Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

Research Hotspots and Trends of Knowledge Graph Question Answering System----Visual analysis based on CiteSpace

Yong Xu, Xizhi Lv, Hengna Wang*, Xiaoyu Li, Yunke Peng
School of Management Science & Engineering, Anhui University of Finance & Economics, Bengbu 233030, China
*Corresponding Author.

Abstract:

Knowledge graph question answering is the combination of knowledge base technology and intelligent question answering system. It has gradually become a new research direction in the field of artificial intelligence and natural language processing. To explore the research status, development trend and research hotspots in the field of knowledge graph question answering at home and abroad, CiteSpace software was used to draw the visual map, and the time series distribution, core author distribution, main organization distribution and author cooperation network of 662 documents related to the research of knowledge graph question answering at home and abroad in recent ten years were described and compared, and the problems and challenges in domestic knowledge graph question answering research were summarized; based on co-occurrence and cluster analysis of key words, the research hotspots of knowledge graph question answering at home and abroad were summarized, and some suggestions for the next research work were put forward.

Keywords*: Knowledge graph question answering; CiteSpace; Bibliometrics; Cluster analysis*.

I. INTRODUCTION

According to the 47th Statistical Report of China's Internet Development released by China Internet Network Center (CNNIC) on February 3rd, 2021, the number of Internet users in China reached 989 million by the end of 2020, and the penetration rate of Internet use reached 70.4% [1]. With the rapid increase of Internet users in China, the network information resources also show explosive growth, which brings certain challenges for the vast number of information users to get the required information timely and accurately. The traditional way of obtaining information on the Internet is mainly based on search engine, which searches and sorts webpage documents through keyword matching, and returns user query results [2]. These methods have some problems such as difficulty in capturing user's questioning intention and low retrieval efficiency. To improve the information service quality of Internet users, intelligent question answering, as a brand-new way of knowledge acquisition, has gradually

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

become a hot issue in the field of artificial intelligence and natural language processing.

Intelligent question answering can be described as the following tasks: users input natural language questions into a system, and the system automatically identifies the user's intention to ask questions, matches and searches in the built knowledge base, and finally generates the correct answer [3]. Intelligent question answering system can return high-quality answer information to users in a more direct and accurate way, which is the basic form of a new generation of search engines [4], also regarded as one of the key technologies of information intelligent service in the future, and an important means to realize human-computer interaction [5]. An efficient intelligent question answering system depends on the establishment of a perfect knowledge base. As a highly structured knowledge representation and knowledge base construction method, knowledge graph can represent entities and relationships among entities in the objective world in the form of graphs [6], which provides a high-quality data source for the intelligent question answering system. At the same time, it also improves the ability of semantic comprehension of questions and reasoning of answers in the intelligent question answering system. The academic research on knowledge graph question answering started late. However, with the continuous development and improvement of knowledge graph technology, it shows an obvious upward trend. At present, the research on knowledge graph question answering mainly focuses on knowledge graph construction technology, question comprehension, and knowledge graph answer reasoning. In this paper, taking CNKI and the core set database of Web of science as data sources, and combined with bibliometric methods and visual analysis tools, statistical analysis was made on the relevant literature information in the field of knowledge graph question answering from the aspects of the distribution of published papers, the distribution of authors' institutions, the distribution of keywords and cluster analysis, and hot issues in the research of knowledge graph question answering at home and abroad in recent ten years were explored from qualitative and quantitative perspectives. Besides, opportunities and challenges in existing research were summarized, and relevant research suggestions were put forward.

II. DATA SOURCES AND RESEARCH METHODS

2.1 Data Source and Processing

The data of this paper come from CNKI database and Web of Science database, and the retrieval time is limited from January 1st, 2008 to August 30th, 2021. In CNKI database, "topic = knowledge graph and question answering" was used to retrieve and obtain Chinese literature information. English literature information was searched and obtained with the topic "knowledge graph and question answering" in the Web of Science database. A total of 680 original documents were searched, and documents such as meeting notices, news, information and those without no keywords were manually excluded from the original documents. Meanwhile, documents whose abstract content was irrelevant to the "knowledge graph question answering" were excluded, and a total of 662 documents were obtained,

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

including 445 Chinese documents and 217 English documents. They were used as analysis data samples of the visual analysis tool CiteSpace [7].

2.2 Research Methods

Bibliometric analysis is a method that combines the knowledge of mathematics, statistics and other disciplines to study characteristics such as quantity distribution, cooperation, change law of literature in a certain field and quantitatively reveal the development process and research status of an academic field [8]. In this paper, Excel and CiteSpace, a scientific knowledge graphing tool, were used to make statistical and visual analysis of the screened knowledge graphing question answering documents. Firstly, the basic feature information such as the year, author and institution of the document data were statistically analyzed to understand the development history and current situation of research in this field. Secondly, CiteSpace was used to draw the co-occurrence network map of authors in this field at home and abroad, so as to understand the core authors, institutions and cooperation relationships in this field. Finally, the hot research issues in this field were combed and summarized with the characteristic atlas such as literature high-frequency keywords and keyword clustering, and suggestions for future research were put forward.

III. ANALYSIS OF LITERATURE CHARACTERISTICS

3.1 Distribution of Number of Published Papers

The number of published papers is a direct representation of the changes in the number of documents in a certain discipline field [9], which can help us to understand the development process and heat changes of knowledge graph question answering research to a certain extent. At the same time, the comparative analysis of the published papers in this field at home and abroad is also helpful to accurately locate the position of Chinese scholars among international peers.

