluate the results of emergency information extraction by qualitative and quantitative methods. This paper constructs three aspects of the indicators, namely comprehensive accuracy of emergency geographic information (CAEGI) [19], toponym recognition accuracy (TRA) [20], [21] and toponomy spatial accuracy (TSA) [22], [23]. The detailed quantitative assessment process is shown below (refer to Figure 6).
Fig. 6 Evaluation model of emergency information extraction.

(1) Precision ($P$): The ratio of emergency information extraction accuracy to test sample [24].

$$P = \frac{\text{Extract results (E)}}{\text{Samples (S)}} \times 100\%$$ (1)

(2) Recall ($R$): The correct extraction result and the extraction accuracy ratio [25].

$$R = \frac{\text{Correct number (C)}}{E} \times 100\%$$ (2)

(3) F-factor ($F$): Precision and recall rate of the weighted average. $F$ is the overall average of the system's extraction performance [26].

The formula for calculating the harmonic mean is:

$$F_{\beta} = \frac{(\beta^2 + 1)(P + R)}{\beta^2 \times P + R}$$ (3)

The $\beta$ is the weighting coefficient. The three phases of emergency information extraction are critical. This article treats the importance of precision and recall rates equally. So, the weight coefficient of $F$ is set to 1 in this paper. $P$ is the precision rate; $R$ is the recall rate.

(4) Mean synthetic average ($M$): $M$ is the expectation of $F$ [27].

$$M(F) = \sum_{i=1}^{n} f_{i} p_{i}$$ (4)

$f$ is the F-factor calculated by (3). $p$ is the weight of the emergency information extraction result. In the order of extraction, the weights are 1/3, $i=1,2,3$.

As shown in Figure 6, we use (1) to get the $P$ values for CAEGI, TRA, and TSA. We employ (2) to acquire the $R$ values for CAEGI, TRA, and TSA. We apply (3) to obtain the $F$ values for CAEGI, TRA, and TSA. Finally, we exploit (4) to achieve $M$ results for CAEGI, TRA, and TSA.

4. Results and Discussion

Driven by the emergency events' frame, the emergency information extraction results are shown in the figure (refer to Figure 7). In this paper, 114284 nonempty original web pages are used as test samples.
Test samples from China National Emergency Broadcasting Network. As shown in the following table (see Table 1), the data sources of the site are highly reliable, authentic and real-time update of emergency data.

![Figure 7. Emergency information extraction results.](image)

| Type   | S   | E     | C     | P     | R     | F     |
|--------|-----|-------|-------|-------|-------|-------|
| CAE    | 114284 | 113662 | 113255 | 99.10% | 99.64% | 99.37% |
| TRA    | 113662 | 92846  | 92808  | 81.65% | 99.96% | 89.88% |
| TSA    | 92846  | 90589  | 90566  | 97.54% | 99.97% | 98.74% |
| M      | -    | -     | -     | -     | 92.76% | 99.86% | 95.99% |

Notes: Please refer to heading 3.3 on the abbreviations for the table.

The results of the analysis show that the main error of CAEGI is derived from the noise of the test data. There are website science articles and pictures on the site, which pages do not contain incident information. TRA's error is mainly that some pages do not contain toponomic information. At the same time, this test sample is a Chinese page. Chinese segmentation dictionary does not completely cover foreign names information [28]. This is the biggest source of error for TRA. TSA is the main error of TRA caused. TSA is using a domestic map service. When the query to foreign TSA, the error appeared. At the same time, some of the contents of TSA cannot be found [29]. The results of the three types of measurements show that the comprehensive extraction accuracy based on the emergency information events' frame is 95.99%. The emergency geographic information extraction system is reliable and conforms to the actual engineering needs.

5. Conclusion and Future work
There are many information extraction manners exist today, and it is an unattractive operation for us to select the appropriate way. The main purpose of this study is to propose a suitable procedure for
emergency geographic information extraction. Event frame technology which including data, models and show layer is for this target. The datasets come from crowdsourcing. The function of the EIEM in the framework is the structural extraction of the emergency information, perfectly matching the address and accurate access to the spatial information of location. The effect of the AEMEI in the frame is to evaluate the extraction results of the EIEM. The final result which meeting the AEMEI is displayed. This framework provides a one-stop operation for the extraction of emergency information. There are still shortcomings in this paper. First, in the EIEM, different data sources need to compile distinct rules to extract the emergency information. This makes the regulations too time consuming. In the future, we want to use machine learning to automatically draw information. Second, in the TSA, we get the convenience of querying spatial coordinate. For those which cannot query the spatial location information, this article does not propose a fitting method. Third, this article returns the results based on the user's query. We should use the recommended system to intelligently decision the results. For the future work, this is still the direction of further research.

6. References

[1] Yang, J. (2017). Environmental Emergency Response Plan. Environmental Management in Mega Construction Projects. Springer Singapore.

[2] Etzioni, O., Fader, A., Christensen, J., Soderland, S., & Mausam, M. (2011). Open Information Extraction: The Second Generation. IJCAI 2011, Proceedings of the, International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July (Vol.14, pp.3-10). DBLP.

[3] Yang, P., Wang, W., & Dong, C. (2009). Application of Emergency Case Ontology Model in Earthquake. International Conference on Management and Service Science (pp.1-5). IEEE.

[4] Jiang, D. L. (2010). Research on extraction of emergency event information based on rules matching. Computer Engineering & Design, 31(14), 3294-3297.

[5] Min, L. I., Zhao-Long, L. I., Huo, J., & Yong-Qiang, L. I. (2015). Research on technology of extraction and rapid release of key information about earthquake emergency. Journal of Seismological Research.

