Transactions in GIS

WILEY

Transactions in GIS

RESEARCH ARTICLE

Mapping a Knowledge Graph of Flooding in Academic Literature Through Full-Text Entity Extraction

Min Zhang[1,2,3] | Juanle Wang[1,4,5] | Xiaodong Zhang[2]

[1]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China | [2]Beijing Municipal Institute of City Planning & Design, Beijing, China | [3]School of Earth and Space Sciences, Peking University, Beijing, China | [4]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China | [5]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

Correspondence: Juanle Wang (wangjl@igsnrr.ac.cn)

Received: 1 March 2025 | **Revised:** 1 March 2025 | **Accepted:** 1 April 2025

Funding: This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0740200-02-01); National Natural Science Foundation of China (42050105); Construction Project of China Knowledge Center for Engineering Sciences and Technology (CKCEST-2023-1-5).

Keywords: academic literature | deep learning | entity extraction | flood disaster corpus | method sentence

ABSTRACT

Academic literature with long-tail characteristics contains rich knowledge resources but is not easily discovered through limited manual knowledge extraction capabilities. This study proposed a refined extraction strategy, transitioning from “full text to sentence,” in the field of flood disaster risk reduction. Sentences describing research methods were identified from the full texts of 5180 articles published between 1990 and 2020. Research method entities—including algorithms, software, and data—were extracted using optimal deep learning models. A flooding knowledge graph was constructed and applied to several flood control scenarios. The results showed that the BiLSTM-CRF model outperformed more complex alternatives. In all, 2144 research methods, 291 software tools, and six types of remote sensing data sources were obtained based on extracted usage method sentences. The flooding knowledge graph contained 42,420 nodes and 78,242 edges. The proposed refined knowledge entity extraction method provides a reference for related knowledge graph mapping based on big data.

1 | Introduction

Academic literature serves as a crucial, verifiable, and transmittable source of knowledge. Presented in standardized formats, it contains peer-reviewed insights, accumulates innovative achievements, and constitutes a valuable professional data resource (Jiang 2021). Over the past half-century, there has been a significant global increase in the volume of academic publications. From 2008 to 2018, the number of studies published in science and engineering journals and conferences worldwide grew at an average annual rate of approximately 3.8% (Oxford 2021). Similarly, the International Association of Scientific, Technical, and Medical Publishers (STM) reported that the number of peer-reviewed journals increased by 5%–6% annually during this period, with 1.5 to 3 million articles published each year (Johnson et al. 2018). However, manually retrieving relevant knowledge from this extensive body of academic literature is time-consuming and labor-intensive. Consequently, new methods for mapping knowledge from available literature are crucial for advancing scientific research.

Knowledge graph can effectively integrate and organize “scattered” research methodology entities within academic literature, supporting intelligent retrieval and recommendation, knowledge reuse, and innovative discovery. Knowledge graph is a structured knowledge base that represents entities and their relationships in a networked form. Research methodologies—such as algorithms, tools, means, data, or technology—are essential assets in academic literature, providing solutions to problems in diverse research fields (Gupta and Manning 2011; Kovačević et al. 2012; Singh et al. 2017). By leveraging knowledge graph construction technologies—such as natural language processing (NLP), machine learning (ML), and database systems—to efficiently extract these entities or relevant sentences from vast academic literature, researchers can be equipped with domain-specific knowledge, thereby enhancing their problem-solving capabilities and efficiency.

We are experiencing a shift in demand from data services to knowledge services in disaster risk reduction (Wang et al. 2020). The issue of "big data, little knowledge" is particularly evident in the field of flood disaster prediction and mitigation. With rapid climate change and frequent extreme weather events, reducing flood disaster risk has become a priority for sustainable global development. The number of natural disasters worldwide increased significantly between 2000 and 2019, especially climate-related disasters, with floods rising from 1389 to 3254 when compared to the preceding 20-year period from 1980 to 1999 (UNDRR 2020). As global climate change and extreme weather events intensify, a significant amount of knowledge regarding flood defense is required. However, there is a stark imbalance between the ability to collect data from multiple sources and the capacity to process this data into meaningful information (Innerebner et al. 2017). Only a small portion of the data are utilized within a narrow scope (Li et al. 2016; Zhu et al. 2016), making it difficult to effectively implement scientific responses to flood disasters.

In response to this challenge, we focused on academic literature related to flood disasters and aimed to use deep learning algorithms to mine flood disaster–related knowledge from full texts. This study constructed vertical academic knowledge graphs in the field of flood disasters to provide a global perspective on flood disaster research and explore potential application scenarios. The primary contributions of this study were summarized as follows:

- We distinguished the sentence types in literature and proposed a "full text-sentences-entities" strategy to extract the entities of flood disaster research methods.
- We constructed a method set of flood disaster research based on the literature, which can provide a valuable reference for recommending methods in flood disaster research.
- A flood disaster knowledge graph was constructed consequently and applied in multi flooding control scenarios, providing a clear and accessible presentation of the current research landscape in flood disaster studies.

The remainder of this study is structured as follows: Section 2 reviews research progress from two perspectives: entity extraction methods and the construction of natural disaster knowledge graphs. Section 3 outlines the data sources, entity annotation rules, and extraction methods used. Section 4 presents the entity extraction results and constructs a flood disaster knowledge graph. Section 5 discusses the construction of flood disaster research methodology collection, the application analysis of various flood scenarios based on the knowledge graph, and the importance and shortcomings of this study. Finally, the conclusions are presented in the last section.

2 | Related Work

Knowledge graph has the capability to integrate a vast array of academic literature related to flood disasters, thereby constructing a knowledge network for disaster prevention and mitigation. Entity extraction is a pivotal step in the construction of knowledge graphs, and the quality of extraction significantly impacts the accuracy of the knowledge graph. This section will provide an overview from two perspectives—entity extraction methods and existing knowledge graphs on natural disasters—to grasp research trends.

2.1 | Evolution of Extraction Technology for Research Method Entities Based on Literature

Manual annotation is a primitive paradigm with high accuracy but low efficiency. Early studies on entity extraction primarily relied on experts to manually annotate methodological entities in literature. For instance, Chu and Ke (2017) systematically analyzed 1900 literatures in the field of library and information science for the first time by manual annotation and established a classification system of research methods. Qasemi and Schumann (2016) manually annotated the research methods and tools in 300 abstracts and evaluated the quality and difficulty of annotation. Although manual annotation can ensure high-quality labeling, it has significant defects such as time-consuming, labor-intensive, high professional threshold, and difficulty in expansion.

The rule-based approach improves efficiency but sacrifices flexibility. To reduce labor costs, scholars turned to automated methods based on rules and dictionaries. For example, Makino et al. (2022) combined deep learning and rule-based methods to extract entities and relationships from literature, addressing the absence of paragraph-level analysis of material synthesis processes. Timakum et al. (2023) adopted a lexicon- and rule-based approach to extract sentence-level entities and their relationships of bipolar disorder from social media and literature. The rule-based approach achieves rapid extraction through predefined rules (such as keyword matching, syntactic templates) but cannot adapt to term variants (such as "HEC-RAS 2D model" abbreviated as "H2D model").

Traditional machine learning (such as CRF and SVM) improves flexibility by introducing statistical models but is limited by feature engineering. Cheng and Li (2017) employed CRF to extract method entities from titles and abstracts, relying on manually designed features (such as part-of-speech tags and word boundaries). Lv et al. (2022) designed an entity-relation extraction method based on the bootstrapping method to reduce labeling requirements through iterative training. While traditional machine learning partially addresses the rigidity of rule-based approaches, it still suffers from critical limitations, including an overreliance on feature engineering, high dependency on labeled data, and weaknesses in modeling long-distance dependencies.

Deep learning models (e.g., BiLSTM-CRF and pretrained language models) have become mainstream due to their advantages in automatically capturing contextual features. Ammar et al. (2017) used a model combining BiLSTM and CRF based on character vectors to identify questions, methods, and data in the full text, avoiding problems in pipeline methods and reducing reliance on feature engineering. Ma and Yuan (2019) used the BiLSTM-CRF model combined with a feature-based named entity knowledge base to extract information such as time, place, and technical entities from the full text. Zheng et al. (2021) developed a joint framework for term extraction and relation extraction tasks using the SciBERT encoder, which performed significantly better than other baselines. However, few scholars distinguished the method entity types of "used" and "mentioned" in entity extraction research based on deep learning, which might cause extraction errors. Therefore, it is urgent to design a low-resource, high-precision entity extraction framework for professional literature to achieve comprehensive and accurate recognition of research method entities under limited annotated data.

