AGILE: GIScience Series, 5, 37, 2024. https://doi.org/10.5194/agile-giss-5-37-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

Constructing Spatio-temporal Disaster Knowledge Graph from Social Media

Yen-Ching Liu[1] and Chiao-Ling Kuo[1,2]

[1] Department of Geography, National Taiwan University, Taipei, Taiwan
[2] Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan

Correspondence: Chiao-Ling Kuo (kuo@chiaoling.com)

Abstract. Social media enables the disclosure of real-time crowd situations and provides high accessibility to the public. Thus, social media has emerged as a promising resource for discovering and managing disasters. This study aims to realize an effective process of constructing a spatio-temporal disaster knowledge graph (ST-DKG) from social media for disaster semantic interpretation and achieving an effective solution for disaster management. ST-DKG is constructed using PTT, one of Taiwan's most popular social network platforms, and natural language processing with artificial intelligence (AI) models based on Bidirectional Encoder Representations from Transformers (BERT). ST-DKG addresses issues such as ellipsis and coreference resolution, entity recognition, relation extraction, and the identification of subject-predicate-object triples. Our method not only enhances the access efficiency and interoperability of social awareness information, but also provides bottom-up spatio-temporal disaster relevant knowledge for disaster management.
Keywords. Spatio-temporal disaster knowledge graph, social media, natural language process, disaster management.

1 Introduction

Disasters are defined as *events that seriously disrupt the normal functioning of a community or society. They can be caused by lots of different hazards* (IFRC, 2023). In Taiwan, natural disasters occur frequently because of the location and topography. These disasters often lead to injuries, economic losses, and environmental destruction. Thus, disaster management has become vital in Taiwan. Social awareness information from social media networks has played a vital role in disaster management. Social media allows users to share user-generated contents (O'Reilly, 2007). Thus, it helps reveal real-time crowd situations and enables users to convey emergent messages to others (Lee, 2017). Emergency responders can obtain frontline information directly from crowds by extracting disaster-related information from social media. In this way, communication can be accelerated, and public thoughts can be conveyed to help disaster management with information gathering, response, and distribution of human and material resources. Furthermore, social media can help disaster response to avoid unnecessary deaths and economic losses (Huang et al., 2010).

Disaster-related and spatio-temporal information that frequently appears in free-text social media requires further processing to acquire meaningful details, such as factors causing the disaster, the spatial extent of damages, and the number of injuries or deaths. Disaster information and knowledge can be obtained by applying natural language processing (NLP) and statistical analysis approaches. For example, the evolution (Kitazawa and Hale, 2021), response (Kitazawa and Hale, 2021), and geographic information (Deng, 2011; Kitazawa and Hale, 2021) of a disaster can be recognized.

A knowledge graph should be employed to describe and interpret disaster knowledge formally. The knowledge graph is an instance of domain knowledge that represents domain knowledge based on its advantages in expressivity, standardization, performance, and interoperability (Ontotext, 2023). The basis for constructing a knowledge graph lies in identifying subject-predicate-object (SPO) triples. The method chosen to identify SPO triples depends on the type and characteristics of materials. For instance, NLP is used to extract SPO triples for textual resources. NLP relevant approaches, such as keyword extraction and matching, correlation methods, and machine learning, are employed to construct disaster knowledge graphs by applying formal documents as textual materials (Du et al., 2020; Ma et al., 2022; Zhang and Wang, 2022).