The numbers of Chinese and English documents published were counted separately, and their changes year by year were shown in Figure 1. As can be seen from Figure 1, the knowledge graph question answering research at home and abroad can be divided into three stages from the perspective of the number of published papers: the initial stage, the gradual rising stage and the rapid growth stage. From 2008 to 2014, the research on knowledge graph question answering system was in the initial stage. At this time, due to the imperfect knowledge graph and knowledge base construction technology, the number of papers published in foreign countries over the years was less than 10, and the number of papers published in Chinese literature was zero. Foreign research mainly focuses on the improvement of traditional search engine and semantic network technology [10-12] and the search and query of graphic

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

database [13-16], while domestic research is basically in a blank state. From 2014 to 2018, with the proposal of Google knowledge graph [17], the number of published papers at home and abroad showed a slow upward trend at this stage. Domestic and foreign research mainly focuses on the construction of domain knowledge graph [18-20], the improvement of knowledge graph question answering algorithm [21-23], and the construction of question answering system [24-26]. At this stage, the number of Chinese documents in this field achieved a "zero" breakthrough, and gradually surpassed that of English documents, which may be related to the Chinese government's emphasis on accelerating the intelligent upgrading of industries and promoting efficient and convenient intelligent services [27]. From 2018 to 2021, with the introduction of improved deep learning models such as BERT model [28-31] and BiLSTM-CRF model [32-35], and the popularity of graphic database represented by Secondary School, the number of published papers at home and abroad increased rapidly, reaching the peak of 167 and 47 respectively in 2020. It shows that knowledge graph question answering has gradually become a hot research topic in the field of artificial intelligence, and domestic scholars have a stronger research interest and enthusiasm in this field.

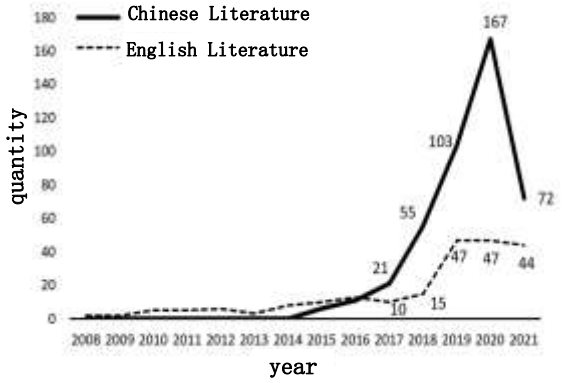

Figure 1: Time distribution diagram of related documents of knowledge graph question answering

3.2 Distribution of Authors

3.2.1 Core author analysis

Core authors refer to a few scholars who have published a leading number of papers in a certain discipline field and lead the development of the discipline. Statistical analysis of the distribution of core authors will help us to understand the academic leaders in this field and track the hot spots in the forefront of disciplines. Firstly, according to Price's Law [36], this paper judged the core authors in the field of knowledge graph question answering. The formula is as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

$$N = 0.749 \times \sqrt{N_{max}} \quad (1)$$

Wherein, N_{max} is the number of published papers issued by the author with the highest number of published papers, and N is the number of published papers issued by the core author at least. According to the statistics of 662 selected documents, the maximum number of published papers at home and abroad is 6, and the judgment standard of core authors in this field at home and abroad is 2. According to statistics, the number of foreign and domestic core authors is 72 and 41 respectively. As can be seen from Table I, the authors with the largest number of published papers abroad and in China are Wang Patrick (6 papers) and Liu Yijun (4 papers), respectively. Wang Patrick mainly focuses on the realization of knowledge graph question answering system in biomedical field and the research of answer reasoning algorithm in question answering system. Liu Yijun mainly discusses the realization path of knowledge graph in intelligent question answering and intelligent service of library [37].

From the statistics of the distribution of core authors, we also found that although the total number of published papers in China was larger than that in foreign countries, the number of published papers by core authors was less than that in foreign countries and the total number of published papers by core authors was also lower than that in foreign countries. The number of published papers by most scholars was 1, showing that the research on knowledge graph question answering by domestic scholars was still basically at the initial stage. The research mainly focuses on the basic concepts, principles and applications of knowledge graph question answering, with a lack of depth and focus. However, the research focus of foreign core scholars has shifted to the bottom-level development, algorithm improvement, and model optimization. Therefore, in the future, domestic scholars can conduct more in-depth research in these aspects to promote the further development of knowledge graph question answering research in China.

TABLE I. Core authors at home and abroad (number of published papers≥3)

No.	Foreign author	Number of published papers	Domestic author	Number of published papers
1	Wang Patrick	6	Liu Yijun	4
2	Bizon Chris	5	Yang Zhihao	3
3	Wang Meng	5	Xiong Taichun	3

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

4	Fecho Karamarie	5	Jia Lirong	3
5	Balhoff James	4	Liu Lihong	3
6	Morton Kenneth	4	He Sheng	3
7	Tropsha Alexander	4	Gao Bo	3
8	Cox Steven	3	Liu Jing	3
9	Kebede Yaphet	3	Wang Xinlei	3
10	Lehmann Jens	3	Li Shuaichi	3

3.2.2 Author's cooperative network analysis

The author co-occurrence network map can directly reflect the influence and cooperation of authors in the field. In this paper, CiteSpace 5 software was used to draw the cooperative network maps of authors in the field of knowledge graph question answering at home and abroad. The year per slice was set to 1, and the top N per slice was set to 10%. The clipping type of maps is Pathfinder. Each node in the graph represents one author, and the connection between nodes reflects the cooperation among authors. There is a positive correlation between the font size and the number of papers published by authors, and there is a positive correlation between the connection thickness and the degree of cooperation among authors.

