The Construction of Knowledge Graph of Newspaper Distribution in Yan'an Period and Frontend Visualization

Yifan Du(✉)

Modern College of Northwest University, No. 1, Chenbei Road, Science and Education Park, Luanzhen Street, Chang'an, Xi'an 710100, Shaanxi, China
952173411@qq.com

Abstract. In order to solve the problems of scattered data of CPC newspapers and periodicals in Yan'an period, weak data connection, linear text presentation hindering the use of literature, and low utilization rate of distribution data, this paper proposes a method to construct the knowledge graph of CPC newspapers and periodicals in Yan'an period. This method studies the key technologies such as information extraction, knowledge fusion, knowledge processing, and knowledge graph storage and construction in the process of knowledge graph construction. On the basis of the constructed knowledge graph, the front-end visualization system of the query, analysis and return results of the CPC newspapers and periodicals during the Yan'an period was realized. The constructed knowledge graph includes 15 ontologies, 27349 relationships, 4074 newspapers, and 10616 entities. This study provides new research perspectives and methods for the research of newspaper distribution, and implements a knowledge graph construction and front-end visualization system for newspaper distribution.

Keywords: newspaper distribution · Neo4j knowledge graph · Visualization · Text data extraction · AI model training

1 Introduction

From October 19, 1935, when the Central Committee of the Communist Party of China arrived in Wuqi, Shaanxi Province during the Long March, to March 23, 1948, when they crossed the Yellow River and left Shaanxi Province, the 13 years in between are known as the Yan'an period (1935–1948). The Yan'an period was a time of formation for the ideological and media practices of the Communist Party of China, especially under the leadership of Mao Zedong. It was also when the system for distribution of party newspapers was established as a unique feature of Chinese journalism. During the Yan'an period, the Party faced complex challenges including external Japanese imperialism and internal Kuomintang (Nationalist) reactionary rule. Nevertheless, they managed to build a strong team for newspaper publication which covered the whole country and played a crucial role in leading the people's struggle and achieving great victories that would never be forgotten. While research on the content and editorial practices of newspapers

R. B. B. M. Hussain et al. (Eds.): ICHSSR 2023, ASSEHR 765, pp. 776–789, 2023.
https://doi.org/10.2991/978-2-38476-092-3_95

during the Yan'an period has been comparatively deep and fruitful, much less attention has been paid to the distribution system of Communist Party newspapers during that time [1]. Moreover, most of the research methods have followed the traditional paradigm of revolution history, with few studies that combine the approach of new humanities and "Internet+". With a focus on creating new pathways of research in this area, our study will use knowledge graph technology to conduct a deep and systematic investigation of the newspaper distribution system of the Communist Party during the Yan'an period, based on existing databases.

The history of Communist Party newspapers during the Yan'an period is rich and diverse, with a huge quantity of source materials that include at least nine databases dedicated solely to newspapers. Among them, the earliest established are the "Database of Party History Periodicals" and the "Database of Communist Party Newspapers Archives". Subsequent databases include the "Database of Red Historical Documents" and the "Classical Red Periodicals Database". The most professional and recent database is the "Full-text Database of the Central Committee of the Communist Party of China's Newspapers during the Yan'an Period", although it has not yet been made publicly available [2]. Despite the significant progress in digitization of Communist Party newspapers, current usage of data remains limited. There are problems with data fragmentation, inconsistent standardization of structured data, low organizational efficiency and data redundancy in the database construction. From the perspective of capacity for user information retrieval, most databases only support string queries for newspaper or article names [3]. The underlying cause of this situation is the lack of a compact and efficient knowledge system to organize the data, making it difficult for deep-level data mining and application. To address this problem, our study will introduce knowledge graph technology into the research field of newspaper distribution for the Communist Party during the Yan'an period.

Knowledge graph was first proposed by Google in 2012 with the aim of improving Google's search engine using knowledge graph technology [4]. In the past, searches defined the search content as a string and then performed regular matching between strings, displaying search results from high to low according to the degree of matching. With the use of knowledge graph, the search content is no longer defined as a string, but as an entity. For example, when searching for "Dawn Newspaper," the search engine is not searching for the string "Dawn Newspaper," but for the actual newspaper and providing not only its basic information but also displaying similar information related to it. The basic information of the "Dawn Daily" includes the founding time, category, publication and distribution cycle, and distribution location. Related information includes other newspapers and magazines similar to "Dawn Daily" in the region, as well as personal information about the founder, Peng Xuefeng. Currently, knowledge graphs have become one of the most important ways of expressing knowledge due to their excellent performance in semantic search, intelligent question answering, and recommendation systems, as well as their fundamental position in the field of machine learning. In this study, we introduced knowledge graphs and used information processing technology to extract entities, attributes, and their relationships from textual data, constructing structured knowledge and visualizing it in a graphical form. This made the knowledge more in

line with human habits, and made the key knowledge and its relationships and structures clearer, laying the foundation for further machine learning and semantic computing.