However, few studies have deeply discussed disaster information and spatio-temporal knowledge within the knowledge graph, which is a crucial aspect of effective disaster management. Moreover, research using social media texts to construct a spatio-temporal disaster knowledge graph (ST-DKG) automatically is notably absent. Such research is essential because of the prevalence of big data during a disaster and the necessity to discover knowledge from crowds. In addition, texts shared by crowds are often written in a more unrestricted and casual way than traditional media (i.e., news media). Thus, issues, encompassing uncompleted sentences, replaced words, and incorrect grammar, complicate the extraction of SPO triples because of the ellipsis of subjects and occurrences of pronouns in a sentence (Wu et al., 2022). Hence, this study aims to realize an effective and automatic process of constructing an ST-DKG from social media texts for disaster management using NLP methods

2 Materials and Methods

As illustrated in Fig. 1, a workflow is proposed to develop an ST-DKG for disaster management. The workflow comprises six steps: data acquisition, data pre-processing, ellipsis and coreference resolution, entity recognition, relation extraction, and SPO triple extraction. During the data acquisition, free-text social media data are collected. Data pre-processing step involves filtering records that relate to disaster. Ellipsis and coreference resolution restore replaced and missing subjects and objects in sentences. Entity recognition recognizes disaster-related entities. Relation extraction involves identifying relations, adopted as predicates, through pattern comparison. SPO triple extraction combines entity recognition and relation extraction to help construct the knowledge graph. Other details are provided below.

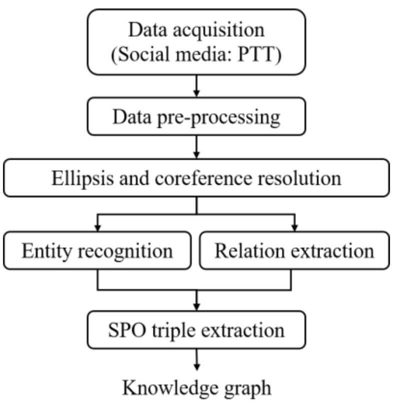

Figure 1. Workflow of this study.

[1] https://term.ptt.cc/

[2] 她們 (they) is the pronoun for a group consisting of females, whereas 他們 (they) is used for a group consisting of males and females.

2.1 Materials and Data Pre-processing

The study materials are collected from PTT Bulletin Board System[1] , the fifth most used social media platform in Taiwan (Taiwan Network Information Center, 2020). PTT Gossiping Board is chosen because it is characterized by nearly real-time activity and high accessibility. The posts are collected by a self-developed web crawler. The dataset used in this study was posted from January 2020 to January 2024. Two keywords, “earthquake” and “disaster” are used for filtering. Posts with a title starting with “news” and “announcement” are excluded because these two types of posts refer to news and announce violations of PTT rules, which are unrelated to this study.

2.2 Ellipsis and Coreference Resolution

Ellipsis and coreference resolution are adopted to construct complete sentences, thereby facilitating the extraction of SPO triples successfully. Ellipsis resolution involves finding the missing entity and restoring it in a sentence. Coreference resolution can restore replaced words, usually pronouns, in a sentence to the referring entity (CKIP, 2023). In the ellipsis process, the Bidirectional Encoder Representations from Transformers (BERT) model is applied to detect whether a sentence is an ellipsis case (Lin et al., 2019). For ellipsis resolution, the BERT question-answering task is used to restore the ellipsis entities (Aralikatte et al., 2019). The input record consists of a question and a paragraph. If a sentence is an ellipsis, it would serve as a question, and the five sentences before and after the ellipsis sentence would be selected as a paragraph when the search window is set as five. For coreference resolution, if a sentence includes a pronoun found by string comparison listed in Table 1, the pronoun would serve as a question, and the five sentences before and after the sentence containing the pronoun would be selected as a paragraph. The answer is the word that should be restored.

Table 1. Types and examples of pronouns.

Type	Example
Person	他 (he), 她 (she), 他們 (they), 她們 (they)[2]
Place	這裡 (this place), 那裡 (that place), 此地 (this place)
Time	這時 (this moment), 此時 (this moment), 那時 (that moment)
Thing Event	牠 (it), 祂 (it), 它 (it)[3], 這 (this), 這個 (this one), 這次 (this time), 這些 (these), 那 (that), 那個 (that one), 那次 (that time), 此次 (this time), 那些 (those)

[3] 牠 (it) is the pronoun used for animals, whereas 祂 (it) is used for gods. Non-living entities can be referred to with 它 (it).