According to the cooperation network of foreign authors, it is found that there are 1707 connections among 787 nodes, with a network density of 0.0055. There are many connections and dense networks, and the connections between nodes are thick, which shows that the cooperation among foreign scholars is relatively close. At the same time, combined with the distribution of authors' institutions, we find that there are a large number of cross-institutional cooperation among foreign scholars, a large number of joint publications, and few small-scale cooperation sub-networks. We can find two scholars, Chris Bizon and Patrick Wang, are at the core of the network, and have a large number of papers published in cooperation with other scholars, which have a high influence in the field of knowledge graph question answering abroad.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

According to the cooperation network of authors in China, there are 920 connections among 733 nodes, with a network density of 0.0034. The number of connections is small and the network density is low, which shows that the research of Chinese scholars in the field of knowledge graph question answering is scattered, and the cooperation among core authors is sparse. The cooperation is mainly among scholars in the same institution, and mostly exists in the form of "sub-networks" in the map. As can be seen from Figure 2, there are four four-person teams, including a team composed of Wang Xinlei, Li Shuaichi and Lin Hongfei, with Yang Zhihao as the center. Four papers have been published; there are five three-person teams, including a team composed of Lu Qi (the core), Xie Jun and Pan Zhisong. Three papers have been published. The teams formed by these authors are in the leading position in domestic knowledge graph question answering research.

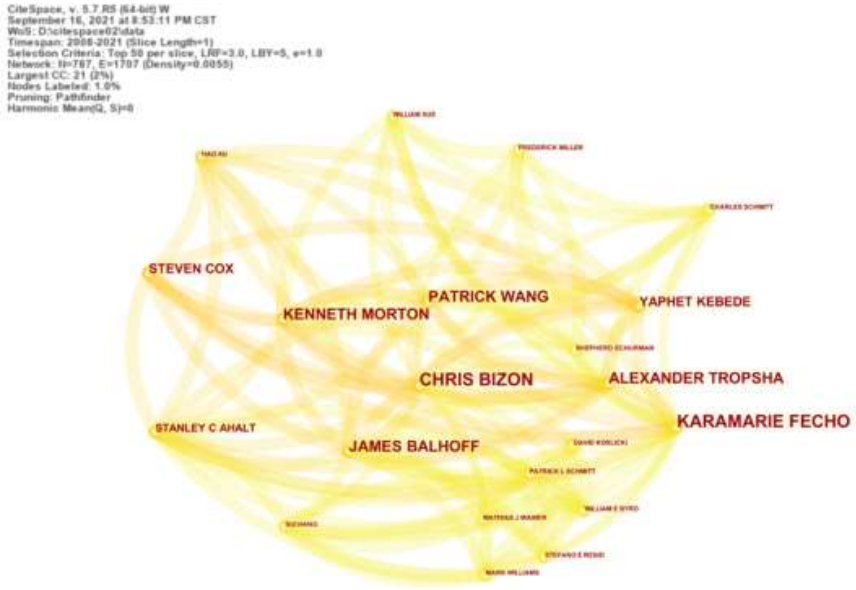

Figure 2: Author co-occurrence network of foreign knowledge graph question answering research

3.3 Institution Distribution

A comparative analysis of the main research institutions at home and abroad in the field of knowledge graph question answering will help us to understand the main research teams in this field and the main force of paper writing. After screening, 662 papers were counted according to the number of papers published by institutions, and the main research institutions in this field at home and abroad were obtained. It can be seen from Table II that the institutions with the largest number of publications abroad and in China are University of California (7 papers), French National Centre for Scientific Research (6 papers), IBM (5 papers), Beijing University of Posts and Telecommunications (21 papers), University of Chinese Academy of Sciences (15 papers) and Harbin Institute of Technology (11 papers).

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

Based on the statistical analysis of the types of major publishing organizations at home and abroad (see Table III for the results), we find that the departments in universities and colleges at home and abroad are the "main force" in the field of knowledge graph question answering research and literature publication, which account for 72.33% and 67.65% of all institutions respectively. Scholars of various departments in universities and colleges mainly focus on the theoretical and algorithmic research of knowledge graph question answering. In addition, research institutions account for 16.99% and 16.18% of the total respectively. Enterprise institutions account for 8.74% and 11.76% of the total respectively. Among these enterprises, there are many large-scale technology enterprises such as Google, IBM, Huawei, etc. They have made some achievements in the research of knowledge graph question answering. For example, Google has fully applied knowledge graph technology to its own search engine technology, and IBM has developed a question answering system, Watson system, on the basis of knowledge graph technology.