2.2 | Research Status and Challenges of Disaster Knowledge Graph

In recent years, the research and application of knowledge graph in the field of natural disasters has grown rapidly, but there are still significant limitations in construction methods and application in real scenarios. From the perspective of data sources, knowledge graph construction overly relies on structured data and expert experience. Ge et al. (2022) constructed a disaster prediction knowledge graph by integrating terrain data, meteorological data, and professional knowledge in the field of disasters. Zhang et al. (2023) used agrometeorological databases to construct a knowledge graph for drought warning. Liu et al. (2023) constructed a knowledge graph of basin flood control based on the historical database of flood and expert-predefined scheduling rules. Jiao and You (2023) relied on 3000 earthquake scene pictures annotated by experts to construct an earthquake rescue knowledge graph. A few studies have attempted to construct knowledge graph based on literature, but the depth of mining was insufficient. Li et al. (2020) mentioned literature in the construction of debris flow disaster knowledge graph, but the entities were still predefined by experts and automatic extraction was not achieved. Du et al. (2020) constructed a knowledge graph for flood disaster emergency response based on Chinese abstracts and disaster event data from the website, but did not mine the full text of literature. Consequently, the mining depth of literature related to natural disaster still needs to be further strengthened.

There is a lack of practical application scenarios of disaster knowledge graphs in disaster risk reduction. Hofmeister et al. (2024) developed a cross-domain knowledge graph for flood risk assessment and tested it on historical cases in simulated environments without implementing it in practical applications. The knowledge graph of debris flow disaster constructed by Li et al. (2020) supported three-dimensional visualization but could not generate emergency decision-making recommendations. Jiao and You (2023) validated the inference algorithm of the earthquake rescue knowledge graph solely on synthetic datasets without connecting to the real rescue process. Only Johnson et al. (2022) attempted to connect the flood disaster knowledge graph with the New York City's building database to support real-time assessment of flood impacts; yet, its entities still required manually entered, reflecting low automation levels.

In summary, current research on disaster knowledge graph is highly dependent on structured data and expert rules, and lacks in-depth mining of unstructured literature data. In addition, the applications of knowledge graph lack deep coupling with various real scenarios.

3 | Methodology

This study proposed a "full text–sentences–entities" strategy to achieve refined extraction of research method entities. In the first stage, sentence-level text was obtained from the full text of academic literature. By distinguishing between usage method sentences, citation method sentences, and nonresearch method sentences, a classification algorithm was employed to extract usage method sentences from the full text. Compared to abstracts, the full texts contain more comprehensive research method entities. In the second stage, entity annotation rules were designed, and a deep learning algorithm was used to extract research method entities in usage method sentences. This approach can avoid the impact of entities in citation method sentences on the accuracy of extraction results. By adopting the three-stage strategy for method entity extraction, a more comprehensive and accurate resource library of flood disaster research methods could be constructed.

Annotation rules for research method entities were established, and *Doccano* (https://github.com/doccano) was used to manually annotate the entities related to algorithms, tools, and data in flood disaster research to create a corpus for flood disaster entity extraction. *Doccano* is an open-source text annotation tool for humans. It provides annotation features for text classification, sequence labeling, and sequence to sequence tasks. Subsequently, various combination models—Bidirectional Encoder Representations from Transformers (BERT), BiLSTM, iterated dilated convolutional neural networks (IDCNN), and CRF—were used to extract the research method entities. Model comparisons and selections were conducted using indicators such as Precision (P), Recall (R), Accuracy (ACC), and F1. This process ultimately enabled the batch extraction of research method entities, forming a collection of flood disaster research method entities. For the construction of the flood knowledge graph, metadata from the articles, such as authors, organizations, and keywords, were integrated with the extracted knowledge, including methods, software tools, and data. On this basis, the applications such as knowledge graph analysis of typical flood disasters and applicable scenario reasoning for specific methods were conducted to discern research hotspots and global trends in the field of flood disasters. The detailed technical process is illustrated in Figure 1.

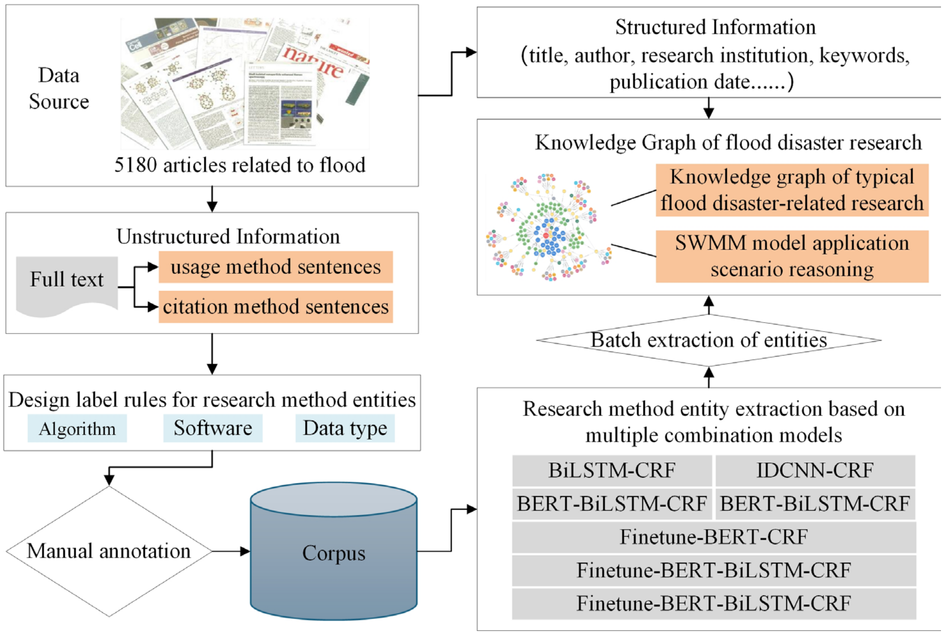

FIGURE 1 | Technical roadmap.

3.1 | Dataset and Preprocessing

The research data were obtained from the Scholar module of the International Disaster Risk Reduction Knowledge Service (DRRKS), affiliated with UNESCO. The DRRKS integrates metadata on disaster-related literature from the Web of Science database by designing a literature search formula based on subject terms and disciplines. Using the keyword "flood," 14,076 articles published between 1990 and 2020 were retrieved. In addition to studies on flood emergency response, the search results also included geochemical, physical, and biological research on flooding. To eliminate the literature unrelated to flood, this study classified flood research topics based on the abstract and obtained 5180 open-access documents that were directly related to floods in the context of disaster prediction, response, and mitigation. To accurately capture the method entities used in each document, referring to relevant studies (Zhang et al. 2020), this study categorized sentences into three types: usage method sentences, citation method sentences, and nonresearch method sentences (Zhang and Wang 2023). Among these, usage method sentences are the target sentences, describing the methods adopted by each study, whereas citation method sentences mainly appear in the literature review section, merely mentioning or citing the methods used by others. We used the Enhanced Representation through kNowledge IntEgration (ERNIE) model to extract 124,196 usage method sentences from the full text for method entity extraction.

3.2 | Design of Annotation Rules for Research Method Entities

Research method entities, including theories, models, algorithms, formulations, tools, data, and software, are typically nouns that appear in method description sentences. In this study, research method entities were characterized into three types: algorithm model (method), software tool (software), and dataset (data). These were labeled using the B-I-O (Begin, Inside, Outside) scheme.

The main guidelines for labeling research method entities were as follows: (1) Method phrases ending with "model," "approach," "algorithm," etc. were labeled as method entities. (2) Method phrases such as "...-based methods/approach/model/algorithm," were also included in the scope of method entity labeling. (3) General terms without qualifiers, such as "investigation," "statistics," "statistical analysis," "experiment," "evaluation," "sorting," and "preprocessing," were not marked unless paired with qualifiers. (4) Resources such as databases and datasets used in the study were labeled. (5) Content in parentheses following a method entity was labeled if it represented an alias but was not labeled if it served as a supplementary description. (6) If the research method entity appeared multiple times in a sentence, it was labeled only once.

Seventeen graduate students mastering geography were recruited in this study to train and carry out entity labeling according to the main guidelines. Double-checking was used to ensure the accuracy and consistency of the annotated corpus. The annotation of ambiguous content was determined through discussion. Examples of method entity annotations are listed in Table 1. In total, 4862 method entities were labeled from 3137 research method sentences. The corpus was divided in a ratio of 9:1 to train the entity extraction model.