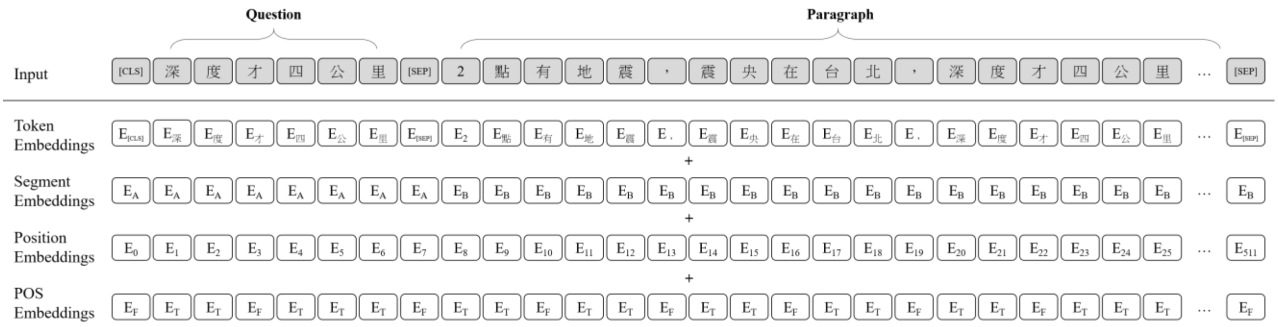

Figure 2. Embeddings of the resolution model in this study.

Table 2. Definition and example for types of entity.

Type	Pronoun	Definition	Example
Formal place name	NS	Administrative districts and road names	台北 (Taipei), 台北市 (Taipei City), 信義路 (Xinyi Road)
Point of Interest (POI)	POI	Landmarks, tourist spots, and stations	大安森林公園 (Daan Park), 台北車站 (Taipei Main Station), 南港展覽館 (Nangang Exhibition Center)
Time instant	TIME	Time with zero duration	下午 12:15:00 (PM 12:15:00), 12/25, 今天 (today), 星期五 (Friday), 中秋節 (Mid Moon Festival)
Time interval	DURATION	Time with an extent or duration	早上 (morning), 晚上 (night)
Number	NUM	Numeric expression	芮氏規模四級 (magnitude 4), 深度 5 公里 (depth 5 km), 500 毫米 (500 mm), 東經 121.5 度 (longitude 121.5°E)
Person	PERSON	Person	市長 (mayor)、台北市民 (Taipei citizen)
Thing	THING	Things, buildings, facilities, and products	大樓 (building), 捷運 (MRT)
Event	EVENT	Disaster-related events and actions	地震 (earthquake), 死亡 (death), 救援 (rescue)
Organization	ORG	A group assembled for a specific purpose	慈濟 (Tzu Chi), 學校 (school), 搜救隊 (rescue team)

Furthermore, the resolution model should include the part of speech (POS) as a feature for improving the performance based on the truth that the resolution words are usually nouns. Thus, POS embedding is included in addition to the three original embeddings utilized. If the input word is a noun or an entity, it would be denoted as 1; otherwise, it would be 0. Then, it is transferred into the embedding format. The embeddings are illustrated in Fig. 2.

2.3 Entity Recognition

Entity recognition is used to identify subjects and objects for SPO triples. The BIO (Beginning, inside, outside) scheme is applied for sequence labeling. The scheme can facilitate a machine learning mechanism to investigate the features of words and achieve token classification. BERT single sentence tagging task (Devlin et al., 2018) is chosen to perform token classification because it benefits context-understanding ability. Table 2 shows the types of entities with examples. The types of entities are defined based on disaster-related studies (Lu, 2019), the Disaster Prevention and Protection Act, and Chinese Knowledge and Information Processing (CKIP) Named Entity Recognition to cover critical aspects of disasters and enhance the spatio-temporal components.