Table II. Major institutions at home and abroad

No.	Foreign institutions	Number of published papers	Domestic institutions	Number of published papers
1	University of California system	7	Beijing University of Posts and Telecommunications	21
2	CNRS (Centre national de la recherche scientifique)	6	University of Chinese Academy of Sciences	15
3	IBM (International Business Machines Corporation)	5	Harbin Institute of Technology	11
4	NanYang Technological University	5	University of Electronic Science and Technology of China (UESTC)	11
5	University of North Carolina	5	Dalian University of Technology	10
6	University of Texas System	5	Wuhan University	10

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

7	Cover Appl Technol	4	East China Normal University	10
8	Google Incorporated	4	Chinese Academy of Sciences	8

Table III. Distribution of main institutions at home and abroad

Type of institution	Number of foreign institutions	Proportion (%)	Number of domestic institutions	Proportion (%)
Departments in universities and colleges	149	72.33%	92	67.65%
Research institutions	35	16.99%	22	16.18%
Enterprise	18	8.74%	16	11.76%
Other	4	1.94%	6	4.41%
Total	206	——	136	——

IV. RESEARCH HOTSPOT ANALYSIS

4.1 High-Frequency Keyword Analysis

The research hotspot is a common concern and focus of many related literatures in a certain period of time [38]. The key words are the induction and summary of the core content of the literature, and the key word analysis helps us to explore the high-frequency hotspots in the field of knowledge graph question answering. Co-occurrence analysis of keywords was carried out on 662 domestic and foreign literatures. The year per slice was set to 1, and the top N per slice was set to 10%. The clipping type was Pathfinder. The network map of keyword co-occurrence was drawn (see Figure 3). The node size and font size are positively correlated with the frequency of keyword occurrence, and the thickness of connection line is

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

positively correlated with the co-occurrence degree of keyword. Centrality is one of the indicators to measure the importance of keywords. First, the keywords with the same research topic, such as "knowledge graph" and "Intelligent Question Answering", were eliminated, and then the top ten keywords in the knowledge graph question answering documents at home and abroad were selected (see Table IV). They were analyzed by combining keyword centrality and keyword co-occurrence network map.

Combined with Figure 3 and Table 4, we find that in foreign knowledge graph question answering research, high-frequency keywords mainly include: Knowledge, Ontology, Model, System and Deep Learning. Their centrality in the map is 0.04, 0.27, 0.36, 0.16 and 0.04 respectively. It can be seen that the related research in the field of knowledge graph question answering in foreign countries mainly focuses on the research topic "question answering". By using the characteristics of knowledge graph, namely structured representation and knowledge storage, and combining with deep learning model, the traditional question answering system and algorithm were improved, so as to improve the efficiency of the question answering system.

We can find that in the domestic knowledge graph question answering research, high-frequency keywords mainly include: deep learning, named entity recognition, relationship extraction, entity disambiguation and artificial intelligence. Their centrality index values in the network are 0.2, 0.07, 0.04, 0.06 and 0.16 respectively. It can be seen that domestic knowledge graph question answering research mainly focuses on the research topic "knowledge graph". By exploring the process of knowledge graph construction, such as entity identification, relationship extraction, entity linking and other technologies, knowledge graph technology can be applied to various artificial intelligence services.

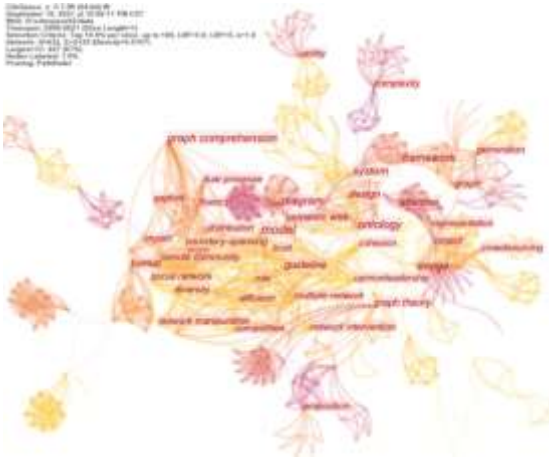

Figure 3: Keyword co-occurrence network of foreign literature of knowledge graph question answering

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

4.2 Keyword Cluster Analysis

Keyword co-occurrence cluster analysis is based on the calculation of keyword co-occurrence frequency. It gathers keywords with high co-occurrence frequency to form a relatively independent "cluster", which can help us to understand the research status and future trend of knowledge graph question answering. In this paper, foreign and domestic keyword maps were clustered separately. The clustering method adopted is LLR algorithm, and the clustering results are shown in Figures 4 and 5.

Combined with Figure 4 and Figure 5, in foreign and domestic keyword clustering maps, the values of clustering modularity are 0.8944 and 0.5888, respectively. Both of them are greater than 0.3, indicating that the clustering structures generated by LLR algorithm are remarkable [39]. At the same time, the average clustering silhouette are 0.969 and 0.8495 respectively. Both of them are greater than 0.5, indicating that the clustering results are convincing.

Table IV. TOP10 high-frequency keywords at home and abroad

No.	Overseas	Frequency	Domestic	Frequency
1	Knowledge	12	Deep learning	64
2	Ontology	12	Named entity recognition	26
3	Semantic web	10	Natural language processing	25
4	Model	10	Relation extraction	17
5	Graph	9	BERT	16
6	System	8	Entity disambiguation	16
7	Deep learning	8	Artificial intelligence	15
8	Algorithm	7	Entity link	14
9	Knowledge graph embedding	7	Graph convolution	14