3.3 | Entity Extraction for Research Methods Based on Multiple Combined Models

This study employed various combinations, including BiLSTM, IDCNN, and BERT, to conduct extraction research on flood

TABLE 1 | Examples of research method entity labeling.

Word	Label	Word	Label
Flood	O	ArcGIS	B-SOFTWARE
Inundation	O	10.3	I-SOFTWARE
Was	O	.	O
Simulated	O	For	O
Using	O	Accuracy	O
A	O	Assessment	O
Two-dimensional	B-METHOD	,	O
Hydrodynamic	I-METHOD	Multiple	O
Model	I-METHOD	Ground	O
And	O	Truth	O
The	O	Points	O
Connectivity	O	Were	O
Between	O	Collected	O
Wetlands	O	From	O
And	O	Landsat8	B-DATA
Rivers	O	OLI	I-DATA
Were	O	And	O
Calculated	O	Sentinel-2	B-DATA
Using	O	Satellite	I-DATA
Geoprocessing	O	IMAGERIES	I-DATA
Tools	O	.	O
In	O		

disaster research methods. The specific conditions were as follows:

1. The BiLSTM-CRF model comprises Embedding, BiLSTM, and CRF layers. The Embedding layer splits research method sentences into words and embeds them as vectors. These vectors are then fed into the BiLSTM layer to effectively extract the features of the method entities by considering the context information. Finally, the CRF layer is used to decode and annotate the tag sequence output by the BiLSTM layer to obtain the method entity extraction results.
2. The IDCNN-CRF model captures more information by increasing the receptive field of the convolutional kernels. The IDCNN generates logits for each label in sentences related to flood disaster research methods, which are then fed into the CRF for decoding using the Viterbi algorithm. Compared to RNNs, the IDCNN-CRF can accelerate GPU parallelism and reduce training time.
3. The BERT-BiLSTM-CRF and BERT-IDCNN-CRF models utilize BERT to vectorize sentences related to flood disaster research methods into matrices and then input them into the BiLSTM-CRF model to complete the extraction of method entities. Compared to the former, the latter only updates the parameters of the upper layers during model training, whereas the BERT parameters remain unchanged. This approach requires fewer training parameters and offers higher training efficiency than fine-tuning BERT.

Additionally, BERT fine-tuned model training, that is, Finetune-BERT-BiLSTM-CRF, Finetune-BERT-IDCNN-CRF, and Finetune-BERT-CRF, was also performed. These models were based on BERT and updated only the parameters of the later layers, saving considerable time compared to training models from scratch.

4 | Results

4.1 | Model Performance Evaluation

This section compared the performances of various models based on multiple indicators to select the most suitable one for entity extraction. Table 2 listed the model parameter settings, and Table 3 listed the model performance evaluation indicators (P, R, F1, ACC, and Time). Specifically, the BiLSTM-CRF model

TABLE 2 | Parameter setting of research method entity extraction model.

Models	Parameters
BiLSTM-CRF	Learning-rate = 0.01, epoch = 30, Dropout = 0.13, hidden_dim = 200, embedding_dim = 300
IDCNN-CRF	Learning-rate = 0.01, epoch = 30, Dropout = 0.13, filter_nums = 256, idcnn_num = 2, embedding_dim = 300
Bert-BiLSTM-CRF	Learning-rate = 0.001, epoch = 30, Dropout = 0.13, hidden_dim = 200
Bert-IDCNN-CRF	Learning-rate = 0.001, epoch = 30, Dropout = 0.13, filter_nums = 256, idcnn_num = 2
Finetune-Bert-BiLSTM-CRF	Learning-rate = 5e-5, epoch = 30, Dropout = 0.1, hidden_dim = 200
Finetune-Bert-IDCNN-CRF	Learning-rate = 5e-5, epoch = 30, Dropout = 0.1, filter_nums = 256, idcnn_num = 2
Finetune-Bert-CRF	Learning-rate = 1e-5, epoch = 30, Dropout = 0.1

TABLE 3 | Performance comparison of research methods' entity extraction models.

	Performance				
Models	P	R	F1	ACC	Time (min)
BiLSTM-CRF	**0.752**	**0.654**	**0.697**	**0.941**	87
IDCNN-CRF	0.652	0.637	0.649	0.931	**9**
Bert-BiLSTM-CRF	0.650	0.489	0.556	0.910	291
Bert-IDCNN-CRF	0.652	0.513	0.579	0.910	399
Finetune-Bert-BiLSTM-CRF	0.640	0.569	0.601	0.911	3559
Finetune-Bert-IDCNN-CRF	0.648	0.572	0.606	0.913	977
Finetune-Bert-CRF	0.606	0.529	0.600	0.911	935

Note: The bold values in Table 3 represent the best values for each evaluation indicator in reserach methods' entity extraction models.

performed the best with P, R, F1, and ACC values of 0.753, 0.654, 0.697, and 0.941, respectively. In terms of running time, IDCNN-CRF required the shortest time (9 min), followed by BiLSTM-CRF (87 min), which was significantly shorter than that of the other models. BiLSTM learned sequence data through two layers of LSTM neurons, which could learn the contextual information of words, recognized more method entities related to flood disaster research, and achieved high recall rates. The IDCNN-CRF expanded the receptive field of the convolution kernel by adding an expansion width, focusing more on the information and features around the entities, and could better distinguish the boundaries of the entities. Therefore, the BiLSTM was superior in terms of accuracy and recall. However, in the BiLSTM calculation process, the subsequent input depended on the previous result and could not be parallelized. The IDCNN structure was relatively simple and could be parallelized; therefore, it was advantageous in terms of the training speed. Compared to BiLSTM-CRF and IDCNN-CRF, the performance of their improved models, including Bert-BiLSTM-CRF, Finetune-Bert-BiLSTM-CRF, Bert-IDCNN-CRF, Finetune-Bert-BiLSTM-CRF, and Finetune-Bert-CRF, had lower performance index values and longer run times. Therefore, the BiLSTM-CRF model was determined to be the most suitable choice for this study.

The Wilcoxon rank-sum test was used to detect differences among the seven combined models for entity extraction. We considered the P, R, and F1 of the model as sample data and combined the BiLSTM-CRF model with the remaining six combined models in pairs for the Wilcoxon test. The null hypothesis, H0, was that there is no difference between the two models tested, and the alternative hypothesis, H1, was that there is a significant difference between the two models. Using the Python SciPy toolkit, $p = 0.0495$ was obtained, which was less than 0.05. Therefore, the null hypothesis H0 was rejected at a confidence level of 0.05, and the alternative hypothesis H1 was accepted. This indicated that the BiLSTM-CRF model differed from the remaining six combined models and was significantly better than the other models. Therefore, the BiLSTM-CRF was used in this study to perform research on the method of entity extraction for flood disasters. Table 4 presents examples of research methods for entity extraction.

4.2 | Analysis of Entity Extraction Results for Flood Disaster Research Methods

The research data in this paper focused on academic literature in the field of disaster risk reduction, which was more specialized and standardized. Therefore, there were fewer ambiguous entities, but issues such as voice and tense do exist. These included the full name and abbreviation of the same entity (analytical hierarchy process [AHP]), singular and plural forms (general extreme value model and general extreme value models), active and passive voice (regression analysis and regression-based approach), and different tenses (Monte Carlo simulation and Monte Carlo simulated). This study achieved alignment and disambiguation of method entities through manual correction.

The extraction results of the research method entities were integrated and analyzed to understand the overall trends in flood disaster research. A total of 2144 research methods were obtained with a frequency distribution of [13554]. Table 5 lists

TABLE 4 | Example of research method entity extraction.

Number	Research method sentence	Method entities
1	This study contributes a new and simple approach by combining the HEC-HMS and CMIP models to investigate the impact of climate change on flood events.	HEC-HMS, CMIP
2	The proposed approach led to a robust optimal operating policy by considering a significant number of rainfalls through a copula-based Monte Carlo method.	Monte Carlo method
3	A hydrodynamic branched one-dimensional model was created of the Scheldt River and its most important tributaries using the MIKE11 software package from DHI.	MIKE11 software
4	To find the rainfall losses as given by the Soil Conservation Society Curve Number Method, the Geographical Information System (ArcGIS) software was used to obtain the curve number.	Geographical Information System (ArcGIS) software
5	A lack of storm tide height data between tide gauges led to the use of synthetic aperture radar (SAR) images to validate the characteristics approach.	Synthetic aperture radar (SAR) images
6	To represent the topography of urban surface and road network, high-resolution airborne LiDAR data of the study site are collected from Shanghai Survey Bureau	LiDAR data

TABLE 5 | Summary of flood disaster research methods (top 100).