2.4 Relation Extraction

Relation extraction is used to find predicates of SPO triples, including spatial relations, semantic relations, and properties. The string comparison method is applied for spatial and semantic relations. When a string is matched, the predicate is determined. Some examples of strings are listed in Table 3. They are designed based on Wang et al. (2019), Guo et al. (2021), and E-HowNet contributed by CKIP (2023). Properties can be extracted based on the types of entities discussed in Table 2.

Table 3. Types of relations.

	Type	Directed	Example
Semantic relation	is	Yes	是 (is), 為 (is), 即 (in other words)
	has	Yes	有 (has)
	is part of	Yes	部分 (part), 一部分 (a part of)
	happen	Yes	發生 (happen), 出現 (occur)
	result in	Yes	造成 (cause), 導致 (lead to)
	result from	Yes	來自 (come from), 因為 (because)
	influence	Yes	影響 (influence), 作用 (affect)
	action	Yes	other verbs
Spatial relation	topological	Yes	相交 (intersect), 包含 (contain)
	distance	No	鄰近 (near), 相距 (apart), 在 (at)
	orientational	Yes	東北方 (north-east), 偏東方 (near east)
Property	time	No	Entity: TIME, DURATION
	location	No	Entity: NS, POI
	value	No	Entity: NUM

Table 4. Rules of SPO triple extraction.

Rule		Subject	Object	Predicate	Relation type
1[4]		PERSON, THING, EVENT, ORG, NS, POI	TIME, DURATION	time	Property
			NUM	value	
		PERSON, THING, EVENT, ORG	NS, POI	location	
2		EVENT	PERSON, THING, EVENT, ORG, NS, POI	result from	Semantic
		PERSON, THING, EVENT, ORG, NS, POI	EVENT	result in, happen	
		PERSON, THING, EVENT, ORG, NS, POI	PERSON, THING, EVENT, ORG, NS, POI	others verbs	
3	topological	PERSON, THING, NS, POI	PERSON, THING, NS, POI	topological words	Spatial
	distance	PERSON, THING, EVENT, ORG, NS, POI	PERSON, THING, EVENT, ORG, NS, POI	NUM (+distance words)	
	orientational[5]	PERSON, THING, EVENT, ORG, NS, POI	PERSON, THING, EVENT, ORG, NS, POI	orientational words	
4		EVENT	PERSON, THING, ORG	has	Semantic
		PERSON, THING, ORG	EVENT	happen	
		EVENT	EVENT	has	

2.5 SPO Triple Extraction

SPO triples are determined by adopting four rules based on heuristics derived from speaking habits on social media to integrate the results of entity recognition and relation extraction. Table 4 presents the components of SPO triples for each rule. After the acquisition of SPO triples, ST-DKG can be constructed by directly connecting the triples.

2.6 Data and Software Availability

The analyses in this study are conducted using Python Transformers and PyTorch library. The BERT pre-trained model used in this study is provided by CKIP: https://huggingface.co/bert-base-chinese. The demo data and code that support the findings of this study are available at: https://gitops.tw/conference/stdkg.

[4] The order of appearance of the two entities does not need to be considered.

[5] Two patterns describe orientational relations in Chinese: (1) *Entity B* + is situated + *Entity A* + *Orientational words* (*B* is subject and *A* is object); (2) *Entity A* + Orientational words + *Entity B* (*A* is subject and *B* is object).

3 Results

The experimental results, including the ellipsis and coreference resolution, and the construction of ST-DKC are provided.

3.1 Ellipsis and Coreference Resolution

A total of 2,798 sentences are retrieved from 2,601 articles used for the ellipsis detection experiment. The training, validation, and test datasets are split by a ratio of 7:2:1. The proportions of ellipsis and non-ellipsis sentences are the same. Table 5 shows the performance of the ellipsis detection model. Here, precision (P), recall (R), and f1-score (F1) are used to evaluate the model. For the resolution experiment, 1,552 records (1,408 records for ellipsis resolution and 144 records for coreference resolution) are used for the training, validation, and test datasets with a ratio of 7:2:1. The average of the input paragraph length is 74 per record. The performance is shown in Table 6.