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

	(Knowledge graph embedding)			
10	Knowledge base	6	Ontology	14

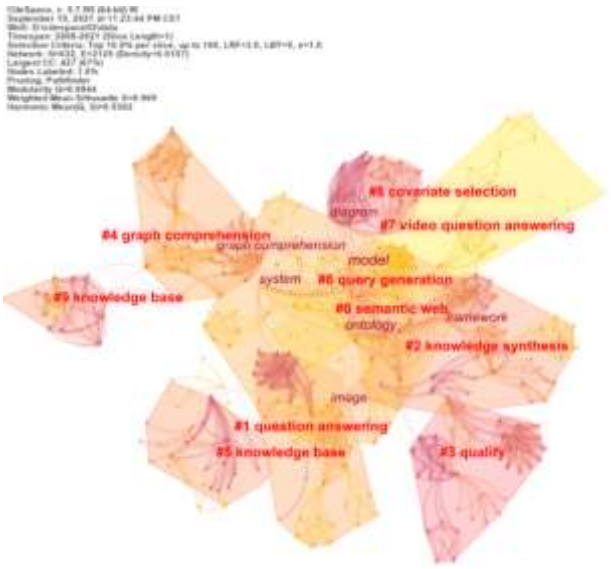

Figure 4: Key words clustering map of foreign literature of knowledge graph question answering

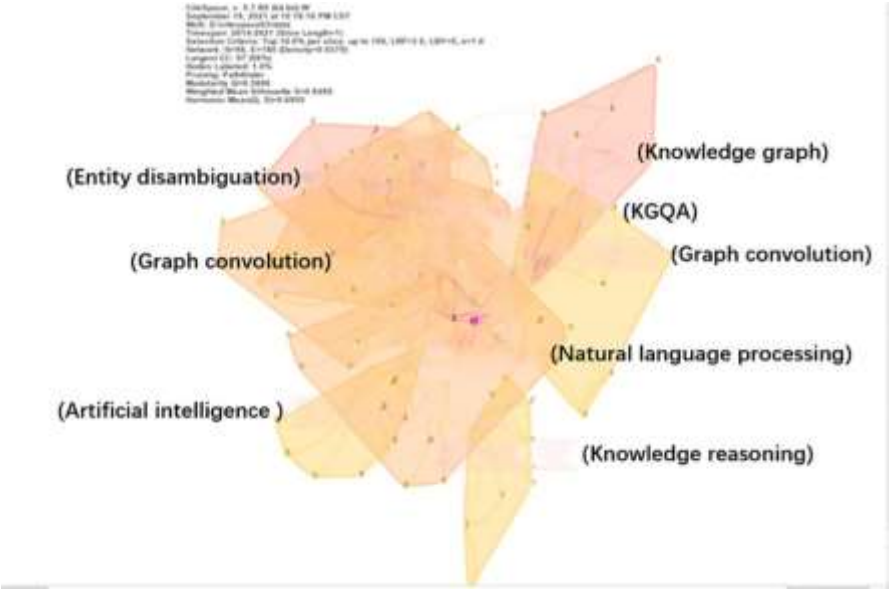

Figure 5: Cluster map of keywords in domestic literature of knowledge graph question answering

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

4.3 Analysis of Research Topics

As can be seen from Figure 4 and Figure 5, foreign and domestic literature keywords are clustered into nine categories and seven categories respectively. Seen from the foreign and domestic keyword clustering results, the top five keywords with LLR values are selected, and Table V and Table VI are obtained. According to the synthesis and summary of keyword clustering results, three research topics of knowledge graph question answering research are obtained: knowledge graph construction, question comprehension and answer reasoning.

Table V. Foreign keyword clustering results

Clustering number and name	Clustering consistency	Key words of the top 5 LLR values
#0 Semantic web	0.886	knowledge graphs、knowledge graph embedding、knowledge graph completion、knowledge extraction、knowledge representation
#1 Question answering	0.979	rdf、ontology、question comprehension、hierarchical recurrent neural network、knowledge-based system
#2 Knowledge synthesis	0.953	network configuration management、training、natural language generation、network security、information graphics
#3 Quality	0.999	cognitive conflict、knowledge re-use、contextual graphs、user interaction、q&a systems
#4 Graph Comprehension	0.975	knowledge graph embedding、lstm、attention model、feature extraction、memory network
#5 Knowledge base	0.987	graph similarity、computational modeling、knowledge base completion、knowledge extraction、knowledge storage
#6 Query generation	0.959	network interventions、multiple networks、semantic query graph、sparql、graph theory

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

#7 Video question answering	0.988	adaptation models、 image segmentation、video captioning、video description generation、multi-interaction
#8 Covariate selection	0.986	directed acyclic graphs、knowledge discovery、semantics、crowdsourcing、link prediction

Table VI. Domestic keyword clustering results

Cluster number and name	Clustering consistency	Key words of the top 5 LLR values
#0 Natural Language Processing	0.98	Knowledge base question answering, deep learning, knowledge graph, map volume, knowledge question answering
#1 knowledge graph Q&A	0.78	Automatic question answering, information retrieval, map database, semantic network, medical question answering
#2 Graph convolution	0.77	Attention mechanism, neural reasoning, language model, neural network, question classification
#3 Entity disambiguation	0.774	Entity link, optimal path selection, predicate matching, feature enhancement, relational link
#4 Knowledge extraction	0.889	Representation learning, named entity recognition, relation extraction, attribute extraction, triples
#5 Artificial Intelligence	0.737	Big data, e-commerce field, medical knowledge graph, intelligent customer service, health and wellness
#6 Knowledge base	0.884	Ontology, self-attention mechanism, problem triad, knowledge organization, knowledge storage
#7 Knowledge Reasoning	0.948	Open domain dialogue, coding-decoding model, rule reasoning, semantic comprehension, ranking learning

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

4.3.1 Construction of knowledge graph

The concept of knowledge graph was first proposed by Google in 2012, and was initially applied to information retrieval and search engine platform development. Generalized knowledge graph is a structured knowledge representation method, while narrow knowledge graph is a knowledge base that represents and stores objective entities (concepts, people, things) and relationships among entities through graph network. Knowledge graph mainly consists of two parts: ontology layer and data layer (reference is based on ontology ……). Ontology layer defines the conceptual structure, relational structure, domain scope, and representation specification in the constructed knowledge graph, which has a certain constraint and normative effect on the data layer. The data layer aims to extract knowledge triples (< entity, relationship, entity > or < entity, attribute, attribute value >) from structured or unstructured multi-source heterogeneous data according to the constraints of the ontology layer, and store and visually display the data through graph databases such as neo4j.