2 Construction and Visualization System of Knowledge Graph for the Publication of Chinese Communist Party's Newspapers and Magazines During the Yan'an Period

The existing database of the CPC newspapers and periodicals has the following problems in terms of construction: First, from the perspective of knowledge sources, the existing database mainly digitizes the catalog, introduction, and text of the Red newspapers and magazines, with little focus on their distribution. Second, from the perspective of database construction, the traditional relational database is used, but the logic of the current state-of-the-art graph database, i.e., the knowledge graph, is lacking. Traditional regular operations are used for knowledge extraction, and the semantic relationships between entities and entities mainly come from manual annotation, without using AI models, making it difficult to extend to large-scale publication data. Third, from the perspective of knowledge visualization, there is currently no newspaper database presented in a visual way.

Based on the above analysis, this study started from the original text of the "Catalog of Historical Newspapers and Magazines of the Chinese Communist Party," took publication as the focus, and proposed and implemented a knowledge graph and visualization system for newspaper and magazine distribution, innovatively combining NLP technology-based [5] entity extraction AI models with the relevant research in the field of newspaper and magazine publication, and querying and visualizing the knowledge of newspaper and magazine distribution from three different dimensions: entities, time, and location. The framework of the Chinese Communist Party newspaper and magazine distribution knowledge graph and visualization system proposed in this study (shown in Fig. 1) includes two parts: knowledge graph construction and front-end visualization.

3 Knowledge Graph of CPC Newspapers and Periodicals Distribution

To address the pain points of the existing Chinese Communist Party newspaper and magazine distribution database: inaccurate search function, a heavy reliance on manual labor, poor scalability, this study proposed the construction of a knowledge graph and database for the distribution of the Chinese Communist Party newspapers and magazines. Unlike the current Chinese Communist Party newspaper and magazine database, which defines search content as strings and then matches them with regular expressions, this database presents search results based on the similarity of entities. This study aims to construct a knowledge graph where search content is defined as entities rather than strings, greatly improving search accuracy. For example, when searching for "Fuxinbao" the search engine will not search for the string "Fuxinbao" but rather for the newspaper "Fuxinbao" In addition to displaying its basic information, information related to and similar to the newspaper will also be shown. The basic information of "Fuxinbao"

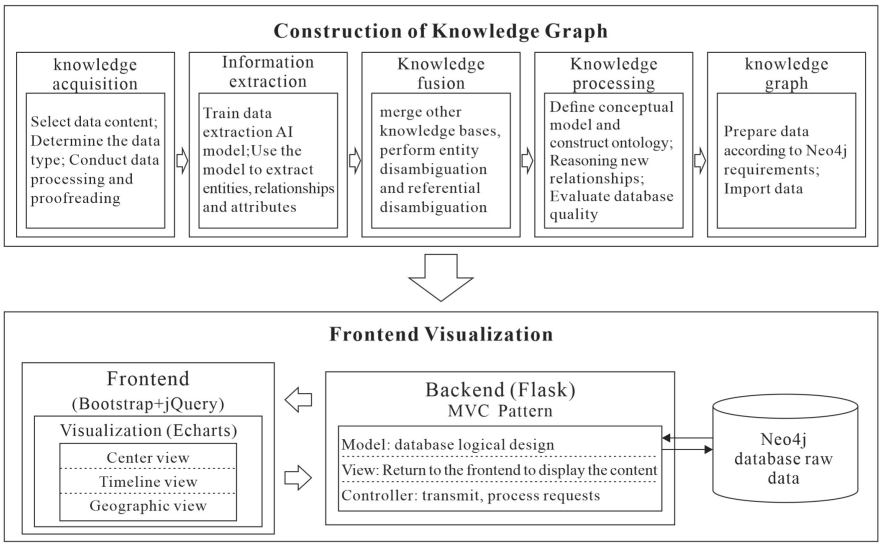

Fig. 1. System framework

includes the founding time, category, publishing cycle, and location, while related information includes other newspapers and magazines in the same region or with a similar founding time, as well as personal information about Fuxinbao's founding person, Peng Xuefeng.