Table 5. Performance of ellipsis detection.

P	R	F1
0.724	0.705	0.714

Table 6. Performance of resolution.

Exact match	P	R	F1
0.590	0.596	0.651	0.622

3.2 Entity Recognition

The dataset information of the entity recognition model is shown in Table 7. The ratio for training, validation, and testing datasets is 7:2:1. The average of entities is approximately six per record. Table 8 shows the performance of the model. Type NS and NUM exhibit the best f1-score. However, THING has the worst f1-score, possibly because the descriptions of THING are diverse.

Table 7. Statistics of entity recognition datasets.

	Training	Validation	Test
Number of records			
	3,429	980	490
Number of entities			
NS	5,769	1,591	768
POI	463	460	58
TIME	1,303	393	177
DURATION	298	108	49
NUM	5,308	1,372	709
PERSON	785	210	117
THING	1,637	472	220
EVENT	4,284	1,274	647
ORG	464	142	53
Total	20,311	6,022	2,798

Table 8. Performance of entity recognition.

Type	P	R	F1
NS	0.754	0.735	0.745
POI	0.523	0.596	0.557
TIME	0.445	0.460	0.453
DURATION	0.596	0.571	0.583
NUM	0.752	0.771	0.761
PERSON	0.432	0.557	0.487
THING	0.360	0.352	0.356
EVENT	0.616	0.659	0.637
ORG	0.484	0.585	0.530
Average	0.645	0.663	0.653

3.3 SPO Triples and Knowledge Graph

Three examples of the knowledge graph construction are shown in Fig. 3. The results for each process step are displayed.

In case 1, the pronoun "這 (this)" is identified to refer to its original words "地震 (earthquake)" using the resolution model. The extracted triples are *[地震 (earthquake), has, 捷運 (MRT)]* (R4), *[地震 (earthquake), 造成 (cause), 停駛 (suspend)]* (R2), and *[捷運 (MRT), happen, 停駛 (suspend)]* (R4).

In case 2, an ellipsis occurs in the third sentence; that is, the word "地震 (earthquake)" should appear at the beginning. Here, spatial relations exist in the third sentence, showing the successful example of extracting spatial relations from the social media sentence. The extracted triples are *[地震 (earthquake), location, 桃園 (Taoyuan)]* (R1), *[地震 (earthquake), time, 2020/06/11 09:38:26]* (R1), *[地震 (earthquake), has, 震央 (epicenter)]* (R4), *[宜蘭縣政府 (Yilan County Government), 東偏南方 (east-south), 震央 (epicenter)]* (R3-o), and *[宜蘭縣政府 (Yilan County Government), 69.6 公里 (69.6km), 震央 (epicenter)]* (R3-d).

In case 3, the second and third sentences are ellipsis. After restoring the sentences, the SPO triples can be extracted as *[朋友 (friend), time, 今天 (today)]* (R1), *[朋友 (friend), location, 愛河 (Love River)]* (R1), *[朋友 (friend), 搭 (take), 輕軌 (light rail)]* (R2), *[輕軌 (light rail), 因為 (because), 地震 (earthquake)]* (R2), *[地震 (earthquake), has, 停駛 (suspend)]* (R4), and *[朋友 (friend), 被趕下 (be kicked off), 輕軌 (light rail)]* (R2).

3.4 Application

Applying SPARQL Protocol and RDF Query Language (SPARQL) to ST-DKG can assist in disaster management by providing answers to queries and retrieving relevant knowledge about disasters. Two examples are provided below.