There are three main methods of knowledge graph construction: bottom-up construction, top-down construction and the combination of the two. The bottom-up construction method emphasizes the breadth of knowledge scope, extracts massive triples of information from the Internet through various data collection technologies to fill the data layer, and sums up the main layer of resource pattern construction. It is often used for the construction of large-scale open domain knowledge graphs (such as YAGO, DBPedia, etc.). Top-down construction method emphasizes the depth of knowledge. Professionals and experts first construct the model of ontology layer according to industry rules or industry experience, and then extract high-quality information from multi-source heterogeneous data. This method is often used to construct domain-specific knowledge graphs (such as HowNet).

The construction of knowledge graph generally includes the steps such as data acquisition, knowledge extraction, knowledge fusion and knowledge storage. Data acquisition refers to the process of obtaining structured data from domain knowledge base and ontology base, and obtaining semi-structured and unstructured data from open web pages and websites, which usually requires preprocessing and other operations. Knowledge extraction mainly refers to the operations of named entity identification, relation extraction and attribute extraction on the acquired unstructured data to obtain triple information for construction of "subgraph" of the corresponding knowledge graph. To resolve the conflicts and redundancies between different sources and structural data, knowledge fusion is needed. The process of fusion mainly includes entity disambiguation and entity alignment. Finally, the extracted triple information is stored in a suitable database to facilitate subsequent query and update.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

4.3.2 Question comprehension

Question comprehension is the key step for the question answering system to capture the user's questioning intention, which will directly affect the efficiency of the follow-up processing module of the question answering system. To understand the information needs of users more accurately, it is necessary to transform the natural language problems of users into logical languages and query representations that computers can comprehend. Usually, the research on question comprehension mainly focuses on two aspects: question category comprehension and question content comprehension. Common questions comprehension methods include question classification, question entity identification, question relation extraction, question semantic expansion, and question supplement.

Question classification divides user questions into different categories, which enables the question answering system to choose different ways and mechanisms for obtaining answers according to different question types to improve the efficiency. At present, most scholars use machine learning and deep learning algorithms to train question text classifiers to classify users' questions [40-42]. Question entity recognition and relation extraction mainly realize the accurate positioning of core nouns and relations in user questions. Now, the method of combining circular neural network with conditional random field is often used to extract topic entities and relations [23, 32, 34]. Semantic expansion of questions mainly refers to the semantic supplement of short questions or incomplete questions from different angles, such as synonyms, upper meanings, and lower meanings, to restore the context information omitted from the user's questions, so as to improve the accuracy of question comprehension [43].

4.3.3 Reasoning of answers

Reasoning, as the "ending" step in the knowledge graph question answering, matches the triple information obtained by the question comprehension module with the relational path in the knowledge graph mainly by means of the constructed knowledge graph (knowledge base), so as to retrieve the answer entity or the attribute value corresponding to the answer entity. For a single relational question, that is, a simple question with only one triple information, the extracted question entities can be mapped to the constructed knowledge graph through entity links, and then the extracted text entity relationships are matched with many relationship names of corresponding entities in the knowledge graph, so as to return the fruit entities and related attribute values. For multi-relation questions, that is, questions with multiple triples of information, it is necessary to reason the relation path in the knowledge graph. At present, there are three methods for reasoning multi-relation answers: graph-based search, weak-supervision-based learning and logic-based language query [44].

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

V. CONCLUSIONS AND SUGGESTIONS

Knowledge graph question answering is a natural language processing task that combines knowledge graph (knowledge base) technology with traditional intelligent question answering system. Compared with traditional search engines, it has higher efficiency of question answering feedback. On the one hand, the introduction of knowledge graph technology can provide users with higher quality knowledge content and faster feedback of questions and answers; on the other hand, the question comprehension and answer reasoning module in the question answering system can gain insight into the intentions contained in the user's questioning sentences and return more accurate answers. With the popularization and application of knowledge graph question answering technology in medical question answering, e-commerce customer service, search engine and other scenes, it has gradually become a system with favorable research value and application prospect.

In this paper, based on the data of 662 knowledge graph question answering documents in CNKI and Web of science core database, Excel and CiteSpace were used to compare and analyze the development trends and research hotspots of knowledge graph question answering research at home and abroad, so as to explore the problems and challenges encountered by domestic research in this field. From the perspective of development trend, although the number of papers published by domestic scholars shows a good development trend, the overall number of papers published by individual authors is small. Besides, the researches are not in-depth, mainly staying at the theoretical research of knowledge graph and the application of knowledge graph question answering. Therefore, domestic scholars may strengthen more in-depth research such as algorithm improvement and model optimization in the future. Judging from the distribution of core authors and institutions, there is a lack of inter-agency cooperation and communication among domestic research scholars. Moreover, the researches are scattered, and there are only a small number of influential scholars. At the same time, domestic researches are mainly concentrated in institutions of higher learning, and large-scale research institutes and enterprises have little investment in this field compared with similar institutions abroad. From the perspective of research hotspots, domestic research mainly focuses on hot issues such as knowledge graph construction and question comprehension, while foreign researches mainly focus on answer reasoning and question answering system algorithm model in knowledge graph. Based on the above conclusions, further research work can be carried out around the following aspects.

