Research on the Construction of Knowledge Graph of Traditional Medical Intangible Cultural Heritage

Ruihui Huang¹, *, Hang Wei¹, Dongdong Bei¹, Haoyang Liu¹, Zejiong Zhou²

¹ School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, Anhui 233030, China
² School of Economics, Anhui University of Finance and Economics, Bengbu, Anhui 233030, China

*Corresponding author: 932949990@qq.com

Abstract

By constructing a knowledge graph based on traditional medical intangible cultural heritage, comprehensive organization and storage of traditional medical intangible cultural heritage can be achieved, thereby achieving the protection of traditional medical culture. Firstly, Python's web crawler technology was used to obtain relevant data on traditional medical intangible cultural heritage, and it was cleaned and organized. Subsequently, using knowledge graph technology, thematic extraction, fusion, and correlation analysis were performed on the data to construct a knowledge graph centered on traditional medicine. In the process of constructing the knowledge graph, ontology modeling and semantic network technology were used to model and represent traditional pharmaceutical related concepts and relationships, achieving structured storage of the knowledge graph. At the same time, using visualization technology, the knowledge graph is presented in a graphical form, enabling the intuitive presentation of the knowledge and value of traditional medical intangible cultural heritage. Ultimately, a knowledge graph based on traditional medical intangible cultural heritage is obtained, which covers a wealth of traditional medical knowledge and related resources, and can help users gain a more comprehensive understanding and learning of traditional medical culture. By storing and visualizing the knowledge of traditional medical intangible cultural heritage in a structured manner, it provides important support and foundation for the protection and inheritance of traditional medicine.

Keywords

Intangible Cultural Heritage; Knowledge Graph; Traditional Medicine; Python.

1. Introduction

Traditional medicine includes three parts: traditional Chinese medicine, ethnic medicine, and other folk medicine [1]. Traditional medical intangible cultural heritage refers to the traditional medical concepts, knowledge, skills, and practical forms inherited and practiced by various ethnic groups, regions, or communities. It covers a rich range of traditional medical theories, treatment methods, traditional Chinese medicine production techniques, and other aspects [2]. The Chinese government attaches great importance to the protection of traditional Chinese medicine and intangible cultural heritage. In 2005, the National Administration of Traditional Chinese Medicine applied for the "packaging" of traditional Chinese medicine for cultural heritage. Traditional Chinese medicine projects were successfully selected into the first batch of national lists and actively applied for the "Representative List of Human Intangible Cultural Heritage" [3]. In 2006, China established the "China Traditional Medicine Application for World Cultural Heritage Committee" and its office, specifically responsible for the application and
protection of traditional medicine intangible cultural heritage. In the report of the 20th National Congress of the Communist Party of China, it was clearly pointed out that we should "promote the inheritance, innovation and development of traditional Chinese medicine, and promote the construction of a healthy China", which further strengthens our confidence and determination in promoting the high-quality development of traditional Chinese medicine.

However, at present, there are still some problems in the protection of traditional medical intangible cultural heritage. The inheritance of traditional medical knowledge is usually carried out through oral transmission and apprenticeship transmission, with a single transmission method. Influenced by modern lifestyles, fewer and fewer young people choose to engage in the study and inheritance of traditional medicine, which has led to a decrease in the number of inheritors of traditional medicine [4]. At the same time, China's traditional medicine has a long history of development, and some traditional medicine knowledge may not be fully recorded and organized due to being passed down for too long, leading to knowledge gaps and deficiencies, posing the risk of precious traditional knowledge and technology being lost [5]. Secondly, in modern times, with the establishment of the Western medical system and the spread of Western medicine in China, traditional Chinese medicine was greatly impacted. In the process of diagnosis and treatment, unlike Western medicine, which pursues a rigorous axiomatic system, traditional Chinese medicine theory has strong speculative power but low accuracy, which can easily lead to ambiguity and ambiguity. In medical practice, the emphasis on personal experience has led to doubts about the accuracy of traditional Chinese medicine diagnosis and treatment [6].

A knowledge graph is a graph based data structure designed to represent the relationships between entities, and these entities and relationships can be clearly understood through semantics [7]. The construction of a knowledge graph mainly involves ontology knowledge representation, semantic annotation, entity recognition, and relationship extraction. The entities in a knowledge graph typically represent objects in the real world, such as people, places, events, etc. Relationships describe various interactions between entities.