Number	Research method	Quantity	Number	Research method	Quantity
1	Regression model	3284	51	Analysis of variance (ANOVA)	107
2	General extreme value model (GEV)	1794	52	Kriging	107
3	Hydrologic Engineering Center—River Analysis System (HEC-RAS)	1660	53	Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)	101
4	Bayesian approach	1320	54	Fuzzy C-means (FCM)	99
5	Monte Carlo method	1138	55	Action Message Format (AMF)	88
6	Likelihood method	1116	56	El Niño-Southern Oscillation (ENSO)	71
7	Roughness coefficients	859	57	MK test	70
8	Hydrologic Engineering Center—Hydrologic Modeling System (HEC-HMS)	842	58	EM algorithm (EMA)	68
9	Storm Water Management Model (SWMM)	826	59	TUFLOW	68
10	Analytic Hierarchy Process (AHP)	826	60	Hydrologic Engineering Center—Geospatial Hydrologic Modeling System (HEC-GeoHMS	62
11	Gaussian function	706	61	Extended Finite Cloud (EFC) method	61
12	Artificial Neural Network (ANN) model	646	62	Wilcoxon test	61
13	Pearson type III distribution	635	63	*t*-test	60

(Continues)

TABLE 5 | (Continued)

Number	Research method	Quantity	Number	Research method	Quantity
14	Support vector machines (SVMs)	584	64	Particle Swarm Optimization (PSO) algorithm	59
15	Soil Conservation Service (SCS)	541	65	Pettitt test	57
16	Global Climate Model (GCM)	535	66	Continuous ranked probability score (CRPS)	55
17	Shuttle Radar Topography Mission (SRTM)	494	67	Dynamic system response curve (DSRC)	53
18	Peak-over-threshold (POT) model	372	68	Quantile regression technique (QRT)	52
19	Soil and Water Assessment Tool (SWAT)	349	69	Shallow water equations (SWEs)	52
20	Nash–Sutcliffe efficiency (NSE)	343	70	MIKE	50
21	Digital Terrain Model (DTM)	328	71	Maximum likelihood estimation (MLE)	50
22	Genetic algorithm	306	72	Backpropagation neural network (BPNN)	47
23	Receiver operating characteristic (ROC)	299	73	Clustering algorithm	45
24	Akaike information criterion (AIC)	296	74	Parameter regression technique (PRT)	38
25	LISFLOOD Flood Propagation (LISFLOOD-FP)	286	75	Latin hypercube sampling (LHS)	37
26	Variable Infiltration Capacity (VIC) Model	285	76	Structural equation modeling (SEM)	34
27	Kendall test	281	77	Qualitative reasoning (QR)	32
28	HAND model	269	78	Cation exchange capacity (CEC)	30
29	Normalized Difference Vegetation Index (NDVI)	264	79	Data envelopment analysis (DEA)	29
30	Mann-Kendall (MK) test	262	80	Quantitative precipitation estimation (QPE) algorithms	29
31	Flood frequency analysis (FFA)	252	81	Dynamical downscaling (DD) method	28
32	Regional flood frequency analysis (RFFA)	225	82	Autoregressive lognormal model	27
33	Topographic Wetness Index (TWI)	217	83	Probabilistic approach	27
34	Advanced Circulation Model (ADCIRC)	216	84	2D hydraulic model	26
35	Kolmogorov–Smirnov (KS) test	207	85	Public domain stormwater management model (PCSWMM)	25
36	L-moment method	205	86	Liuxihe model	24
37	Generalized Pareto distribution (GPD)	201	87	Smoothing algorithm	24
38	Simulating WAves Nearshore (SWAN)	198	88	Continuous Ranked Probability Skill Score (CRPSS)	23

(Continues)

TABLE 5 | (Continued)

Number	Research method	Quantity	Number	Research method	Quantity
39	Hydrologic Engineering Center—Geospatial River Analysis System (HEC-GeoRAS)	189	89	Diffusion model	23
40	Thiessen polygon method	188	90	Flood event-based spatially-distributed-rainfall-runoff transformations—water balance (FEST-WB)	23
41	Intensity duration frequency (IDF)	185	91	K-nearest neighbor algorithm (KNN)	23
42	K-means	171	92	Random forest (RF)	21
43	Inverse distance weighting (IDW)	159	93	1D river channel model	20
44	Adaptive Neuro-Fuzzy Inference System (ANFIS)	149	94	Digital Terrain Model (DTM)	20
45	Quantitative precipitation forecasting (QPF)	147	95	Change detection algorithm	19
46	Triangulated Irregular Network (TIN)	139	96	EBA4SUB	19
47	Principal component analysis (PCA)	130	97	Self-Organizing Maps (SOM) algorithm	19
48	Topographic Wetness Index-based Distributed Hydrological Model (TOPMODEL)	119	98	Weather Research and Forecasting (WRF) Model	19
49	FLOOD	115	99	PCA	18
50	Horton	108	100	Simple Linear Model (SLM)	18

the top 100 model algorithms in terms of frequency of use. Overall, research methods related to flood disasters could be categorized into three groups: methods in the flood disaster field, statistical methods, and machine learning methods. The regression model and general extreme value (GEV) model were the most frequently used statistical methods, with frequencies of 3284 and 1794, respectively. The former was mainly used for influencing factor analysis and classification research on flood disasters, whereas the latter was primarily used in flood frequency and storm intensity analyses. Among the top 20 high-frequency methods, there were 10 in the category of flood hazards, eight statistical methods, and two machine learning methods. Statistical methods played an essential role in the study of flood disasters.

The data extracted from the research methods mainly included monitoring, remote sensing, field surveys, and historical data. Remote sensing data were the primary data source for flood disaster research, likely due to the advantages of remote sensing, such as wide coverage, low cost, and long-term recording. Therefore, it was particularly suitable for the simultaneous large-area and multilevel monitoring of floods. As shown in Table 6, the remote sensing data source with the highest frequency was radar, including synthetic aperture radar (SAR) and light detection and ranging (LiDAR) with frequencies of 1,221 and 843, respectively. SAR has clear surface water characteristics and all-weather advantages and is more suitable for rainfall scenarios (Guo et al. 2022). LiDAR is a remote sensing tool that can acquire terrain information for all weather conditions and is used for flood disaster risk assessment. Sentinel images have the lowest frequency, and the earliest Sentinel satellite was successfully launched in March 2014. Therefore, it was a slightly new satellite in this historical data analysis. The SAR system carried by the Sentinel-1 satellite can acquire data around the clock under all weather conditions. They are unaffected by extreme weather conditions and can penetrate the cloud layers to provide continuous monitoring data (Chen et al. 2024). In addition, compared to MODIS sensors, the Sentinel-1 satellite has a higher spatial resolution and can obtain more accurate flood inundation areas. Therefore, their use is expected to increase with the collection of more historical data.

In total, 291 software tools were extracted from the research methodology section. Table 7 lists the 100 software tools with the highest frequency of use. High-frequency software tools could be divided into three categories: domain, statistical (e.g., MATLAB and SPSS), and programming language (e.g., R and Python). ArcGIS, MIKE FLOOD, and HEC-RAS were the most commonly used software programs for flood hazard analysis in specific domain categories. ArcGIS had the highest usage frequency of 1223 times, which was far higher than that of the other software tools. MIKE FLOOD was the second most common variant with a frequency of 439.

TABLE 6 | Summary of remote sensing data sources for flood disaster research.

Remote sensing data sources	Quantity
SAR	1221
LiDAR	843
Landsat	503
MODIS	395
ASTER	152
Sentinel	42

4.3 | Knowledge Graph Construction for Flood Disaster Research Methodology

In analyzing the literature, this study used Article_ID as a unique identifier to organize information such as the title, author, keywords, publication year, DOI, and flood disaster subcategories. Combined with the model algorithm, dataset, and software tools extracted above, the relationships between entities were sorted and integrated into ternary groups. These were stored in the Neo4j database to map the flood disaster research knowledge graphs.

Figure 2 illustrates a knowledge graph of the flood disaster research constructed in this study, which contained 42,420 nodes and 78,242 edges. The color of the node represents the entity type, blue bubbles represent research topics in the literature, and yellow bubbles represent each article. The knowledge graph establishes connections between studies through the same entity or attribute values and eventually forms an association network.