The construction and visualization of the knowledge graph in this study includes five parts: data acquisition, information extraction, knowledge fusion, knowledge processing, and knowledge graph construction and storage. The techniques involved include natural language processing (NLP) and Neo4j graph database construction technology.

3.1 The Hierarchical Structure of the Knowledge Graph of Distribution of Newspapers and Periodicals of the CPC

The hierarchical architecture of the knowledge graph for the distribution of Communist Party newspapers and magazines during the Yan'an period is composed of the data layer and pattern layer [6]. The pattern layer is the conceptual model and logical basis of the knowledge graph, where the nodes represent ontology concepts and the edges represent relationships. The data layer nodes represent specific entities, and edges represent relationships between entities. By analyzing the classification of Communist Party newspapers and magazines during the Yan'an period and their relationships, the core concepts were determined and the tree-shaped ontology structure chart of the data pattern layer was obtained. (As shown in Fig. 2) This chart consists of 15 large categories and several small categories, which have been omitted in the figure due to their large quantity.

The core unit of the knowledge graph is the "entity-relationship-entity" triple. In the pattern layer, this study constructs vocabulary and main concepts related to newspaper distribution, including the newspaper name, period, founding time, shutdown time,

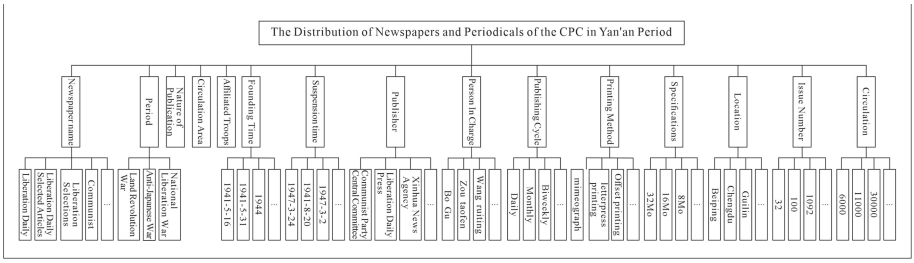

Fig. 2. Structure of Ontology of Knowledge Graph

responsible person, distribution cycle, location, and circulation, among others. Each of the 15 major categories contains several related subcategories. The specific entities, relationships, and triple descriptions are as follows: newspaper name, period, newspaper type, distribution range, affiliation, founding time, shutdown time, publishing body, responsible person, distribution cycle, printing method, specifications, location, issue number, and circulation. The introduction is the attribute of the newspaper name entity and distributes a unique ID as the only identifier.

Relations: mainly centered on the names of newspapers and their relationships with various entities. These relationships include change name, period, type, range, army, start time, close time, distribution subject, responsible person, release cycle, printing mode, specification, location, total number of issues, and circulation.

Triple: The triples composed of entities and relationships include (newspaper name, change name, newspaper name), (newspaper name, period, belonging period), (newspaper name, type, newspaper type), (newspaper name, range, distribution range), (newspaper name, army, affiliated army), (newspaper name, start time, founding time), (newspaper name, close time, suspension time), (newspaper name, distribution subject, publishing subject), (newspaper name, responsible person, person in charge), (newspaper name, release cycle, publishing cycle), (newspaper name, printing mode, printing method), (newspaper name, specification, size), (newspaper name, location, location), (newspaper name, total number of issues, issue number), (newspaper name, circulation, circulation).

3.2 Yan'an Period Chinese Communist Party Newspaper Knowledge Graph Framework Technology

There are two main methods for constructing a knowledge graph: bottom-up and top-down [7]. Although the majority of large-scale knowledge graphs use the bottom-up method, considering the data sources of this study and the fact that it is a research database involving professional personnel in a specific field, the top-down method is adopted.

Data Acquisition

Data acquisition is the first step in building a knowledge graph. The data for this study mainly comes from "A Catalogue of Chinese Communist Party Historical Newspapers and Periodicals (1919–1949)", compiled by Li Yongpu and Lin Zhili, published by Shandong People's Publishing House, and "The History of Chinese Communist

Party Newspaper Distribution" by Wang Xiaolan, published by China Social Sciences Press. "A Catalogue of Chinese Communist Party Historical Newspapers and Periodicals (1919–1949)" is the primary source of data for this study, providing the most important support. The second chapter of "The History of Chinese Communist Party Newspaper Distribution," which covers the distribution of party newspapers during the Land Revolution period, the third chapter on the distribution of CPC newspapers during the Anti-Japanese War, and the fourth chapter on the development of CPC newspaper and periodical distribution during the Liberation War, provide important supplementary data sources as they discuss the secret distribution of party newspapers, the distribution of the "Red Flag Daily," the distribution of "Xinhua Daily," the distribution of CPC newspapers and periodicals in the Kuomintang-controlled areas and the anti-Japanese base areas, the distribution of newspapers and periodicals targeting the Japanese and puppet regimes, the distribution of newspapers and periodicals in Shanghai, Beiping, Hong Kong, and other cities, and the distribution of newspapers and periodicals in the liberated areas. All of the above data are text data and belong to unstructured data.