The key advantage of building a knowledge graph is its ability to store complex entity relationships in a highly structured and semantic manner. This structured data representation enables computers to better understand the complexity of the real world, provide more accurate query results, and help people discover hidden patterns and connections. Therefore, knowledge graphs are widely used in many fields, such as recommendation systems, question answering systems, natural language processing, bioinformatics, and so on. In the current era of information technology, traditional medicine faces challenges that make it difficult to survive and develop. However, there is limited research on the use of relevant information technology to protect traditional medicine. This may be due to the large intangible cultural heritage system of traditional medicine, wide distribution of information, and abundant cultural resources. In order to enable the continuation and development of valuable medical knowledge and technology in traditional medicine, this article focuses on constructing a knowledge graph to represent and analyze the complex network of intangible cultural heritage of traditional medicine, in order to deeply understand the internal connections and structures of intangible cultural heritage of traditional medicine. This has certain practical significance for exploring the protection and development of intangible cultural heritage of traditional medicine.

2. Literature Review

2.1. Current Situation of China's Intangible Cultural Heritage

According to academic consensus, the origin of the concept of intangible cultural heritage can be traced back to the Cultural Property Protection Law enacted by Japan in 1950. In 1998, the UNESCO regulations gradually established the concept of intangible cultural heritage through
multiple revisions and changes, and included it in its provisions. In 2011, the Intangible Cultural Heritage Law of the People’s Republic of China clearly defined “intangible cultural heritage” as various traditional cultural expressions passed down from generation to generation by the Chinese nation, as well as places and things related to traditional cultural expressions. These contents include traditional medicine, traditional skills, traditional sports, etc., providing a legal basis for the protection of intangible cultural heritage.

By August 2023, the State Council of China has announced five batches of national intangible cultural heritage lists, including a total of 1557 items (3610 sub items). Among them, traditional medical intangible cultural heritage is an excellent traditional medicine accumulated through time and practical testing, which has been passed down in a live form to this day. Traditional medicine, as a representative of regional culture with a hundred years of history, has become an important component of China’s “intangible cultural heritage” and ranks ninth in the list. As of November 2022, there are a total of 130 representative inheritors of traditional medicine. The selected items can be divided into 23 categories, of which 13 belong to minority medicine, and the remaining 10 categories are: 1 TCM life and disease cognition method, 34 TCM diagnosis methods, 11 TCM processing techniques, 43 TCM traditional preparation methods, 9 acupuncture and moxibustion and moxibustion, 20 TCM bone setting therapies, 1 Tongrentang TCM culture, 1 Hu Qingyutang TCM culture, 5 TCM regimen, and 9 traditional TCM culture.

2.2. Measures for the Protection of Intangible Cultural Heritage

Intangible cultural heritage faces difficulties in inheriting, developing, and innovating in contemporary society. To further solve the problem, many scholars have seized the opportunities brought by technological development. For example, Bai Guixi et al. [7] have constructed a standard system for the inheritance and protection of intangible cultural heritage in China, adhering to the basic principles of systematization, standardization, quantification, and diversification. Teng Chun’e [8] explored the methods of archival protection of intangible cultural heritage and the organization of ontological knowledge of resources, emphasizing the importance of social memory in the protection of intangible cultural heritage. Shen L P [9] proposed an intangible cultural heritage archive management system based on blockchain technology, considering archive management as an important component of intangible cultural heritage protection. Han Fanfan [10] has associated and integrated knowledge of intangible cultural heritage, and constructed a knowledge graph using a graph database, providing methods and technical support for the research and protection of intangible cultural heritage.

2.3. Problems and Challenges Faced by Traditional Medical Intangible Cultural Heritage

Since the 18th National Congress of the Communist Party of China, Xi Jinping has emphasized on multiple occasions the importance of inheriting and protecting traditional Chinese culture, as well as traditional Chinese medicine. However, there is currently little research on traditional medicine. Wang Weijie [11] criticized the scarcity, unclear attribution, and unclear hierarchy of traditional medical intangible cultural heritage, and reconstructed the classification system of traditional medical intangible cultural heritage. Liu Hong and others pointed out that there is an inadequate protection mechanism in the process of inheriting traditional medical intangible cultural heritage. Lack of school education and insufficient utilization of the role of inheritors.

