KNOWLEDGE GRAPH ENABLED REPRESENTATION AND EXPLORATION FOR URBAN HISTORICAL BUILDINGS: A CASE STUDY IN BEIJING, CHINA

Xian Cao[1,2], Xian Guo[1,2,*], Jie Jiang[1,2]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture (BUCEA), Beijing, 100044, China; southerndog@foxmail.com (C.X.), guoxian@bucea.edu.cn (G.X.), jiangjie@bucea.edu.cn (J.J.)
[2] Key Laboratory of Urban Spatial Information, Ministry of Natural Resources of the People's Republic of China, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
* Correspondence: guoxian@bucea.edu.cn

KEY WORDS: Domain Knowledge Graph, Knowledge base, Ontology, Urban Historical Buildings, Multi-hop Relation.

ABSTRACT:

Urban Historical Buildings (UHBs) preserve the past and memory of a city. Collecting and managing UHB knowledge efficiently is of vital importance to protect local culture heritages and meet the demands of sustainable development. However, UHB knowledge is commonly hidden in the texts from different sources, making it hard to access and manage by general methods efficiently. Meanwhile, domain agnostic knowledge graphs are challenging to express the rich semantic and multi-hop relationships among UHB knowledge. To address the above problems, we proposed a general framework for the extraction and management of the key knowledge from free texts about urban historical buildings. In this study, we attempted to build an UHB knowledge graph from free texts obtained from Internet. Firstly, we designed the UHB domain ontology to regulate the knowledge of urban historical building. Next, the structured knowledge was extracted from web text with Natural Language Processing (NLP) technologies. Finally, we built the UHB knowledge base and developed a Beijing UHB knowledge graph. Furthermore, we demonstrated knowledge retrieval and visualization based on UHB knowledge graph. Our experiment shows that mining UHB knowledge from free text is feasible and important for urban historical building conservation efforts.

1. INTRODUCTION

Historic buildings are the evidence to the development of a nation. Historic buildings are generally considered as buildings or structures with some "historical value", that is, people in the present are connected with them in some way through past events. Urban historical building (UHB) is a subset of historic building, which refer to those historical buildings located in the city. The conservation of urban historic buildings has been the focus of historical and cultural conservation work. Collecting and managing UHB knowledge efficiently is of vital importance to protect local culture heritages and meet the demands of sustainable development. Compared to other structured domain-specific knowledge, such as biological species and meteorology, UHB knowledge encompasses rich semantic relations and multi-hop relations, making it difficult to integrate heterogeneous historical data into modern information systems. If use traditional knowledge managing method to represent UHB with relational databases and web systems, we will confront with problems in less comprehensive representation and slow updating.

Knowledge Graph (KG) is a cutting-edge technology proposed by Google in 2012 and quickly applied to the industrial field. The core idea of knowledge graph is using graph structure to manage and represent data. Employing a graph-based organization of knowledge has numerous benefits when compared with the traditional methods. Knowledge graph provides a promising approach for facilitating the access and use of the domain information. More specifically, a knowledge graph is an organized collection of entities and concepts linked via their possible relations, supporting various queries especially those focusing on domain knowledge. According to their application areas, KGs can be divided into open KGs and vertical domain KGs.

Open Knowledge Graphs are universal and encyclopedic, such as DBpedia (Auer et al., 2007), YAGO (Suchanek et al., 2007), Wikidata (Vrandečić et al., 2014), Freebase (Bollacker et al., 2008), which have proven their advantages for various applications. Open KG is not suitable for all application scenarios (Yu et al., 2021). For fields requiring more specialized knowledge, such as medicine, biology, and E-commerce, vertical domain knowledge graphs can better exploit the advantages of knowledge graphs.

It is challenging to integrate UHB information into a KG as the data quality varies and rarely one-size-fits-all solutions can be applied. Whereas domain-agnostic KGs treat UHBs as iconic buildings or a tourist attraction(Colucci et al., 2020), which neglect and omit UHBs' cultural attributes. Worse still, the semantic and multi-hop relationships among UHBs cannot be fully presented by the conventional website-based query provided by domain-agnostic KG.