(1) Construction of limited domain knowledge base

As the source of knowledge and answers of question answering system based on knowledge graph, knowledge base greatly affects the efficiency and accuracy of question answering system. Compared with the traditional open domain knowledge base, the limited domain knowledge base is more suitable to meet the question answering needs of a specific user group because of its small data scale and more

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

concentrated data domain scope, and it provides users with more personalized and accurate information services. It has been used in various application scenarios. The construction of limited domain knowledge base mainly includes steps such as domain data acquisition, domain knowledge extraction, knowledge storage, etc. How to acquire domain data from the vast unstructured corpus and extract knowledge from the data to obtain domain triple information is the key to improve the quality of domain knowledge base construction.

(2) Question classification algorithm

Question classification is the first step in question comprehension, and the accuracy of the classification results will directly affect the efficiency of question processing in question answering system [45]. Question classification mainly optimizes the efficiency of question answering system from two aspects: on the one hand, question classification can reduce steps such as answer reasoning as well as the computation of similarity calculation and improve the speed of answer feedback; on the other hand, the question answering system can adopt different answer inquiry methods according to different question classification results, thus improving the answer reasoning speed. Question classification mainly depends on manual experience to set the corresponding question category labels and select appropriate classification models to classify questions. Then, the setting of classification labels and the selection of classification models are the key factors to determine the accuracy of question classification.

(3) Question feature extraction algorithm

Question feature extraction is one of the key steps in question comprehension, with the aim of improving the effect and efficiency of the answer reasoning. By extracting the question features or question entity features, and calculating the similarity with the domain entity features stored in the constructed knowledge base, the answer closest to the user's question intention can be screened out from the domain knowledge base. How to choose the question features and feature extraction model reasonably is the key to improve the efficiency of answer reasoning.

(4) Question matching and similar question recommendation

When users cannot well describe the problem, the recommendation of similar and related questions is particularly important in the interactive question answering system. Through the comprehension of questions based on context, users can ensure continuous information exchange with the system in natural language; an adaptive model for real environment is established to meet diversified and personalized information needs.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

(5) nswer recommendation/generation

It is one of the basic functions of intelligent question answering system to feed back accurate answer information to users' questions. However, in a specific application scenario, the provision of this very accurate answer information only is not the most desirable way for users to interact. Users are no longer satisfied with obtaining a single answer to a question, but want to get comprehensive knowledge from higher-quality answers. Therefore, it is necessary to reasonably fuse multiple relevant correct answers to the same question. Further, on the premise of accurately perceiving users' emotions, we should make further exploration of users' interests and preferences, build user models, cluster similar users, and provide extended chat topics that users are interested in through collaborative question recommendation, so as to improve users' stickiness and enhance the attraction of applications.

ACKNOWLEDGEMENTS

This research was supported by Anhui Natural Science Foundation (Grant No. 1808085mf194); Philosophy and Social Science Planning project of Anhui Province(Grant No. AHSKF2021D31); Postgraduate Research and Innovation Fund of Anhui University of Finance (Grant No. ACYC2020365).

REFERENCES

[1] (2021) The 47th Statistical Report on Internet Development in China was released. News World, (03): 96.
[2] Wang J, Xu H X (2016) Re-explore the working principle of search engine. Computer Knowledge and Technology, 12 (25): 165-166.
[3] Wang D S, Wang W M, Wang S, et al. (2017) Summary of natural language comprehension methods for domain-specific question answering system. Computer Science, 44 (08): 1-8+41.
[4] O. Etzioni (2011) “Search needs a shake-up,” Nature: International weekly journal of science, vol. 476, no. 7358.
[5] Wang Z Y, Yu Q, Wang N, et al. (2020) Summary of intelligent question answering based on knowledge graph. Computer Engineering and Application, 56 (23): 1-11.
[6] Huang H Q, Yu J, Liao X, et al. (2019) Summary of knowledge graph research. Application of Computer System, 28 (06): 1-12.
[7] Chen Y, Chen C M, Liu Z Y, et al. (2015) The methodological function of Citespace knowledge graph. Science Research, 33 (02): 242-253.
[8] Qiu J P, Su J Y (2008) Bibliometric analysis of competitive intelligence research literature in China. Information Science, 26 (12): 1761-1765.
[9] Wang X H, Ren X F (2018) Bibliometric analysis of tacit knowledge research in China based on

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

CSSCI. Journal of Management, 15 (12): 1854-1861.

[10] Dogrusoz U., Cetintas A., Demir E., and Babur O. (2009) Algorithms for effective querying of compound graph-based pathway databases, Bmc Bioinformatics, vol. 10, Nov.

[11] M. Nikravesh (2008) Concept-based search and questionnaire systems, Soft Computing, vol. 12, no. 3, pp. 301-314, Feb.

[12] T. Berners-Lee, D. Connolly, L. Kagal, Y. Scharf, and J. Hendler (2008) N3Logic: A logical framework for the World Wide Web, Theory And Practice Of Logic Programming, vol. 8, pp. 249-269, May.

[13] Fouquier G., Atif J., and Bloch I. (2012) Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations, Computer Vision And Image Understanding, vol. 116, no. 1, pp. 146-165, Jan.