5 | Discussion

5.1 | Analysis of Entity Extraction Strategy of Flood Disaster Research Methods

This study innovatively adopted the "full text-sentences-entities" extraction strategy to integrate the methods used in flood disaster research over the past 30 years and formed a knowledge graph. This provided comprehensive and rich information for flood disaster research recommendations. Du et al. (2020) extracted method entities from Chinese abstracts and combined event data from professional websites to construct an emergency knowledge graph for flood disasters. Qiu et al. (2023) considered the Wenchuan earthquake as an example; extracted geological-related entities based on the national geological hazard database, professional websites, and literature; and constructed a knowledge graph of the geological disaster chain. At present, few studies on constructing disaster knowledge graphs were based only on academic literature, lacking in-depth mining of the full text of literatures. The extraction studies with academic literature as one of the data sources included abstract-based and full text–based entity extraction. The abstract was a condensation of the full text, and it was difficult to ensure the comprehensiveness of method entity extraction. However, extracting all method entities based on the full text may extract the "mentation" method entities, rather than the "used" method entities, which will cause erroneous and redundant extraction. The extraction strategy proposed in this study was to obtain the usage method sentences from the full text and then accurately extract the research methods used in the literature. This strategy avoided the interference of citation method sentences and realized comprehensive and refined extraction. The method collections constructed in this paper could recommend models and tools with reference value for flood disaster research.

5.2 | Analysis and Application of Research Methods for Different Types of Floods

The flood knowledge graph could be applied to detailed studies of flood control. There are three types of inland flooding with apparent differences in topography and terrain: urban, flash, and riverine floods (Mishra et al. 2022). We used them as research objects to map their academic knowledge graphs (shown in Figure 3) and conducted a scenario application analysis. In Figure 3, tan-colored bubbles represent academic literature related to flooding research, green bubbles represent the methodological models used, and pink bubbles represent software tools.

Figure 3a presents a knowledge graph of urban flood–related research, which contains 259 nodes and 373 edges and involves 151 studies, 88 modeling approaches, and 19 software tools. The more frequently used modeling methods included the AHP, hydrological model, digital terrain model (DTM), ANN, and the storm water management model (SWMM). Commonly used software tools included ArcGIS, MIKE ENVI, SOBEK, MATLAB, Python, and R. Figure 3b shows the knowledge graph of flash flood–related research, which contains 523 nodes and 789 edges and involved 339 studies, 159 modeling methods, and 24 software tools. In flash flood-related research, the hydrological model, AHP, Bayesian algorithm, HEC-HMS, Gaussian model, ANN, AIC, and line regression models were the most commonly used modeling methods. The most commonly used software packages were ArcGIS, R, SPSS, ENVI, Python, MATLAB, and QGIS. Figure 3c shows a knowledge graph of river flood–related research, which has a significantly smaller number of research studies. It includes 129 nodes and 168 edges, and involves 84 studies, 35 modeling methods, and nine software tools. In river flood research, the Bayesian algorithm, 2D hydraulic model, DTM, GCM model, HEC-RAS, and line regression model appeared more frequently, and the software tools involved were relatively few.

Research on urban flooding, flash flooding, and river flooding differed in its methodological approaches. Urban and flash floods primarily focused on hydrological models to simulate complex hydrological processes and phenomena. In contrast, river flood studies generally used hydraulic models to simulate the dynamics of water in natural environments, such as the diffusion of pollutants in water. Across all three types of flooding, statistical models, including AHP, Bayesian algorithms, Gaussian models, and linear regression models, were commonly used, highlighting the important role of statistical methods in

TABLE 7 | Summary of flood disaster research tools (top 100).

Number	Software	Quantity	Number	Software	Quantity
1	ArcGIS	1223	51	PEST software	3
2	MIKE FLOOD	439	52	SMADA software	3
3	R	159	53	STREAM_2D	3
4	MATLAB	135	54	Surface Water Modeling System (SMS) software	3
5	Python	92	55	Surfer software	3
6	HEC-RAS Software	82	59	TauDEM software	3
7	SOBEK	74	57	Triangle software	3
8	SPSS	71	58	VIIRS NOAA GMU Flood	3
9	ENVI	61	59	Adobe Photoshop CS3 software	2
10	Google Earth	54	60	Agisoft Photoscan software	2
11	Quantum GIS (QGIS)	54	61	Arc/Info software	2
12	Google Earth Engine (GEE)	49	62	ArgusONE commercial software	2
13	HEC-HMS Software	29	63	CES-AES software	2
14	ERDAS Imagine Software	19	64	ClimGen software	2
15	Weka Software	14	65	Fusion software	2
16	Floodwater Depth Estimation Tool (FwDET)	13	66	GAMMA software	2
17	HBV software	11	67	GATE's ontology tools	2
18	Iber software	11	68	GeoCA-Urban software	2
19	EasyFit software	9	69	GeoHECRAS software	2
20	FLO-2D software	9	70	HAZUS-MH software	2
21	eCognition	8	71	HEC-2 software	2
22	FLIKE software	8	72	HEC-SSP software	2
23	FLOW-3D	8	73	HUGIN software	2
24	Concept Model Developer (CMD)	7	74	Hydrospect software	2
25	Super Decision software	7	75	ICPR v.4 software	2
26	Watershed Modeling System (WMS) software	7	76	IDRISI software	2
27	Hydrologic Engineering Center—Geospatial River Analysis System (HEC-GeoRAS)	6	77	Info Works River System software	2
28	Environmental Protection Agency Storm Water Management Model (EPA SWMM)	5	78	IRIS software	2
29	Hydrologic Engineering Center's Geospatial Hydrologic Modeling System (HEC-GeoHMS)	5	79	LATIS tool	2
30	NVivo qualitative software	5	80	Leica Geo Office GPS software	2
31	PeakFQ software	5	81	Library	2
32	SWMM software	5	82	Meteonorm V7 software	2

(Continues)

TABLE 7 | (Continued)

Number	Software	Quantity	Number	Software	Quantity
33	WINFAP-FEH software	5	83	OpenBUGS software	2
34	DassFlow	4	84	OSTRICH (Optimization Software Toolkit for Research Involving Computational Heuristics)	2
35	FLIAT tool	4	85	PC-ORD 6 software	2
36	Hydrognomon software	4	86	Pix4DMapper software	2
37	Hyfran software	4	87	PostgreSQL	2
38	ILWIS software	4	88	RainyDay software	2
39	iPresas UrbanSimp software	4	89	Rating Curve based Automatic Flood Forecasting (RCAFF)	2
40	PCRaster software	4	90	SMAA-TRI software	2
41	Sentinel Application Platform (SAR)	4	91	Structure from Motion software (SfM)	2
42	Sentinel Application Platform	4	92	SGeMS geostatistical software	2
43	XPSWMM software	4	93	Water Modeler software	2
44	Blue Kenue	3	94	SmartPLS 3.0 software	2
45	DFlow FlexibleMesh (DFLOW-FM)	3	95	Sobek 1D-2D Rural software	2
46	DORIS software	3	96	Soil and Water Assessment Tool (SWAT)	2
47	FloodCalc Urban	3	97	Stata	2
48	Hazus-MH	3	98	SYSTAT software	2
49	Java	3	99	Tapenade software	2
50	OpendTect software	3	100	TELEMAC software	2

flood research. However, neural network models, such as ANN, were used less frequently in comparison.

The software tools used in related studies on the three types of flood disasters were relatively similar. Commonly used geographical software, such as ArcGIS, ENVI, QGIS, 3D geological modeling software, and Photoscan, were also applicable to flood disaster research. This indicated that the software tools used in the study of the three inland flooding areas were universal software for disaster science, geography, and statistics. However, there were clear differences in the models used for different emergency tasks in flood disasters.

The constructed knowledge graph could also be used to map specific flood emergency event knowledge. Using Hurricane Katrina in New Orleans (Louisiana, USA) in 2005 as an example, this study revealed the research dynamics related to urban flooding events based on a knowledge graph of flood disasters. Hurricane Katrina, a Category 5 hurricane, struck the United States in August 2005, resulting in the flooding of residential areas in the city of New Orleans, leaving millions of homes without power, turning it into an "ocean city." Figure 4 shows the knowledge graph of the Hurricane Katrina event, which contained 182 nodes and 255 edges and involved 113 studies, 56 modeling methods, and 12 software tools. The high-frequency research methods used included the ADCIRC model, Bayesian algorithm, GEV, DTM, AHP, calibration algorithm, hydrological model, family copula approach, HEC-GeoRAS, ANOVA, CMIP3, Gaussian function, linear regression, and principal component analysis (PCA). The software tools used in this event study were ArcGIS, SLOSH, MATLAB, MIKE, Python, AMOS, iPresas, ERDAS IMAGINE, SPSS, Risk Scale, SPAD, and TELEMAC. The models and tools used in the event study have been included in urban flood–related research. This suggested that the flood hazard knowledge graph can recommend more modeling methods and tools for specific applications and provide references for different flood event studies.