To address the above issues, we present a UHB Knowledge Graph (UHBKG) for extracting, managing, analyzing, and visualizing the key knowledge from a large amount of UHB data in this paper. The main research and contributions of our article are as follows:

- The UHBKG, a domain knowledge graph of urban historical building knowledge, is specifically developed. Taking the famous historical and cultural city of China, Beijing as the study area. The data for building KG are free texts containing UHB information from websites for government, encyclopedias, and tourism.
- UHBKG designs a domain ontology specifically for describing knowledge of urban historical buildings. Following the 'seven-step' approach proposed by Stanford University(Noy et al., 2001), we defined the UHB domain ontology, which covers nine concepts and eight relationship categories to

standardize the concepts, attributes, and relationships of UHB knowledge.

- We proposed a systematic and semi-automated process for mining UHB knowledge from free text. The whole construction of UHBKG is based on a semi-automatic approach, using deep learning models.
- We explored the visualization and retrieval methods based on Neo4j for supporting the application of UHBKG.

The remainder of this paper is structured as follows. Section 2 provides a review of related work on domain ontology and vertical knowledge graphs. Section 3 presents our general workflow for building the UHB knowledge graph and describe the knowledge extraction methods we used. In Section 4, we present the UHBKG-based retrieval and visualization. Finally, Section 5 summarizes this study and discusses our future directions.

2. RELATED WORK

Knowledge graphs aim to describe the real world by mapping entities, concepts, and the relationships between them into an abstract space. (Song et al., 2021). The vertical domain KGs require high quality domain knowledge (both in accuracy and depth) to assist various complex analysis applications and data support. The ontologies are essential to the KGs, especially the vertical domain KGs. Ontology is an abstract model describing the objective world, an aggregation of restrictions, and a framework for a knowledge graph. An ontology encompasses definitions, formal naming, classifications, and attributes of concepts and entities, as well as categories of relationships and rule constraints.

Vertical domain KGs have been widely applied in Medicine, Biology, E-commerce and other fields, for their customizability and excellent knowledge management capabilities. Among them, Medicine is one of the most widely applied fields of knowledge graphs. For instance, Patrick Ernst et al. developed KnowLife (Ernst et al., 2014), including diseases, symptoms, causes, risk factors, drugs, and side effects , using automatic information extraction methods to populate the relations in the knowledge base. In the biological field, KGs are constructed to research biodiversity. Ozymandias (Page, 2019) describes the relationships of taxa, publications, people, places, specimens, sequences, and institutions. In terms of E-commerce, AliMeKG (Li et al., 2020) is a KG developed for optimizing customers' buying process. It captures user questions, points of interest, product information and their relationship from free text to understand user needs and answer pre-sales questions.

Compared to existing KGs in other fields, KGs concerning historical buildings and cultural heritage remain rare. The British Museum Knowledge Graph (Haase et al., 2019) is considered to be the first case in the field of cultural heritage. Its ontology standardizes various metadata about museum artifacts, such as their names, descriptions, designers and past owners. Worthy of special mention, Valentina et al. proposed ArCO (Carriero et al., 2019), covering millions of Italian cultural heritage knowledge. ArCO provides a practical case implementation of the ontology testing method, which is helpful for students and researchers. In recent years, several studies have attempted to establish a general ontology to regulate cultural heritage knowledge. The CIDOC Conceptual Reference Model (CIDOC CRM) (Bruseker et al., 2017) provides the definitional and normative structure for describing the concepts and relationships of cultural heritage knowledge. It standardizes high-level concepts, so it is largely used as the basis of ontology construction in the field of cultural heritage (includes historical building and landscape heritage). All the works mentioned above have contributed to constructing a knowledge graph of historical heritage. However, their methods still cannot be directly applied to urban historical buildings because of the mismatch of ontologies. Each of these mentioned success cases inspired us to build a vertical domain knowledge graph of urban historic buildings.

3. METHOD

Our primary research process consists in six parts as follows: data acquisition and pre-processing, domain ontology building, knowledge extraction, attribute filling, knowledge fusion and knowledge validation. Figure 1 shows the proposed framework for constructing knowledge graph of urban historical buildings.