[6] Kalashnikov, D. V., Ma, Y., Mehrotra, S., Hariharan, R., & Butts, C. (2006). Modeling and querying uncertain spatial information for situational awareness applications. ACM International Symposium on Geographic Information Systems, Acm-Gis 2006, November 10-11, 2006, Arlington, Virginia, Usa, Proceedings (pp.131-138). DBLP.

[7] Bergen, J. R., Anandan, P., Hanna, K. J., & Hingorani, R. (1992). Hierarchical Model-Based Motion Estimation. European Conference on Computer Vision (Vol.11, pp.237-252). Springer-Verlag.

[8] Grishman, R. (2015). Information extraction. Foundations & Trends in Databases, 30(5), 8-15.

[9] Cowie, J., & Lehnert, W. (1996). Information extraction. Information extraction. : Springer.

[10] Cowie, J., & Lehnert, W. (1996). Information extraction. Foundations & Trends in Databases, 1(3), 261-377.

[11] Wang, J., Yan, W., & Ngueyn, M. (2017). A Framework of Event-Driven Traffic Ticketing System. IGI Global.

[12] Kurra, S. (2014). Solar power plant design and location selection for tamuk using gis and nlp. Dissertations & Theses - Gradworks.

[13] Downey, D., Etzioni, O., Soderland, S., & Weld, D. S. (2004). Learning text patterns for web information extraction and assessment. National Conference on Artificial Intelligence (pp.50–55).

[14] Seymore, K., Mccallum, A., & Rosenfeld, R. (1999). Learning hidden markov model structure for information extraction. Proc Aaai 99 Workshop Machine Learning for Information Extraction, 37–42.

[15] Bian, F., & Wan, Y. (2009). A novel spatial co-location pattern mining algorithm based on k-nearest feature relationship. Geomatics & Information Science of Wuhan University, 34(3), 331-334.

[16] Johnson, D. B. (1973). A note on dijkstra's shortest path algorithm. Journal of the Acm, 20(3), 385-388.

[17] Zheng, F., Simpson, A. R., & Zecchin, A. C. (2011). A combined nlp-differential evolution algorithm approach for the optimization of looped water distribution systems. Water Resources Research, 47(8), 2924-2930.

[18] Mehlhorn, Schäfer, & Tamaki. (2003). A heuristic for dijkstra's algorithm with many targets and its use in weighted matching algorithms. Algorithmica, 36(1), 75-88.

[19] Yandell, M. D., & Majoros, W. H. (2002). Genomics and natural language processing. Nature Reviews Genetics, 3(8), 601.
[20] Jagannathan, V., Mullett, C. J., Arbogast, J. G., Halbritter, K. A., Yellapragada, D., & Regulapati, S., et al. (2009). Assessment of commercial nlp engines for medication information extraction from dictated clinical notes. International Journal of Medical Informatics, 78(4), 284-291.

[21] Cogley, J., Stokes, N., & Carthy, J. (2013). Exploring the effectiveness of medical entity recognition for clinical information retrieval. International Workshop on Data and Text Mining in Biomedical Informatics (Vol.5, pp.3-4). ACM.

[22] Okoye, I., Mahmud, J., Lau, T., & Cerruti, J. (2011). Find this for me: mobile information retrieval on the open web. International Conference on Intelligent User Interfaces (pp.3-12). ACM.

[23] Willis, A. (2015). Using NLP to Support Scalable Assessment of Short Free Text Responses. Tenth Workshop on Innovative Use of Nlp for Building Educational Applications (pp.243-253).

[24] Dimitriadis, S., & Goumopoulos, C. (2008). Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications. Panhellenic Conference on Informatics (pp.100-104). IEEE Computer Society.

[25] Jijkoun, V., Rijke, M. D., & Mur, J. (2004). Information extraction for question answering: improving recall through syntactic patterns. International Conference on Computational Linguistics (pp.1284). Association for Computational Linguistics.

[26] Makhouli, J., Kubala, F., Schwartz, R., & Weischedel, R. (1999). Performance measures for information extraction. Proceedings of Darpa Broadcast News Workshop, 249–252.

[27] Pfeiffer, P. E. (1990). Mathematical Expectation. Probability for Applications. Springer New York.

[28] Florczyk, A. J., Nogueras-Iso, J., Zarazaga-Soria, F. J., & Béjar, R. (2012). Identifying orthoimages in web map services. Computers & Geosciences, 47(4), 130-142.

[29] Abdulkarim, B., Kamberov, R., & Hay, G. (2014). Supporting urban energy efficiency with volunteered roof information and the google maps api. Remote Sensing, 6(10), 9691-9711.

Acknowledgment

This work was supported by the Incident Information Fusion and Visualization Based on Public Source Data (Grant No. 7771615).

Bo Yang was born in Anhui, China, in 1992. Currently, he is pursuing his M.S. degree at the School of Mapping and Geographical Science, Liaoning Technical University, Liaoning, China, from September 2015 to June 2018. His research interests include data mining, geographical information science.

Jizhou Wang was born in Hebei, China, in 1976. He obtained his Ph.D. in cartography and GIS at Wuhan University in 2005. He leads the local GIS research group which focuses on research in the field of geographical information systems.

Weijun Ma was born in Hebei, China, in 1976. He obtained his M.S. in cartography and GIS at Chinese Academy of Surveying & Mapping in 2004. His current research interests include population GIS, especially spatialization of population, constructing population information grid and census, and data mining, emergency surveying and mapping.

Xi Mao was born in Zhejiang, China, in 1983. He obtained his B.S. Ph.D. in cartography and GIS at Wuhan University, Peking University in 2005, 2010, respectively. His current research interests include urban geospatial modeling, geospatial information search engine, geospatial information services.