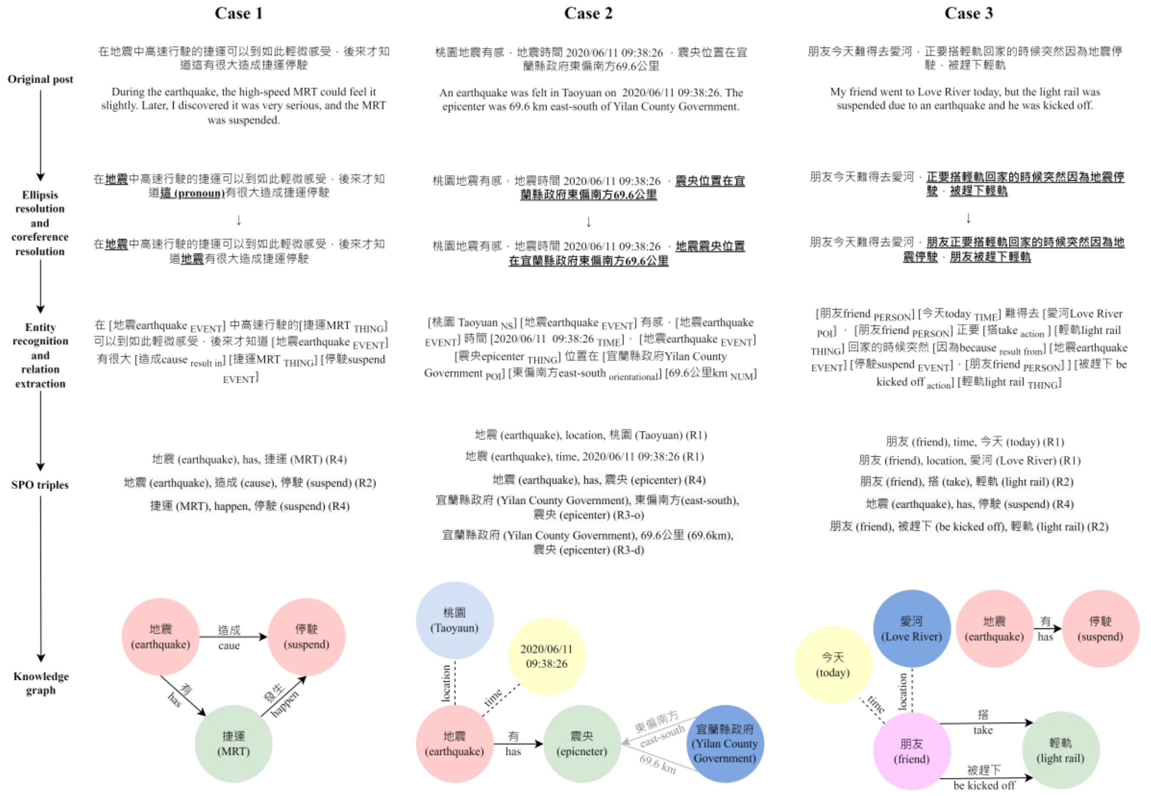

Figure 3. Examples of knowledge graph results.

Query 1. A query for retrieving spatial and temporal information for an earthquake (地震 in line 5 is the Chinese term for earthquake). The result is shown in Table 9.

```
1  PREFIX : <http://stdkg.sgis.tw>
2  SELECT ?event ?time ?location WHERE{
3      ?event :time ?time .
4      ?event :location ?location .
5      FILTER CONTAINS(STR(?event),'地震')
6  }
```

Table 9. Result for Query 1.

event	time	location
地震 (earthquake)	2020/06/11 09:38:26	桃園 (Tauyuan)

Query 2. A query for retrieving what earthquake-related event occurred. The result is shown in Table 10.

```
1  PREFIX : <http://stdkg.sgis.tw>
2  SELECT  ?subject ?relation ?event WHERE {
3      ?subject :happen|:resultIn|:resultFrom|:influence|:action ?event .
4      ?subject ?relation ?event .
5  }
```

Table 10. Result for Query 2.

subject	relation	event
輕軌 (light rail)	happen	停駛 (suspend)
捷運 (MRT)	happen	停駛 (suspend)
地震 (earthquake)	resultIn	停駛 (suspend)
停駛 (suspend)	resultFrom	地震 (earthquake)

3.5 Discussion

The proposed approach can process most situations; however, it may still extract some incorrect SPO triples. For case 1 and case 3, the errors are attributed to spoken Chinese on social media. Thus, the order of words differs from the formal one, thereby complicating the application of the rules for extracting triples, as mentioned in section 2.5.