3.1 Data acquisition and pre-processing

UHB knowledge includes basic information about an urban historical building (such as name, location, and completion date), historical persons, and historical events related to the UHB. UHB knowledge is hidden in a large amount of text data. These text data in various formats are obtained from different sources, including government websites, encyclopedia websites, and travel and tourism websites. Our data sources mainly include the ‘List of Historical Buildings in Beijing’ published by Beijing Municipal Commission of Planning and Natural Resources, the descriptions of historical buildings on the Baidu Baike and Ctrip websites. From these free texts, we obtained the structured data, semi-structured data and unstructured data through web crawlers. Next, data pre-processing is executed, including data cleaning and integration, format conversion, and data reduction, and the valid UHB data is obtained.

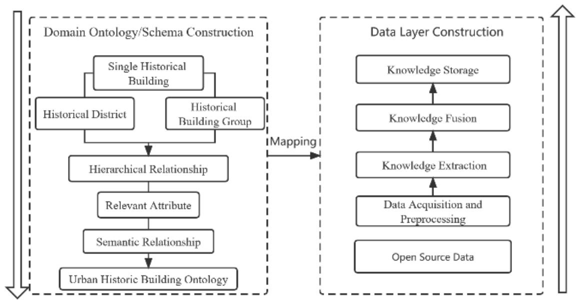

Figure 1. Workflow of the proposed framework for constructing KG for urban historical buildings

3.2 Urban historical building domain ontology building

Domain ontology is intended to describe concepts and relationships between concepts in vertical domains (such as E-commerce and Biology). In this work, to regulate the concepts, attributes and relationships of UHB knowledge, the UHB domain ontology is defined with ‘seven-step’ method. The UHB ontology aims to answer the following questions:

1. What is the definition of urban historical building?
2. What are the basic concepts in the ontology?
3. What are the attributes of urban historical buildings?
4. What are the relationships between entities?

In the ‘Regulations on the Protection of Historical and Cultural Names’, historical buildings are defined as buildings that are more than 50 years old and have certain conservation value. In order to completely describe the information of historical buildings, persons (such as designers, historical personages) and historical events are introduced into the UHB ontology. In addition, the group of historical buildings and the historical district are added to the ontology of UHB in accordance with the ‘List of Historic Buildings Identified in Beijing’ developed by Beijing Municipal Commission of

Planning and Natural Resources. The lack of semantic and multi-hop relations in the ontology are improved as the relations among historic buildings, historic building groups and historic districts are introduced into the UHB ontology.

Concept	Entity example	Attribute example
Actor: Designer	Liang Sicheng, Zhao Dongri, Zhang Bo	Name; Introduction
Actor: Historical Personage	Zhou Enlai, Mao Zedong, Wu Han	Name; Introduction
Actor: Organization	Training Bureau of the General Administration of Sport of China, Tsinghua University	Name; Established Time
Historical Building: Historical Building Group	Hufanglu Community, Zhenwumiao Community	Name; Aliases
Historical Building: Historical District	Zhangzizhonglu North Historical District	Name; Area
Historical Building: Single Historical Building	Former residence of Wu Han, The National Hotel	Address; Aliases; Completion Time
Place: Administrative Area	Xicheng District, Dongcheng District	Name; Area
Place: Street	Tianqiao Street, Qinghuayuan Street	Name; Length
Historical Event	National People's Congress, The May 4th Movement	Occurrence Time

Table 1. Details of all concepts and attributes in UHB Ontology

The core content of the UHB ontology is expressed in the equation:

$$HistoricalBuildingOntology = [SingleHistoricalBuilding, HistoricalBuildingGroup, HistoricDistrict, Relation] \quad (1)$$

Where, Single$Historical$$Building$ is the ontology of single historical building; H$istoricalBuilding$Group refers to the ontology of historical buildings; $HistoricDistrict$ denotes the ontology of historical district; Relation is defend as the semantic relation among single historical building, historical building group and historical district. We define nine subdivision concepts and eight relationships in the UHB ontology. Table 1 shows the details of all concepts and their attributes in UHB Ontology. Table 2 shows the categories of relationships between entities in the UHB ontology.