5.3 | Analysis of Applicable Scenarios of SWMM

The knowledge in the flood knowledge graph could also be mined from the dimensions of model application. The SWMM is primarily used for single-event or long-term (continuous) simulations of runoff flow and water quality in urban areas. SWMM is capable of hydrodynamic modeling, accounting for hydrological processes, and pollutant loading estimates. It has evolved into a desktop program for Microsoft Windows

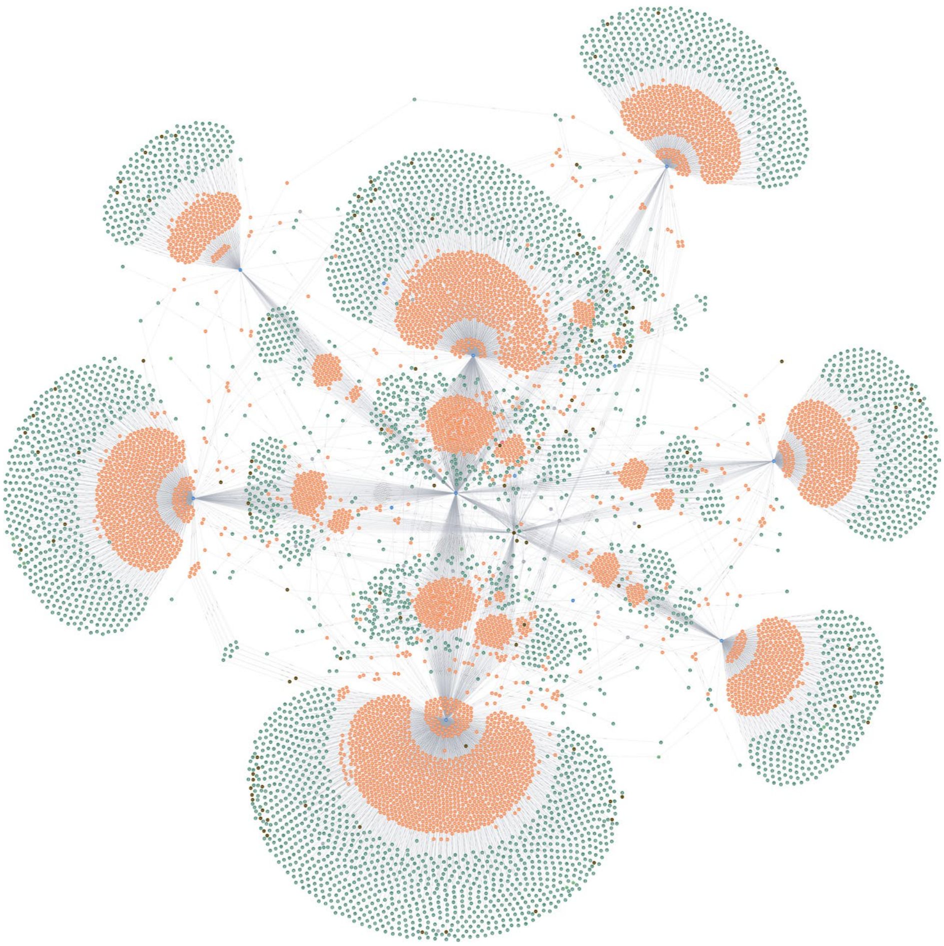

FIGURE 2 | Visualization of knowledge graph for flood disaster research.

and has been widely used in research of urban flood simulation (Barbosa et al. 2012; Chow et al. 2012; Xu et al. 2016; Zeng et al. 2017).

Using SWMM as the root node, the application status of the SWMM was inferred using a flood disaster knowledge graph. A knowledge graph of SWMM model–related research is shown in Figure 5. The graph contains 149 nodes and 398 edges, involving 114 studies, 18 software tools, nine flood hazard subcategories, and seven topics. In terms of flood subcategories, the SWMM was primarily used for urban, storm surge, flash, ponding, and coastal floods. The SWMM model covered all research topics but mainly focused on disaster management, simulation, warning, and disaster risk. The SWMM model was mainly combined with software tools (e.g., ArcGIS and ENVI) to study flood disaster problems. This demonstrated that the knowledge graph of the constructed flood disaster can be used to invert and launch the main areas and scenarios of any model and method for practical applications.

5.4 | Limitations and Prospects of Knowledge Graph Construction

There are some deficiencies in the construction of knowledge graphs that need to be improved. To build a high-quality annotated corpus for model training of entity extraction, 17 graduate students mastering geography were recruited to train and carry out entity labeling according to the designed labeling rules, taking more than 40 days. It can be seen that manual labeling of data required high costs, including manpower and time, and it has a certain degree of subjectivity. Therefore, machine learning models or large language models will be used for automatic annotation to reduce dependence on manually annotated data, form a higher quality and more objective corpus, and lay a good data support for accurate entity extraction.

In the model selection stage of entity extraction, this paper comprehensively considered domain adaptability, training efficiency, and implementation cost. Although large models like GPT excel in general domains, BiLSTM-CRF demonstrated superior stability for processing highly specialized terminology in flood disaster literature, particularly under limited annotated data (3137 labeled samples). Its advantages stem from BiLSTM's sensitivity to local context and CRF's sequence constraint capability. The training time of BiLSTM-CRF was much shorter than that of other combination models (Table 3). In addition, large models (e.g., GPT), with their massive parameters and heavy reliance on general corpora, require extensive fine-tuning to achieve domain adaptability, demanding larger datasets, greater computational resources, and higher training costs. This study prioritizes building a domain-adapted knowledge extraction method over pursuing cutting-edge

FIGURE 3 | Knowledge graph of research methods used to study different types of flood disasters. (a) Knowledge graph of urban flood–related studies. (b) Knowledge graph of flash flood–related research. (c) Knowledge graph of fluvial flood–related research.

model novelty. Future work will explore lightweight fine-tuning approaches for large models, injecting domain-specific knowledge into LLMs to enhance adaptability.

In the process of knowledge graph construction, disambiguation is a crucial step. This study extracted over 2000 method entities from 5180 documents and efficiently achieved alignment and disambiguation through manual correction. In the future, when more entities are extracted based on massive data, manual alignment will no longer be feasible. Instead, machine learning methods will be considered to achieve rapid and automatic alignment.

6 | Conclusions

Driven by the demand to transform data services into knowledge services for disaster risk reduction, this study proposed a "full text–sentences–entities" strategy to extract the entities of flood