In Case 1, the triple *[地震(earthquake), has, 捷運(MRT)]* needs to be deleted because it is semantically incorrect in Chinese. In case 3, the incorrect triples are extracted from the second sentence because they should be divided into multiple relations with the correct order. The order should be " 朋友正要搭輕軌回家的時候突然停駛因為地震". The correct triples would be *[輕軌 (light rail), happen, 停駛(suspend)]* and *[停駛(suspend), 因為(because), 地震 (earthquake)]*. The comparison of the original knowledge graph and the revised one is illustrated in Figs. 4 and 5.

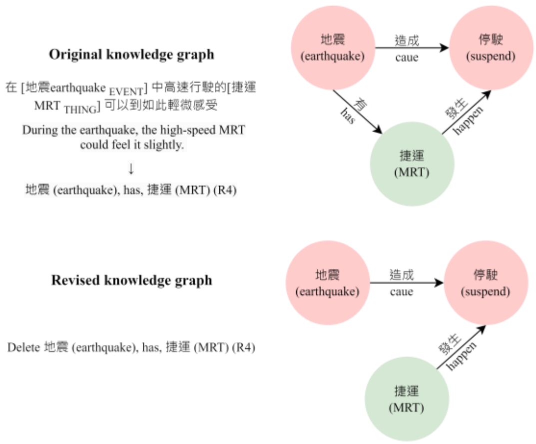

Figure 4. Comparison for case 1.

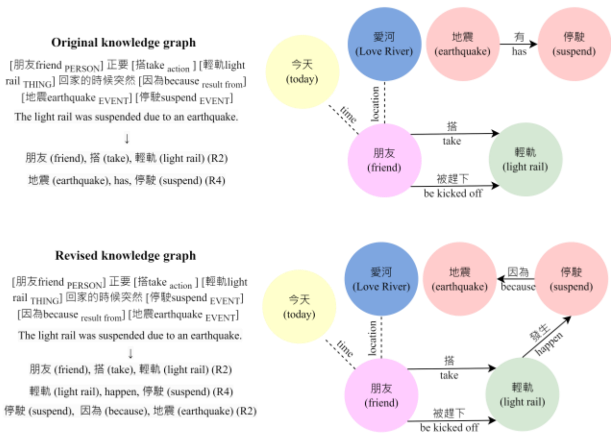

Figure 5. Comparison for case 3.

4 Conclusions

This study focuses on constructing an ST-DKG for the semantic interpretation of disasters using the information gathered from social media. ST-DKG is constructed by applying NLP with AI models. Ellipsis and coreference resolution helps restore the missing part in social media sentences. Moreover, the combination of entity recognition based on BERT, relation extraction, and SPO triple rules contributes to the development of ST-DKG. This process enhances the efficiency and interoperability of accessing social awareness information, thereby providing a bottom-up and spatio-temporal knowledge aspect that traditional knowledge graphs lack in the context of disaster management. Overall, this study provides disaster relief workers and decision-makers with disaster information, thereby empowering them to provide proper reactions by taking well-informed actions during disasters or prevent possible destruction proactively in future disasters by leveraging past knowledge as warnings.

Acknowledgments

The research is supported by the National Science and Technology Council, Taiwan [grant nos. NSTC 111-2121-M-001-001- and 112-2121-M-001-005-].

References

Aralikatte, R., Lamm, M., Hardt, D., & Søgaard, A.: Ellipsis resolution as question answering: An evaluation, arXiv preprint arXiv:1908.11141, 2019.