Relation	Instance
is designed by	Historical Building: Single Historical Building, Historical Building: Historical Building Group ⇒ Actor: Designer
be owned by	Historical Building: Single Historical Building, Historical Building: Historical Building Group ⇒ Actor: Organization
is included in	Historical Building: Single Historical Building, Historical Building: Historical Building Group ⇒ Historical Building: Historical District
is a part of	Historical Building: Single Historical Building ⇒ Historical Building: Historical Building Group
near to	Historical Building ⇒ Place: Street
address	Place: Street ⇒ Place: Administrative Area
participated in	Actor: Historical Personage ⇒ Historical Event
took place in	Historical Event ⇒ Historical Building

Table 2. Categories of relationships between entities in the UHB ontology

3.3 Knowledge Extraction

In general, the data layer is designed to host the graphical information of KG. The knowledge of UHBKG is stored in the graph database with the expression form of knowledge being $<$ head entity, relation, tail entity $>$ or $<$ entity, attribute, attribute value $>$. However, the raw UHB information obtained after pre-processing is still messy and cannot be directly stored in database. To regulate the raw UHB information, we have to extract useful knowledge from semi-structured and unstructured data, and store them in graph database according to the specifications of the ontology layer. The traditional approach, collecting key knowledge manually with the help of experts, can ensure the quality of knowledge. However, this approach is very labor-intensive and inefficient. To reduce the manual dependence, we use a series of NLP algorithms to extract the required knowledge. Knowledge extraction mainly includes two parts, namely Named Entity Recognition and Relation Extraction.

3.3.1 Named Entity Recognition (NER): NER refers to locating and classifying named entities in the text into predefined categories, including names of people, places, and proper nouns within a specific domain.

We choose to adopt the Bidirectional LSTM-CRF (BiLSTM-CRF) model (Huang et al., 2015) to extract entities from the raw data. BiLSTM-CRF model can effectively utilize past and future input features and exhibit high robustness in small datasets (Limsopatham et al., 2016). The BiLSTM-CRF model consists of an embedding layer, a BiLSTM layer and a CRF layer. The embedding layer converts the imported characters into vectors. The BiLSTM layer learns the past and future information on the dependencies of a character in the sequence by forward and backward transmission. The CRF layer is used to learn the dependency information between tags. Figure 2 shows the structure of BiLSTM-CRF network. BiLSTM-CRF requires the training data to be in sequence labeled format. A part of the pre-processed data is marked according to the previously defined UHB ontology.

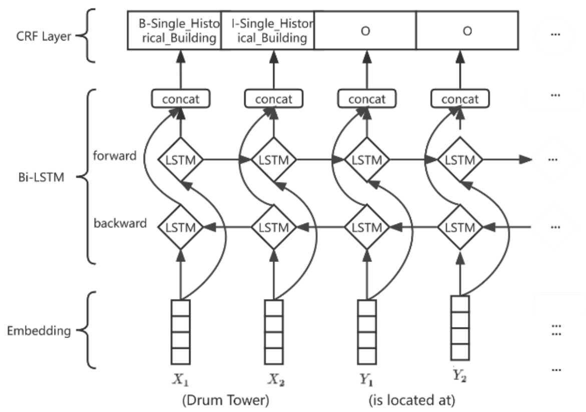

Figure 2. The structure of BiLSTM-CRF network

Next, we convert the format of the labeled data to 'BIO' encoding (Ramshaw et al., 1999). Figure 3 shows the format of the training data. The first letter of each line represents a Chinese character. Each character corresponds to a separate token, e.g., the letter B represents the first character of an entity name, the letter I represents the non-first character of an entity name, and the letter O represents a non-entity. Street represents the category of an entity. For instance, B-Street indicates that the character is the first character of an entity in the Street category. The entities obtained after the above process are inaccurate. We get the accurate results of named entity recognition by manually checking and modifying. A total of 1060 training data are used in our experiments.

3.3.2 Relation Extraction (RE): Relation Extraction aims to recognize the types of relations between entities or concepts. Over the years, researchers have experimentalized various methods on relation extraction. From early pattern matching to current neural networks, existing RE methods have achieved significant progress. Traditional methods suffer from a low recall rate and require a large corpus, which is very tedious and impractical for our case.

For the above reasons, we choose the R-BERT (Wu et al., 2019) model to learn from training data with predefined relations. It is a model merging information from the target entities on the basis of the pre-processed Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) model, which recognizes the relations between entities in a classification approach. For a sentence, we mark the position of the *Subject* in the sentence by inserting an identifier sign '$' before and after the subject. *Object* is marked with the identifier '#' in the same way. At the same time, we place the label at the beginning of the sentence which represents the category of the relation between the subject and the object. Figure 4 shows the structure of the R-BERT model.