"A Catalogue of Chinese Communist Party Historical Newspapers and Periodicals (1919–1949)" was published in 1991, with 500,000 words and a print run of only 3,000 copies, and has not been reprinted since. The version used in this study is the facsimile edition, consisting of 454 pages. This book is an introduction to historical materials and not a research book, compiling and recording the history of the distribution of Chinese Communist Party newspapers and periodicals in the form of a bibliographic directory. The book includes 4,505 kinds of Chinese Communist Party newspapers and periodicals. They are classified according to historical periods, with 332 publications during the period of the founding of the Chinese Communist Party and the first revolution, 884 publications during the Land Revolution period, 1,579 publications during the Anti-Japanese War period, and 1,710 publications during the Liberation War period. The entries are arranged in chronological order by the founding time of the newspaper or periodical, and if they were published during the same historical period, they are arranged in order of the publishing organization system. For publications that span two or more historical periods, the founding time is used as the basis for recording. The title of this research is "The Distribution of Chinese Communist Party Newspapers and Periodicals during the Yan'an Period," covering the period from October 19, 1935, when the Central Red Army of the Chinese Communist Party arrived in Wuqi County of northern Shaanxi Province, to March 23, 1948, when the CPC Central Committee left northern Shaanxi Province. This period covers the Land Revolution, the Anti-Japanese War, and the Liberation War, so publications from the first revolution period are not included. The data collected for this research only includes publications that ceased publication after October 19, 1935, and began publication before March 23, 1948. The entries include the name of the newspaper or periodical, the publishing and distribution unit, the issue number, size, printing method, location, responsible person, and founding, cessation, resumption, and revision time. However, due to the special environment during the war, many newspapers and periodicals were omitted or incomplete, so "The History of Chinese Communist Party Newspaper Distribution" is used as an important supplementary source. In the end, data from 4,074 kinds of Chinese Communist Party newspapers and periodicals were collected. In terms of data processing, both "Catalog

of Historical Newspapers and Magazines of the Communist Party of China 1919–1949" (compiled by Li Yongpu and Lin Zhili from Shandong People's Publishing House) and "The History of the Publication of Newspapers and Magazines of the Communist Party of China" (written by Wang Xiaolan from China Social Sciences Press) are photocopy version PDFs. Therefore, the first step is to convert the PDFs into Word documents. Since both books were published early and the quality of the photocopies was poor, OCR technology [8] was used to improve the conversion accuracy. After the conversion, manual proofreading was carried out multiple times.

Information Extraction

The data used in this research is unstructured data, involving 4,074 unstructured newspaper entries in text format. In order to automatically extract information from them and form structured data, this study excluded laborious manual extraction and replaced it with computer iterative learning extraction. With the help of natural language processing technology, structured information was extracted, including entity extraction, relationship extraction, and attribute extraction. According to statistics, up to 10,616 pieces of data need to be extracted for this knowledge graph, which is a huge amount. Manual tagging of each item not only has high cost but also low efficiency. Therefore, it was decided to use EasyDL [9], a deep AI development platform. Based on the ontology, relationships, and attributes that have been pre-defined at the pattern layer, and with the help of NLP development products, an AI model for the study of the publication of Communist Party newspapers during the Yan'an period was trained to enable the machine to automatically extract entities and to serve the study of newspaper publication in different periods. However, the application of AI technology requires a large-scale annotated dataset to train the AI model and improve the efficiency of the training. Therefore, this study used text intelligent tagging to provide only a small number of annotated text data to complete the annotation of the large-scale dataset. The model training for the publication of Communist Party newspapers includes three steps: dataset preparation (tagging), model training, and deployment.

(1) Dataset Preparation
Dataset preparation includes three parts: manual tagging, intelligent tagging, and optimized tagging.