CKIP Lab: Coreference Resolution, available at: https://ckip.iis.sinica.edu.tw/project/coref, last access: 1 February 2024, 2024.

CKIP Lab: E-HowNet, available at: http://ehownet.iis.sinica.edu.tw/, last access: 1 February 2024, 2024.

Deng, D. P.: Harnessing Social Media for Mapping Disaster Relief Information, Journal of Cartography, 21(2), http://dx.doi.org/10.30006/JC.201110.0001, 2011.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, 2018.

Du, Zhiqiang, Li, Yu, Zhang Yeting, et al.: Knowledge Graph Construction Method on Natural Disaster Emergency[J], Geomatics and Information Science of Wuhan University, 45(9), 1344-1355. http://dx.doi.org/10.13203/j.whugis20200047, 2020.

Guo, X., Qian, H., Wu, F., and Liu, J.: A Method for Constructing Geographical Knowledge Graph from Multisource Data, Sustainability, 13(19), 10602, https://doi.org/10.3390/su131910602, 2021.

Huang, C. M., Chan, E., and Hyder, A. A.: Web 2.0 and Internet social networking: a new tool for disaster management?-Lessons from Taiwan, BMC medical informatics and decision making, 10(1), 1-5, 2010.

IFRC: What is a disaster, available at: https://www.ifrc.org/our-work/disasters-climate-and-crises/what-disaster, last access: 20 January 2024, 2024.

Kitazawa, K. and Hale, S. A.: Social media and early warning systems for natural disasters: A case study of Typhoon Etau in Japan. International Journal of Disaster Risk Reduction, 52, 101926, https://doi.org/10.1016/j.ijdrr.2020.101926, 2021.

Ma, Y., Xie, Z., Li, G., Ma, K., Huang, Z., Qiu, Q., & Liu, H.: Text visualization for geological hazard documents via text mining and natural language processing, Earth Science Informatics, 1-16, 2022.

Lee, Kate Sangwon: Explicit Disaster Response Features in Social Media: Safety Check and Community Help Usage on Facebook during Typhoon Mangkhut. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19), 29, 1–12, https://doi.org/10.1145/3338286.3340140, 2019.

Lin, C. J., Huang, C. H., & Wu, C. H.: Using BERT to process Chinese ellipsis and coreference in clinic dialogues, 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), 414-418, 2019.

Lu, Yu-Ting: Constructing data standards for disaster prevention and notification based on knowledge ontology -A Case Study of Kaohsiung City Government, M.S. thesis, Feng Chia University, Taichung, Taiwan, 2019.

Ontotext: What is a Knowledge Graph?, available at: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/, last access: 20 January 2024, 2024.

O'reilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software, Communications & strategies, (1), 17, 2007.

Taiwan Network Information Center: Network Service Application Overview 2020, available at: https://report.twnic.tw/2020/en/TrendAnalysis_globalCompetitiveness.html, last access: 20 January 2024, 2024.

Wang, S., Zhang, X., Ye, P., Du, M., Lu, Y., & Xue, H.: Geographic knowledge graph (GeoKG): A formalized geographic knowledge representation, ISPRS International Journal of Geo-Information, 8(4), 184, https://doi.org/10.3390/ijgi8040184, 2019.

Wu, D. Z., Lu, Y. D., Tung, C. M., Huang, B. Y., Huang, H. H., Lin, C. D., & Lu, W. H.: Intelligent Future Recreation Harbor Application Service: Taking Kaohsiung Asia New Bay as an Example to Construct a Composite Recreational Knowledge Graph, In Proceedings of the 34th Conference on Computational Linguistics and Speech Processing (ROCLING 2022), 155-163, 2022.

Zhang, M. and Wang, J.: Global Flood Disaster Research Graph Analysis Based on Literature Mining, Applied Sciences, 12(6), 3066, https://doi.org/10.3390/app12063066, 2022.