For example, after the label information is inserted, a sentence encompassing 'Great Hall of the People' and its designers is treated as follows

$$\begin{gathered}'[is\ designed\ by]\ The\ \$\ Great\ Hall\ of\ the\ People\$ \\ was\ designed\ and\ built\ by\ \#\ Zhang\ Bo\ \# \\ and\ others' \end{gathered} \tag{2}$$

X_1 B-Single_Historical_Building	Y_1 B-Administrative_Area
X_2 I-Single_Historical_Building	Y_2 I-Administrative_Area
X_3 I-Single_Historical_Building	Y_3 I-Administrative_Area
X_4 I-Single_Historical_Building	Z_1 B-Street
X_5 I-Single_Historical_Building	Z_2 I-Street
X_6 I-Single_Historical_Building	Z_3 I-Street
X_7 I-Single_Historical_Building	Z_4 I-Street
X_8 I-Single_Historical_Building	Z_5 I-Street
M_1 O	N_1 O
M_2 O	N_2 O
M_3 O	N_3 O
M_4 O	N_4 O

Figure 3. The example of training data

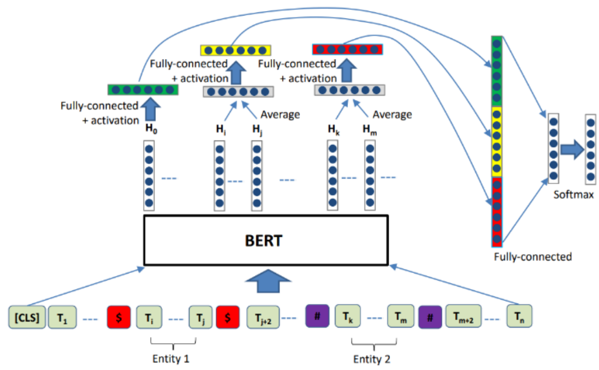

Figure 4. The structure of R-BERT model (Wu et al., 2019)

3.4 Attribute Filling

Attribute filling is to enrich the knowledge density and supplement the semantic information and utility value of the UHBKG data layer, which can support the analysis and visualization of knowledge. On the Internet, encyclopedic knowledge bases hold a large number of entities in a special structure. It is very convenient to crawl the attribute information from these encyclopedic knowledge bases.

We wrote a web crawler to crawl the pages we want from the search engines. The basic process of the web crawler is as follows: firstly, the URLs of the encyclopedia pages linked by entities are used as seeds for collection, and secondly, a depth-first traversal strategy is used to track down the seeds link by link, starting from the seed starting page, and download them to the local machine. Next, we define a linguistic rule-based trigger to extract the attribute values of specific entities automatically.

https://doi.org/10.5194/isprs-archives-XLVIII-3-W2-2022-1-2022

An encyclopedia page usually consists of a string of information blocks. Each information block of the websites can be found by XPath or regular expressions, as each has a specific location on the page and is tagged with a specific label. We only need to write a rule trigger for the valuable blocks of information to complete the attribute extraction.

3.5 Knowledge Fusion

Knowledge fusion refers to the fusion of entities from different KGs or different entities in the same KG. Although the information of entities and relationships are extracted by the above methods, there are large amount of duplicated and redundant information. For instance, ‘Gu Gong’ and ‘Zi Jin Cheng’ are recognized as two objects, but they both represent Forbidden City in fact. In this paper, entity alignment is the main task in the knowledge fusion phase. Entity alignment aims to determine whether two or more entities from different information sources are the same object in the real world. We choose the TransE (Bordes et al., 2013) to align entities. This method maps entities and relations between entities into vectors on the same plane, and determines whether they are the same entity according to the similarity of the vectors. Figure 5 shows the representation of urban historical buildings based on TransE model.

3.6 Knowledge validation

It is essential to validate the knowledge for ensuring the quality of the knowledge graphs. Structured knowledge is generated from various types of UHB data from different sources. However, the incorrect knowledge or invalid entity relations may also be extracted from the UHB data. For improving the quality and credibility of UHBKG, we validated the knowledge with the help of historical architecture experts and knowledge engineers. Next, a knowledge base of urban historical buildings in the context of Beijing is constructed based on the above methods. The content of the knowledge base is continuously updated as new data resources emerge.