Manual tagging is an indispensable step. This study used the text intelligent tagging [10] function of the EasyDL NLP development platform to upload 500 annotated data samples, which saves the laborious work of manually tagging a large number of samples. Then, text intelligent tagging is turned on, using the 500 small amount of manually annotated data and the pre-trained model ERNIE2.0 [11] to predict the large amount of unannotated data, obtaining smart tagging for the data. At this point, some errors may occur in smart tagging and optimized tagging is necessary. The platform provides up to 300 priority verification samples that are difficult for machines to confirm for manual verification. Finally, after the verification is completed, the system uses this part of the manual verification sample to retrain the teacher model, and eventually obtains more accurate intelligent tagging data.

(2) Model Training
After completing the text intelligent tagging, the subsequent model training can be carried out. The model training is shown in the figure, which takes 32 min with an

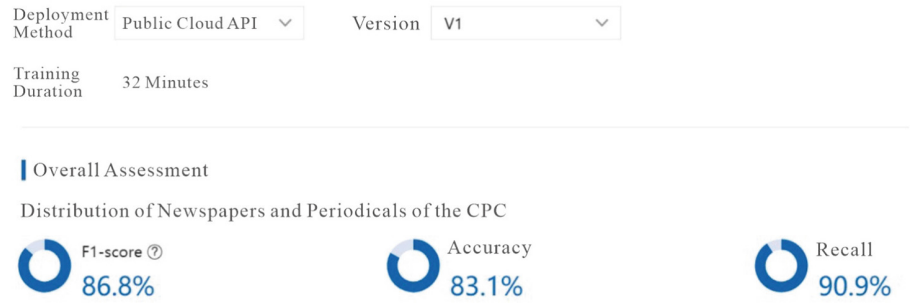

Fig. 3. Evaluation of AI Knowledge Extraction Model

F1-score of 86.8%, accuracy of 83.1%, recall of 90.9%, and excellent performance (as shown in Fig. 3).

(3) Deployment
EasyDL platform provides one-click deployment, which is relatively simple. This model adopts a convenient public cloud deployment to ensure the model's effectiveness, with the selection of high-precision algorithms.

Knowledge Integration

Although the primary data source for this study comes from "Historical Newspaper Catalog of the Communist Party of China 1919–1949", it is not limited to this alone. "The History of Publication and Distribution of Communist Party of China Newspapers and Magazines" is an important supplement to this. To integrate the two data sources into one knowledge base, this study carried out knowledge integration of the two data sources. Knowledge integration first requires entity alignment. Entity alignment is divided into two categories, one is entity disambiguation, and the other is reference disambiguation. During the Yan'an period, there were blockades in various base areas due to the continuous war, and although the number of newspapers and magazines published and distributed was significant, there was a lack of unified management. The phenomenon of repetition of the same name in newspapers and magazines was not uncommon, and the practice of concealed and disguised distribution that concealed the editor and newspaper name made the work of organizing newspapers and magazines even more challenging. To address this, this study conducted meticulous entity disambiguation work, with the rule that in the event of a case with the same name, distinction is carried out based on the date of the founding. If there is a case where the name of a newspaper has changed, differentiation is carried out based on the date of the founding and the publishing entity. This allows for the differentiation of different entities with the same name or the merging of entities with different names, but the same entity. In addition, during the Yan'an period, many newspapers and magazines were threatened with publication and distribution, with many being forced to stop publishing just a few issues after the start of publication. The text frequently includes "discontinued the same year" and other phrases. To address this, reference disambiguation was used to identify words such as "the same year" and "the

same as above", and the entity was determined and replaced based on the principle of proximity.

Knowledge Refinement

After the extraction and fusion of data from the Chinese Communist Party's historical newspaper catalogues between 1919 and 1949, additional refinement is necessary to ensure the quality of the database. Three key components of knowledge refinement include ontology construction, knowledge inference and quality evaluation.

(1) Ontology construction
An ontology describes the conceptual model of terms and their relationships in a knowledge graph. It defines both the terminologies and the hierarchical relationships between these terms. Ontology establishes a basic vocabulary and the rules necessary for defining out-of-vocabulary (OOV) terms. In this study, the field is newspaper distribution, and a top-down approach is utilized for ontology construction. The ontology is pre-determined and includes concepts such as newspaper titles, circulation units, editors, publishers, publishing dates, termination dates, publishing cycles, formats, distribution locations, and circulation quantities.