[14] Chiang M F, Peng W C, and Yu P S (2012) Exploring latent browsing graph for question answering recommendation, World Wide Web-Internet And Web Information Systems, vol. 15, no. 5-6, pp. 603-630, Sep.

[15] Travillian R. S., Diatchka K., Judge T. K., Wilamowska K., and Shapiro L. G., (2011) An ontology-based comparative anatomy information system," Artificial Intelligence In Medicine, vol. 51, no. 1, pp. 1-15, Jan.

[16] Dumontier M., Ferres L., and Villanueva-Rosales N. (2010) "Modeling and querying graphical representations of statistical data," Journal Of Web Semantics, vol. 8, no. 2-3, pp. 241-254, Jul.

[17] Xu Z L, Sheng Y P, He L R, et al. (2016) Overview of knowledge graphping technology. Journal of University of Electronic Science and Technology of China, 45 (04): 589-606.

[18] Dou X Q, Liu T Y, Zhang Z Z (2018) Question answering system based on military knowledge graph. The 6th China Command and Control Conference, 5.

[19] Xu P (2016) Research and implementation of knowledge graph construction method in tourism field. Beijing Institute of Technology.

[20] Ruan T, Sun C L, Wang H F, et al. (2016) Construction and application of TCM knowledge graph. Journal of Medical Informatics, 37 (04): 8-13.

[21] Tang H L (2017) Research on knowledge graph completion algorithm integrating structural and semantic information. Beijing University of Posts and Telecommunications.

[22] Shen C, Huang T L, Liang X (2018) question answering of knowledge graph based on multi-granularity feature representation. Computer and Modernization, (09): 5-10.

[23] Du Z Y, Yan Y, He L (2017) Question answering system in e-commerce field based on Chinese knowledge graph. Computer applications and software, 34 (05): 153-159.

[24] Zhou M (2017) Research and Implementation of Question Answering System Based on knowledge graph. Beijing University of Posts and Telecommunications.

[25] Li Z X (2015) Research on automatic construction of Chinese question answering system knowledge base. Shandong University of Finance and Economics.

[26] Cao Q, Zhao Y M (2015) Technical realization process and related application of knowledge graph.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

Intelligence Theory and Practice, 38 (12): 127-132.
[27] (2018) Development Plan of New Generation Artificial Intelligence. Science and Technology Herald, 36 (17): 113.
[28] Yuan C (2020) Research and implementation of insurance question answering system based on BERT. North University for Nationalities.
[29] Zhang Y, Wang S S, He B, et al. Named entity recognition method of elementary mathematics text based on BERT. Computer application: 1-8.
[30] Wang C T, Ding L K, Yang X X, et al. Named entity identification of Chinese electronic resume based on BERT. Chinese scientific papers: 1-7.
[31] Peng Y, Li X Yi, Hu S J, et al. Three-stage question answering model based on BERT. Computer application: 1-8.
[32] Zhang C T, Chang L, Wang W K, et al. (2020) question answering of fine-grained knowledge graph based on BiLSTM-CRF. Computer Engineering, 46 (02): 41-47.
[33] Luo X, Xia X Y, Ying A, et al. (2021) Chinese clinical entity recognition by combining multi-head self-attention mechanism with BiLSTM-CRF. Journal of Hunan University (Natural Science Edition), 48 (04): 45-55.
[34] Liu Y H, Yang B, Sun Y N, et al. (2019) Natural Language question answering of Marriage Law Based on BiLSTM. Computer Engineering and Design, 40 (04): 1190-1195.
[35] Jiang X, Ma J X, Yuan H (2021) Named entity identification in the field of ecological governance technology based on BiLSTM-IDCNN-CRF model. Computer applications and software, 38 (03): 134-141.
[36] Hu Z, Zhang Y (2016) Analysis of core authors and extended core authors based on Price's law and comprehensive index method–Take Journal of Southwest University for Nationalities (Natural Science Edition) as an example. Journal of Southwest University for Nationalities (Natural Science Edition), 42 (03): 351-354.
[37] Liu Y J, Li R P, Luo Y, et al. (2018) Realization path and innovation mode of artificial intelligence+library knowledge service. Library Science Research, (10): 61-65+42.
[38] Du L, Zuo H M, Li Y S. Analysis of the publishing hotspots and frontiers of domestic smart libraries based on Citespace in recent ten years. Library theory and practice: 1-14.
[39] Yang X H, Sun X B. CiteSpace visual analysis of research progress of pharmaceutical services in China. Chinese Journal of Hospital Pharmacy: 1-8.
[40] Liao K J, Huang Q Y, Xi Y J (2021) Research on knowledge graph construction of online medical community question answering texts. Information Science, 39 (03): 51-59+75.
[41] Cao M Y, Li Q Q, Yang Z H, et al. (2019) Knowledge Question Answering System for Primary Liver Cancer Based on knowledge graph. Journal of chinese information, 33 (06): 88-93.
[42] Liu Y F, Zhang C R (2020) question answering of knowledge base of joint entity identification and relationship prediction. Computer Engineering and Design, 41 (11): 3224-3228.
[43] Meng M M, Zhang K, Lun B, et al. (2019) A semantic query expansion method for knowledge graph

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 967-988
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

question answering. Computer Engineering, 45 (09): 276-283+290.

[44] Zhao X W (2021) Research on Question Answering System Based on Medical knowledge graph. Harbin University of Science and Technology.

[45] Han D F, Turdi Tohti, Eskar Aimudula (2021) Summary of the research on question classification methods in question answering system. Computer Engineering and Application, 57 (06): 10-21.