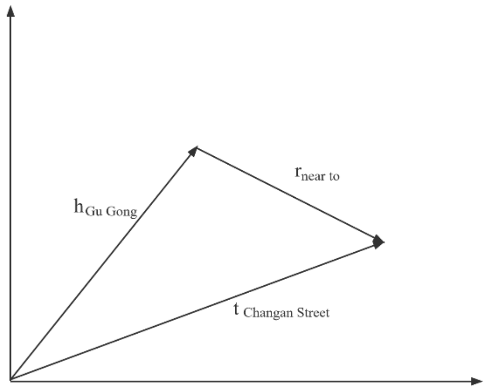

Figure 5. Representation of UHB knowledge in TransE

Where $h_{Gu\ Gong}$ represents 'Gu Gong'; $t_{Changan\ Street}$ represents 'Changan Street'; $r_{near\ to}$ represents the semantic relationship between them.

4. RESULTS

Beijing, the capital of China, is selected as our test area. It has a history of more than 3,000 years as a city and 850 years as a capital. It is the city with the largest number of world heritage sites worldwide and has various historical buildings. At the same time, due to the rapid urban expansion of Beijing, many historical buildings are not well protected and managed. Consequently, it is worthwhile to select Beijing as the experimental area.

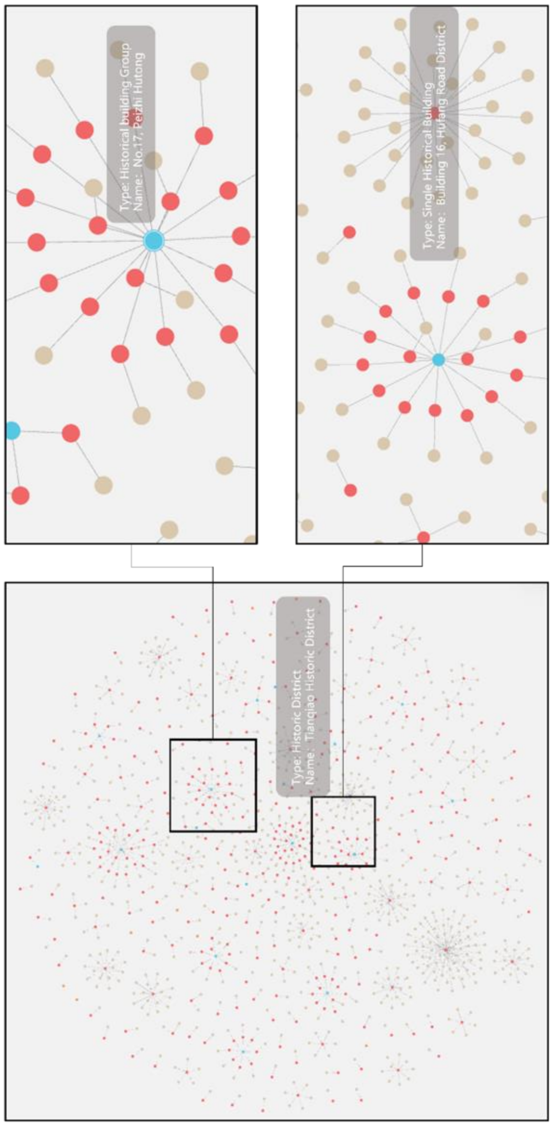

Figure 6. Knowledge graph visualization of Beijing urban historical buildings.

The data of knowledge graph is usually stored and presented in JSON, XML or Table format, which is not intuitive enough. We choose to adopt Neo4j to store structured knowledge. Neo4j is one of the most popular graph databases in existence, specializing in the storage and visualization of KGs. Figure 6 shows the visualization results of UHBKG through Neo4j. UHBKG contains various semantic relationships between historical buildings in Beijing. We use nodes of different colors to denote different concepts. For example, historic districts, historical building groups and single historical buildings are represented in blue, red and brown nodes. The contents shown in Figure 6 delineate all the historical buildings in the western part of Beijing city, as well as their relationships. By analyzing the visualization results, some valuable knowledge can be easily discovered. For instance, we found that the ‘Baiwanzhuang neighborhood’ is a large-scale historical building group, as it has 73 relationships connected to single historical buildings. This knowledge can be verified by consulting the information that the neighborhood is the first residential area in the People's Republic of China and has many Soviet-style buildings. As mentioned, it can be concluded that the constructed UHBKG can well describe the semantic relationships among historical buildings, which reflect the rationality of the ontology we designed.