(2) Knowledge inference
After the ontology is constructed, the prototype of the knowledge graph begins to take shape. However, at this point, most of the relationships between the knowledge graphs are incomplete and require further knowledge discovery through knowledge reasoning. Knowledge reasoning refers to the process of establishing new associations between entities by using existing entity relationship data in the knowledge base and computer inference, thereby expanding the knowledge network. For example, inferring entity relations: If Lu Dingyi is one of the editors of "Frontline Monthly" and one of the chief editors of "Liberation Daily," then through knowledge reasoning, it can be inferred that these two entities' editors overlap. Inferring attribute values: The entity's publishing period can be inferred through the attribute of the publication date of the newspaper. Inferring concepts: Given (Anti-Japanese War battlefield, publishing body, Political Department of the First 139 Division of the Eighth Route Army) and (Political Department of the 139 Division of the Eighth Route Army, affiliation, people's army), one can infer (Anti-Japanese War battlefield, newspaper nature, People's Army Newspaper).

(3) Quality evaluation
Quality evaluation is also a component of knowledge base construction technology. By quantifying the credibility of knowledge and discarding those with low confidence, the quality of the knowledge base is ensured.

Building Knowledge Graph Based on Neo4j

There are two commonly used storage methods for knowledge graphs: RDF (Resource Description Framework) and Neo4j. Considering the emphasis on efficient graph querying and searching in this research, the use of RDF, which stores data in triple format without attribute information, was abandoned in favor of the Neo4j graph database. Neo4j is a platform that stores structured data using nodes and relationships. Its built-in Neo4j Browser visualization tool [12] can display the data in graphical format in real-time

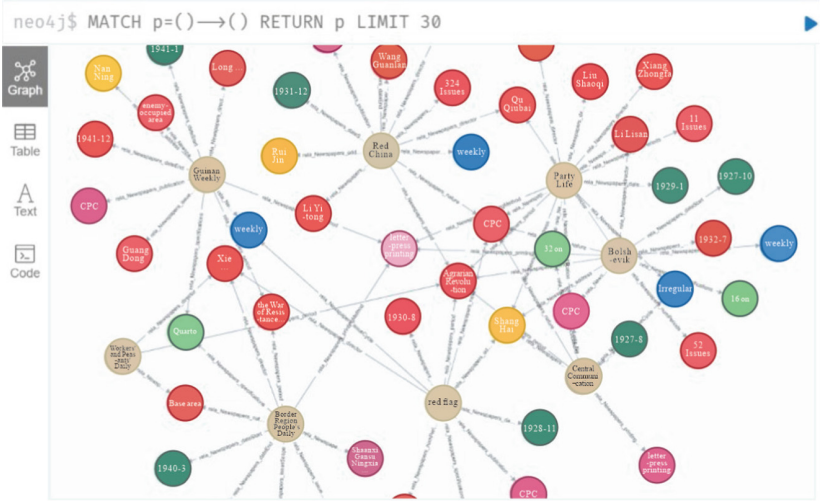

Fig. 4. Screenshot of Neo4j knowledge graph after import.

during development, with various rules and views that allow for analysis from different perspectives, greatly improving analysis efficiency and uncovering hidden structures. Therefore, in the construction of the knowledge graph in this research, the Neo4j platform and Neo4j Browser visualization tool were utilized.

To import data into the Neo4j platform, data must be divided into entities, relationships, and attributes and must be in CSV format with a UTF-8 encoding requirement. This research processed the data according to these requirements and prepared the data. Then, Neo4j was started by inputting "neo4j console" into the command prompt, and the data was imported using the "neo4j-admin import" command in batches. Now the knowledge graph can be visualized in Neo4j Browser (as shown in Fig. 4) with the ability to perform additions, deletions, modifications, and queries. However, at this point, the knowledge graph is only available to developers and cannot be used by users. To make the database usable to users, a frontend and backend need to be set up, data transferred, and ultimately visualized and displayed to users.

4 Knowledge Graph Front-End Visualization

After constructing a knowledge graph in Neo4j, the next step is to visualize it in the front-end. The front-end design concept of this research is centered around the visualization of the database and includes query and analysis functions. Taking into account the needs of research, data volume, and usage scenarios of newspaper distribution, this research chooses to use the Flask framework based on the Python language [13] for design. The front-end design uses the Flask-Bootstrap framework [14] and the jQuery library [15], and uses Echart [16] for data visualization. The back-end uses the Flask framework and develops the back-end based on the MVC pattern. The database used is the constructed knowledge graph of the Communist Party's newspaper distribution

during the Yan'an period in Neo4j. The front-end development of this research uses the Bootstrap framework. Bootstrap is an open-source front-end framework developed by Twitter. Its user interface components can be used to create clean and attractive web pages, and it is compatible with all web browsers. The simplest way to install Bootstrap is through Flask-Bootstrap, so this research implements the installation of Flask-Bootstrap through Flask. Flask is a web development micro-framework based on the Python language.