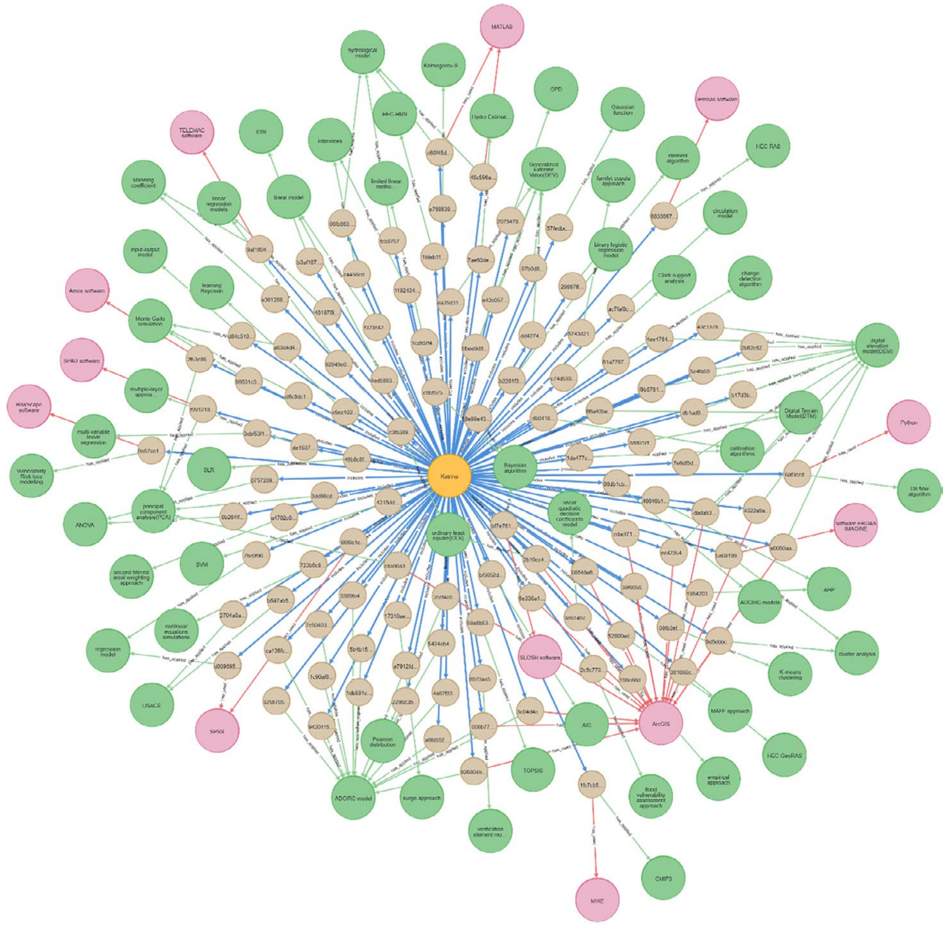

FIGURE 4 | Knowledge graph of Hurricane Katrina research in 2005.

disaster research methods from 5180 articles published between 1990 and 2020 and construct a knowledge graph of flood disaster research. The main conclusions were as follows: (1) The performance of complex algorithms was not always optimal for entity extraction, and dataset scale must also be considered. In this study, the BiLSTM-CRF model proved more suitable for extracting flood disaster research methodologies than more complex models. (2) Besides professional methods, algorithms, and tools, statistical methods played a significant role in flood disaster research, with remote sensing images serving as a key resource for flood control. This was based on the extracted knowledge entities, which included 2144 algorithmic models, 291 software tools, and six types of remote sensing data sources. (3) The emergency tasks and research focuses varied across different types of flood disasters. Urban floods and mountain torrents primarily utilized hydraulic models, whereas river floods also leaned heavily on hydraulic models. The innovations of this study included the following. ① A flood disaster research method repository based on academic literature was constructed for the first time. Based on more than 14,000 literatures related to flood disaster, a rich flood disaster research repository was formed, which was a difficult task for manual reading. On this basis, three innovative application scenarios were proposed, including the recommendation of research methods for different types of floods, the applicable methods recommendation for flood event research, and the applicable scenario reasoning of the classic SWMM, which fully utilized the potential value of literature data. ② A three-stage extraction strategy of "full text–sentences–entities" was proposed. Based on the full text of the literatures, the usage method sentences were first extracted and then the research method entities were extracted, which avoided the interference of the method entities in the citation method sentences. This strategy makes the flood disaster research method repository constructed more comprehensive and accurate. Although the existing model (BiLSTM-CRF) was used in this study, the accuracy of method entity extraction was improved through process optimization, providing a reusable standard framework for disaster literature mining.

This study achieved comprehensive and refined extraction of research method entities from flood disaster literature, addressing a gap in disaster-related research. The constructed knowledge graph created a network of associations between various studies, providing a clear and accessible presentation of the current research landscape in flood disaster studies. Furthermore, the proposed knowledge extraction strategy offered new insights for geographic big data mining and addressed the issue of "big data, little knowledge." In future research, we aim to explore implicit knowledge in the literature, such as spatiotemporal relationships and the formation mechanisms of disaster chains, to enrich the flood disaster knowledge graph.

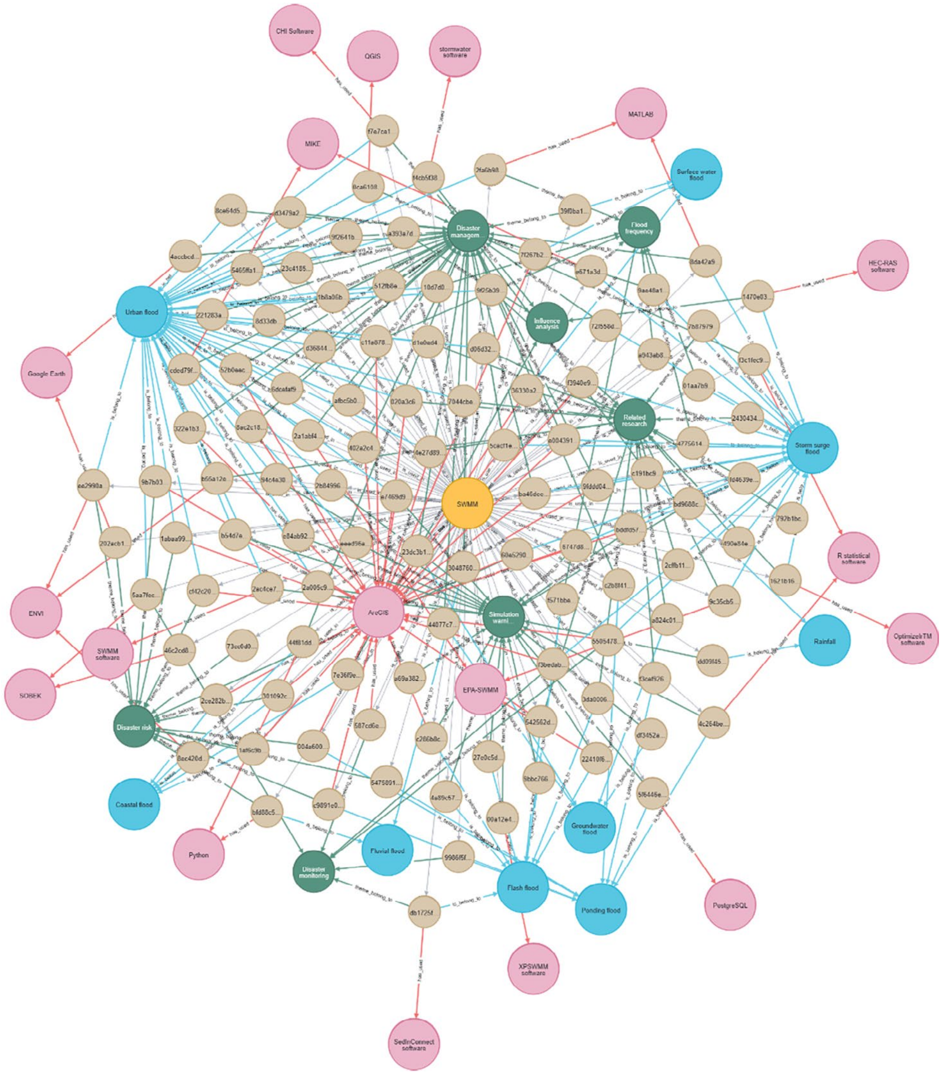

FIGURE 5 | Knowledge graph of SWMM model–related research.

Acknowledgments

This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDB0740200-02-01, National Natural Science Foundation of China under Grant No. 42050105, and the Construction Project of China Knowledge Center for Engineering Sciences and Technology under Grant No. CKCEST-2023-1-5. Thanks to 17 graduate students mastering geography for their support in entity labeling.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ammar, W., M. E. Peters, C. Bhagavatula, and R. Power. 2017. "The AI2 System at SemEval-2017 Task 10 (ScienceIE): Semi-Supervised End-To-End Entity and Relation Extraction." In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 592–596. Association for Computational Linguistics.

Barbosa, A. E., J. N. Fernandes, and L. M. David. 2012. "Key Issues for Sustainable Urban Stormwater Management." *Water Research* 46, no. 20: 6787–6798. https://doi.org/10.1016/j.watres.2012.05.029.

Chen, F., M.-M. Zhang, H. Zhao, W.-G. Guan, and A.-Q. Yang. 2024. "Pakistan's 2022 Floods: Spatial Distribution, Causes and Future Trends From Sentinel-1 SAR Observations." *Remote Sensing of Environment* 304: 114055. https://doi.org/10.1016/j.rse.2024.114055.

Cheng, Q.-K., and X. Li. 2017. *Automatic Recognition of Term Function in Academic Text for Semantic Publishing*. Digital Library Forum.

Chow, M. F., Z. Yusop, and M. E. Toriman. 2012. "Modelling Runoff Quantity and Quality in Tropical Urban Catchments Using Storm Water Management Model." *International Journal of Environmental Science and Technology* 9, no. 4: 737–748. https://doi.org/10.1007/s13762-012-0092-0.