In addition, UHBKG is developed to provide a fully and improved knowledge system to public. Public and urban managers can quickly search and grasp the key knowledge based on UHBKG. As shown in Figure 7, what historical districts in Beijing's Xicheng District is queried through

$$\text{MATCH (n:`Street`)} - [\text{r}] - \text{(m:`Administrative Districts` \{`AdministrativeDistrictsName`: 'XiCheng'\}) RETURN n, r, m} \tag{3}$$

As shown in Figure 8, what historical buildings and historical buildings group in Beijing's Xicheng District is queried though.

$$\text{MATCH (e:`SingleHistoricalBuilding`)} - [\text{f}] - \text{(c: `HistoricalBuildingGroup`)} - [\text{d}] - \text{(a:`Historic District`)} - [\text{b}] - \text{(n:`Street`)} - [\text{r}] - \text{(m: `AdministrativeDistricts` \{`AdministrativeDistricts Name`: 'XiCheng'\}) RETURN c, e, f} \tag{4}$$

As mentioned above, the key knowledge can be easily retrieved through Cypher language in UHBKG. Knowledge graphs provide high-quality knowledge to the public, and people can discover UHB knowledge quickly due to the semantic association of knowledge.

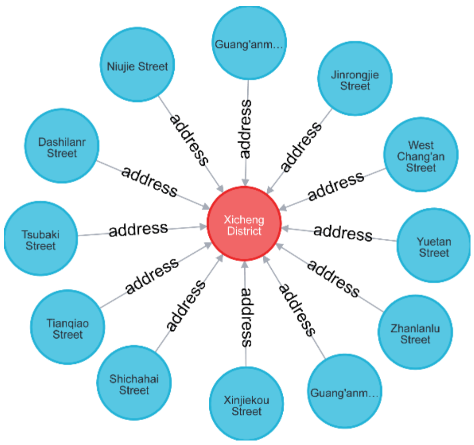

Figure 7 Results of querying 'Historic Districts in Beijing Xicheng District' with our UHBKG.

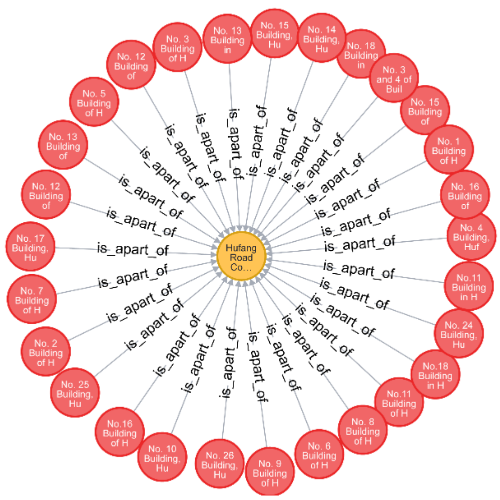

Figure 8 Results of querying 'Historic Buildings and Historical Buildings Group in Beijing Xicheng District' with UHBKG.

5. DISCUSSION AND FUTURE WORK

In this paper, we attempted to build a knowledge graph for urban historical buildings. Firstly, we constructed the UHB domain ontology to regulate the UHB knowledge. Next, the structured knowledge was extracted from web text with NLP technology. Finally, we constructed the UHB knowledge base and developed the Beijing UHBKG by gathering all obtained UHB knowledge, including nine concepts and eight types of relationships. We further demonstrated the preliminary applications based on UHBKG, such as knowledge querying and knowledge visualization.

In the future, we will consider incorporating subtypes of historical buildings into the ontology layer to optimize the structure of UHBKG, which can support the analysis of the distribution of various historical buildings. We will continue to enrich UHB knowledge with the professional literatures on historical buildings to provide more technical support for the conservation of urban historic buildings in the next stage.

6. ACKNOWLEDGEMENT

The research is supported by the National Key Research and Development Program of China (Grant 2021YFE0117500), the R&D Program of Beijing Municipal Education Commission (KM202210016004), the Pyramid Talent Training Project of Beijing University of Civil Engineering and Architecture (JDYC20200322), and the Fundamental Research Funds for Beijing University of Civil Engineering and Architecture (X20044).