The front-end design takes into account the characteristics of various entities and research requirements, and designs three presentation schemes: central view, timeline view, and geographic view. Finally, it realizes structured knowledge queries, analysis, and presentation schemes with different entities as the center.

(1) Central view: The central view presents entity-related knowledge centered around entities such as newspapers, publishing entities, responsible persons, and locations. This view can query all entities, and clicking on a relationship type node will display the connected entities in a pop-up window.
(2) Timeline view: The timeline view presents the founding and discontinuation time intervals of the queried journals, as well as the founding of five other newspapers before and after the founding time interval and the discontinuation time interval. It is presented in the form of a timeline (as shown in the figure). This view can only query the name of the journal.
(3) Geographic view: This view can only query the name of the journal, the period it belongs to, and the founding and discontinuation times. If the name of the journal is queried, the geographic view highlights the location of the queried journal. If the period and founding and discontinuation times are queried, the geographic view displays different degrees of highlighting according to the number of newspapers in each region.

The backend design is based on the Flask framework and developed according to the MVC pattern [17]. MVC is a layered design concept consisting of Model-View-Controller layers. The Model layer (M) is used to provide data and is generally responsible for database access and management operations in the project. The View layer (V) is responsible for returning content to the frontend for display. The Controller layer (C) is responsible for forwarding requests and processing them. Control is required to link the Model and View layers. For example, if the backend executed search and visualization commands. The process design logics are as follows (As shown in Fig. 5): the Control layer controls the process, determines which button the user pressed on the main interface webpage, and what values were entered. It then obtains these values and passes them to the Model layer, where the Model layer queries the Neo4j database based on these values. The Model execution query statement returns the results corresponding to the entities in the database based on the input and passes the results to the View layer. In the View layer, the content is displayed line by line with the help of the Echart visualization framework.

The final project is deployed on a server and is presented in the form of a web frontend (as shown in Fig. 6). The platform supports searching by publication name, period, type, distribution range, affiliated troops, founding time, termination time, publishing organization, responsible person, publishing cycle, printing method, size, location, issue

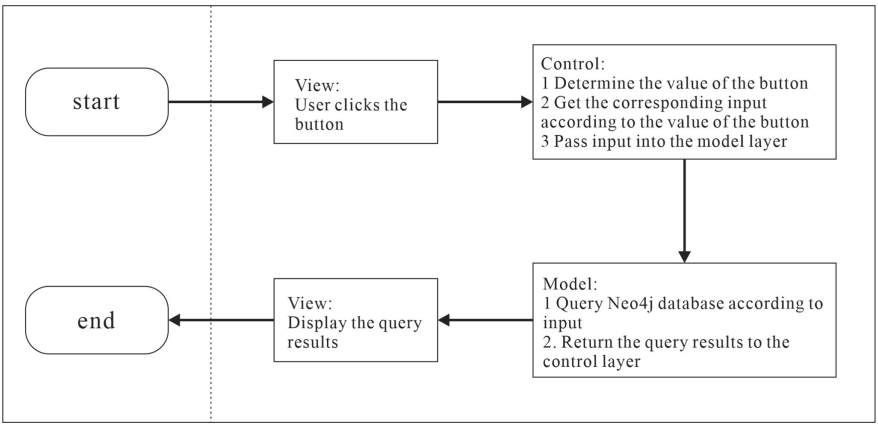

Fig. 5. MVC pattern flow chart

number, and circulation. It presents a complex and interconnected knowledge graph through three types of views: center view, timeline view, and geographical view.