Chu, H.-T., and Q. Ke. 2017. "Research Methods: What's in the Name?" *Library & Information Science Research* 39, no. 4: 284–294. https://doi.org/10.1016/j.lisr.2017.11.001.

Du, Z.-Q., Y. Li, Y.-T. Zhang, Y.-Q. Tan, and W.-H. Zhao. 2020. "Knowledge Graph Construction Method on Natural Disaster Emergency." *Geomatics and Information Science of Wuhan University* 45, no. 9: 1344–1355.

Ge, X.-T., Y. Yang, J.-H. Chen, et al. 2022. "Disaster Prediction Knowledge Graph Based on Multi-Source Spatio-Temporal Information." *Remote Sensing* 14, no. 5: 1214.

Guo, W., H.-Y. Yuan, M. Xue, and P.-Y. Wei. 2022. "Flood Inundation Area Extraction Method of SAR Images Based on Deep Learning." *China Safety Science Journal* 32, no. 4: 177–184.

Gupta, S., and C. Manning. 2011. "Analysing the Dynamics of Research by Extracting Key Aspects of Scientific Papers." In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, 1–9. Asian Federation of Natural Language Processing.

Hofmeister, M., G. Brownbridge, M. Hillman, et al. 2024. "Cross-Domain Flood Risk Assessment for Smart Cities Using Dynamic Knowledge Graphs." *Sustainable Cities and Society* 101: 105113. https://doi.org/10.1016/j.scs.2023.105113.

Innerebner, M., A. Costa, E. Chuprikova, and R. Monsorno. 2017. "Organizing Earth Observation Data Inside a Spatial Data Infrastructure." *Earth Science Informatics* 10: 55–68. https://doi.org/10.1007/s12145-016-0276-0.

Jiang, T. 2021. "A Comparative Study of Term Extraction Schemes in Academic Literature." *Journal of Information Resources Management* 11, no. 1: 112–122.

Jiao, Y.-F., and S.-S. You. 2023. "Rescue Decision via Earthquake Disaster Knowledge Graph Reasoning." *Multimedia Systems* 29, no. 2: 605–614. https://doi.org/10.1007/s00530-022-01002-9.

Johnson, J.-M., T. Narock, J. Singh-Mohudpur, et al. 2022. "Knowledge Graphs to Support Real-Time Flood Impact Evaluation." *AI Magazine* 43, no. 1: 40–45. https://doi.org/10.1002/aaai.12035.

Johnson, R., A. Watkinson, and M. Mabe. 2018. *The STM Report, 5th Edition: An Overview of Scientific and Scholarly Publishing*. STM. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf.

Kovačević, A., Z. Konjović, B. Milosavljević, and G. Nenadic. 2012. "Mining Methodologies From NLP Publications: A Case Study in Automatic Terminology Recognition." *Computer Speech & Language* 26, no. 2: 105–126. https://doi.org/10.1016/j.csl.2011.09.001.

Li, S.-N., S. Dragicevic, F. Anton, et al. 2016. "Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges." *ISPRS Journal of Photogrammetry and Remote Sensing* 115: 119–133. arxiv.org/abs/1511.03010.

Li, W.-L., J. Zhu, Y.-H. Zhang, et al. 2020. "An On-Demand Construction Method of Disaster Scenes for Multilevel Users." *Natural Hazards* 101, no. 2: 409–428. https://doi.org/10.1007/s11069-020-03879-z.

Liu, C.-J., Y.-S. Liu, J.-Q. Wu, et al. 2023. "Exploration on the Knowledge Platform of Digital Twin River Basin Oriented to FEDE for Flood Management." *China Flood & Drought Management* 33, no. 3: 34–41. https://doi.org/10.16867/j.issn.1673-9264.2023075.

Lv, P.-F., Y. Zheng, C.-N. Wang, Y.-Q. Zhu, and W. Liu. 2022. "Research on the Geological Entities Business Relation Extraction Based on the Bootstrapping Method." *Transformations in Business and Economics* 21, no. 2: 322–340.

Ma, J.-X., and H. Yuan. 2019. "Bi-LSTM+CRF-Based Named Entity Recognition in Scientific Papers in the Field of Ecological Restoration Technology." *Proceedings of the Association for Information Science and Technology* 56, no. 1: 186–195. https://doi.org/10.1002/pra2.16.

Makino, K., F. Kuniyoshi, J. Ozawa, and M. Miwa. 2022. "Extracting and Analyzing Inorganic Material Synthesis Procedures in the Literature." *IEEE Access* 10: 31524–31537. https://doi.org/10.1109/ACCESS.2022.3160201.

Mishra, A., S. Mukherjee, B. Merz, et al. 2022. "An Overview of Flood Concepts, Challenges, and Future Directions." *Journal of Hydrologic Engineering* 27, no. 6: 03122001. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002164.

Oxford, E. 2021. "The State of Scientific Research Productivity: The State of Scientific Research Productivity How to Sustain a Critical Engine of Human Progress." https://www.emdgroup.com/content/dam/web/corporate/non-images/scientific-research-productivity/us/Scientific-Research-Productivity-White-Paper-NA.pdf.

Qasemi, Z. B., and A. K. Schumann. 2016. "The ACL RD-TEC 2.0: A Language Resource for Evaluating Term Extraction and Entity Recognition Methods." In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 1862–1868. European Language Resources Association. https://aclanthology.org/L16-1294.

Qiu, Q.-J., L. Wu, K. Ma, Z. Xie, and L.-F. Tao. 2023. "A Knowledge Graph Construction Method for Geohazard Chain for Disaster Emergency Response." *Earth Science* 48, no. 5: 1875–1891.

Singh, M., S. Dan, S. Agarwal, P. Goyal, and A. Mukherjee. 2017. "AppTechMiner: Mining Applications and Techniques From Scientific Articles." In *Proceedings of the Joint Conference on Digital Libraries Joint Conference on Digital Libraries*, 1–8. Association for Computing Machinery. arXiv:1709.03064.

Timakum, T., M. Song, and G. Kim. 2023. "Integrated Entitymetrics Analysis for Health Information on Bipolar Disorder Using Social Media Data and Scientific Literature." *Aslib Journal of Information Management* 75, no. 3: 535–560. https://doi.org/10.1108/AJIM-02-2022-0090.

UNDRR. 2020. "Human Cost of Disasters 2000–2019." https://www.undrr.org/publication/human-cost-disasters-2000-2019.

Wang, J.-L., K. Bu, F. Yang, et al. 2020. "Disaster Risk Reduction Knowledge Service: A Paradigm Shift From Disaster Data Towards Knowledge Services." *Pure and Applied Geophysics* 17: 135–148. https://doi.org/10.1007/s00024-019-02229-w.

Xu, Z.-X., G. Zhao, and T. Cheng. 2016. "'Urban View of the Sea': Challenges and Opportunities Faced by Urban Hydrology." *China Flood & Drought Management* 26, no. 5: 54–55.

Zeng, Z.-Y., Z.-L. Wang, X.-S. Wu, and C.-G. Lai. 2017. "Rainstorm Waterlogging Simulations Based on SWMM and LISFLOOD Models." *Journal of Hydroelectric Engineering* 36, no. 5: 68–77. https://doi.org/10.11660/slfdxb.20170508.

Zhang, C.-Z., Y.-Z. Wang, and R.-P. Wang. 2020. "Constructing the Corpus of Method in the Information Science Domain." *Scientific Information Research* 2, no. 1: 30–45.

Zhang, M., and J.-L. Wang. 2023. "Automatic Extraction of Flooding Control Knowledge From Rich Literature Texts Using Deep Learning." *Applied Sciences* 13: 2115. https://doi.org/10.3390/app13042115.

Zhang, W.-Y., L. Peng, X.-T. Ge, and L. Yang. 2023. "Spatio-Temporal Knowledge Graph-Based Research on Agrometeorological Disaster Monitoring." *Remote Sensing* 15, no. 18: 4403. https://doi.org/10.3390/rs15184403.

Zheng, A.-P., H. Zhao, Z.-H. Luo, C. Feng, X.-P. Liu, and Y.-M. Ye. 2021. "Improving On-Line Scientific Resource Profiling by Exploiting Resource Citation Information in the Literature." *Information Processing & Management* 58, no. 5: 102638. https://doi.org/10.1016/j.ipm.2021.102638.

Zhu, J.-Z., Q. Shi, F.-E. Chen, et al. 2016. "Research Status and Development Trends of Remote Sensing Big Data." *Journal of Image and Graphics* 21, no. 11: 1425–1439.