7. REFERENCES

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z., 2007. DBpedia: A Nucleus for a Web of Open Data. *Lecture Notes in Computer Science*, vol 4825. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-540-76298-0_52.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J. , 2008. Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge. *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. 1247-1250. doi.org/10.1145/1376616.1376746.

Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., Yakhnenko, O., 2013. Translating Embeddings For Modeling Multi-Relational Data. *Neural Informatino Processing Systems*, 2787-2759.

Bruseker, G., Carboni, N., Guillem, A., 2017. Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM. *Heritage Archaeology in the Digital Age*, 93-131.

Carriero, V. A., Gangemi, A., Mancinelli, M. L., Marinucci, L., Nuzzolese, A. G., Presutti, V., Veninata, C., 2019. ArCo: The Italian Cultural Heritage Knowledge Graph. *International Semantic Web Conference*, Springer, Cham. doi.org/10.1007/978-3-030-30796-7_3.

Colucci, E., Kokla, M., Mostafavi, M. A., Noardo, F., Spanò, A.2020.Semantically Describing Urban Historical Buildings Across Different Levels of Granularity. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B4-2020, 33-40. doi.org/10.5194/isprs-archives-XLIII-B4-2020-33-2020.

Devlin, J., Chang, M.-W., Lee, K.,Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *north american chapter of the association for computational linguistics, abs/1810.04805*, 4171--4186. doi.org/10.18653/v1/N19-1423.

Ernst, P., Meng, C., Siu, A., Weikum, G. ,2014. KnowLife: A Knowledge Graph for Health and Life Sciences*. 2014 IEEE 30th International Conference on Data Engineering*, 1254-1257. doi.org/10.1109/ICDE.2014.6816754.

Haase, P., Herzig, D. M., Kozlov, A., Nikolov, A., Trame, J., 2019. Metaphactory: A Platform for Knowledge Graph Management. *Semantic Web, 10*(6), 1109-1125.

Huang, Z., Wei, X., Kai, Y., 2015. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv: Computation and Language*. doi.org/10.48550/arXiv.1508.01991.

Li, F.-L., Chen, H., Xu, G., Qiu, T., Ji, F., Zhang, J., Chen, H., 2020. AliMeKG: Domain Knowledge Graph Construction and Application in E-commerce. *Proceedings of the 29th ACM International Conference on Information Knowledge Managemen*t, 2581-2588. doi.org/10.1145/3340531.3412685.

Limsopatham, N., Collier, N. H., 2016. Bidirectional LSTM for Named Entity Recognition in Twitter Messages. *In Proceedings of the 26th International Conference on Computational Linguistics*, 145-152.

Noy, N. F., McGuinness, D. L., 2001. Ontology Development 101: A Guide to Creating Your First Ontology. *Knowledge Systems Laboratory, 32*.

Page, R. D. M., 2019. Ozymandias: A Biodiversity Knowledge Graph. *PeerJ, 7*, e6739. doi.org/10.7717/peerj.6739.

Ramshaw, L., Marcus, M., 1999. Text Chunking Using Transformation-Based Learning. *Natural Language Processing Using Very Large Corpora. Text, Speech and Language Technology.* Springer, Dordrecht. doi.org/10.1007/978-94-017-2390-9_10.

Song, X., Zhang, H., Bai, L., 2021. Entity Alignment Between Knowledge Graphs Using Entity Type Matching. *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, 578-589.* doi.org/10.1007/978-3-030-82136-4_47.

Suchanek, F. M., Kasneci, G., Weikum, G., 2007. YAGO: A Core of Semantic Knowledge. *Proceedings of the 16th International Conference on World Wide Web*, 697-706. doi.org/10.1145/1242572.1242667.

Vrandečić, D., rötzsch, M.2014.Wikidata: A Free Collaborative Knowledgebase. *Communications of The ACM, 57*(10), 78–85. doi:10.1145/2629489.

Wu, S., He, Y., 2019. Enriching Pre-trained Language Model with Entity Information for Relation Classification. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2361-2364. doi.org/10.1145/3357384.3358119

Yu, H., Haisheng, L., Mao, D., Qiang, C., 2021. A domain knowledge graph construction method based on Wikipedia. *Journal of Information Science*., 47(6),783-789.