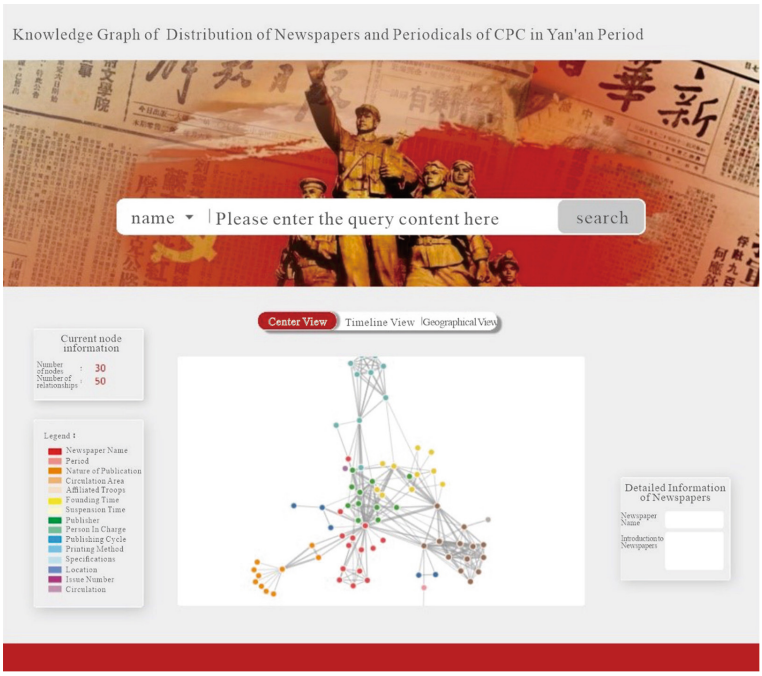

Fig. 6. Distribution of newspaper and periodical web platform.

5 Conclusion

Based on the comprehensive explanation of the development, definition, and general construction process of knowledge graph, this article analyzes the special construction features of the knowledge graph of China Communist Party newspaper distribution during the Yan'an period and proposes a knowledge graph construction method for the newspaper distribution field. This method constructs the ontology of newspaper distribution in advance, constructs and trains information extraction models using AI technology, automatically extracts high-quality entities and relationships in the literature, eliminates contradictions and ambiguities through knowledge fusion, organizes information, and stores the final key knowledge as triples in the Neo4j graph database to form a network-like knowledge graph. In addition, based on the constructed knowledge graph of China Communist Party newspaper distribution during the Yan'an period, this article realizes the front-end visualization query and analysis to assist efficient acquisition of knowledge data in the research of China Communist Party newspaper distribution. The next research work will be to conduct in-depth analysis of the obtained data, mine the intrinsic connections of data in different dimensions, and make breakthroughs.

Acknowledgment. I gratefully acknowledge the financial supports by the general special research project of Shaanxi education department in 2022 under project numbers 22JK0193.

References

1. Wu F. "A century-long history, great achievements, and experience of China Communist Party newspaper distribution". Journal of Henan University (Social Science Edition), 62(05): 136–143+156 (2022).
2. Ren J, Lin W, Li H. "Research on the current situation of digital construction of red literature". Library Journal, (05): 11-15 (2021).
3. Ge Y, Li Q. "The current situation, problems, and countermeasures of the construction of red literature databases: a survey based on 16 red literature data-bases". Library and Information, 40(07): 137–144 (2021).
4. PUJA R A J, MIAO H, GETOO R L, et al. Knowledge graph identification. International Semantic Web Conference. Springer-Verlag New York, Inc. 542–557 (2013).
5. Huang P. "Research on the key technologies of knowledge search for natural language queries". Zhejiang University, (2016).
6. Xu Z, Sheng Y, He L, et al. "Overview of knowledge graph technology". Journal of University of Electronic Science and Technology of China, 45(4): 589–606 (2016).
7. Hu F. "Research on the construction method of Chinese knowledge graph based on multiple data sources". East China University of Science and Technology, (2015).
8. Feng H. "Chinese OCR algorithm and system implementation based on deep learning". University of Chinese Academy of Sciences (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), (2019).
9. Xie Y. "Deep learning development platform and its applications". Artificial Intelligence, (03): 43–53 (2020).
10. Cai Li, Wang Shuting, Liu Junhui, Zhu Yangyong. "A survey of data annotation re-search". Journal of Software, 31(02):302–320 (2020).

11. Li Zhoujun, Fan Yu, Wu Xianjie. “A survey of pre-training techniques for natural language processing”. Computer Science, 47(03):162–173 (2020).
12. Li Xue. “Research and implementation of fuzzy query based on Neo4j graph da-tabase”. Computer Technology and Development, 28(11):16–21 (2018).
13. Ye Feng. “Research on the latest web programming framework Flask based on Python”. Computer Programming Skills and Maintenance, (15):27–28 (2015).
14. Shu Hou, Xiong Yifan, Ge Xuejiao. “Responsive website design and implementa-tion based on Bootstrap framework”. Journal of Beijing Institute of Printing, 24(02):47–52 (2016).
15. Hou Haiping. “Research on improving web user experience based on jQuery front-end framework”. Science, Education and Culture (Bi-Monthly), (02):96–97 (2013).
16. Pei Dandan. “Implementation of data visualization based on ECharts”. Beijing University of Posts and Telecommunications, (2018).
17. Niu Zuodong, Li Handong. “A practical MVC framework based on Python and Flask tools for efficient development”. Computer Applications and Software, 36(07):21–25 (2019).