In response to the current suggestions for the protection of traditional Chinese medicine intangible cultural heritage, Wang Fenglan [13] proposed three points: (1) establish a database of traditional Chinese medicine intangible cultural heritage to achieve source protection; (2) To establish a list of intangible cultural heritage protection for traditional medicine and achieve
graded protection; (3) Protecting inheritors and inheritors is fundamental to protecting traditional Chinese medicine.

Liu Zhiwen et al. [14] proposed the use of information technology as an important means of modernizing the theoretical system of Manchu medicine, providing ideas for the protection and inheritance of Manchu medicine. Tong Lin et al. [15] constructed a knowledge graph of the "Shennong Materia Medica Classic", showcasing its rich content in a visual way, providing reference for the research and inheritance of traditional Chinese medicine intangible cultural heritage. Yin Dan et al. [16] designed a knowledge graph of traditional Chinese medicine prescriptions and utilized the knowledge graph to explore and design a search mode for knowledge graphs of prescriptions that simulated human thinking for retrieval. Yu Tong et al. [17], on the other hand, used the Traditional Chinese Medicine Language System (TCMLS) based on ontology and semantic network technology as a framework to initially construct a knowledge graph of traditional Chinese medicine, and displayed its content through visual graphics. They explored applications such as building a traditional Chinese medicine Wikipedia system or implementing a knowledge map system based on the knowledge graph. Jia Lirong et al. [18] constructed a traditional Chinese medicine knowledge graph from the perspectives of data sources, research content, and display forms, and proposed the application prospects of a retrieval system based on traditional Chinese medicine knowledge graphs.

These viewpoints and representative literature provide important perspectives and methods on the protection and research of traditional medical intangible cultural heritage. By establishing a standard system, applying information technology, and constructing a knowledge graph, we can better protect and inherit the intangible cultural heritage of traditional medicine. In response to the current issues of inadequate protection mechanisms and insufficient education, it is still necessary to vigorously promote the sustainable development and innovation of traditional medical intangible cultural heritage in conjunction with the historical background.

3. Knowledge Graph Construction Method

3.1. Data Source

In the process of conducting research on the protection of intangible cultural heritage of traditional medicine, the main data source selected in this article is the China Intangible Cultural Heritage Network, especially the traditional medicine category in the national intangible cultural heritage representative project list of the website. China Intangible Cultural Heritage Network is a national public welfare cultural website hosted by the Ministry of Culture and Tourism of China. It comprehensively, systematically, and accurately displays the rich connotation and unique value of China's intangible cultural heritage, containing extremely rich and detailed data and information.

The data selected in this article includes names, publication times, etc., forming a comprehensive understanding of traditional medicine. These data have been certified by the Ministry of Culture and Tourism, and have high credibility and authority. In addition, these data are publicly available, and the research process in this article has good replicability.

At the same time, this article also utilizes other supplementary data and knowledge sources, such as referring to local chronicles and important academic works in various regions, in order to have a more comprehensive and in-depth understanding of traditional medicine. Although these sources are not as easily accessible and organized as the China Intangible Cultural Heritage Network, they play a crucial role in enhancing the depth and breadth of research.

Overall, the data sources used in this article have both breadth and depth, as well as high credibility and openness, providing a solid foundation for the research in this article.
In complex text mining and natural language processing projects, word cloud maps can also serve as preliminary exploratory data analysis tools to help analysts quickly understand the characteristics of text data. The font size in a word cloud is usually proportional to the frequency of the word. According to the word cloud chart, it can be seen that the highest proportion of project names among all are traditional Chinese medicine, traditional medicine, therapies, preparations, techniques, etc., which can intuitively identify the key content of intangible cultural heritage projects.

3.2. Software Tools

In this study, a series of software tools and methods were used to collect, process data, and construct a knowledge graph. The following are the main tools used in this article and their applications:

(1) Data collection. In order to collect data from China’s intangible cultural heritage online, this article uses web crawling tools such as Requests, JSON, CSV, etc. These Python libraries can help us automate the retrieval of required information from web pages.
Figure 3. Captured partial data screenshot

(2) Data cleaning and preprocessing. This article uses the Pandas library for data cleaning and preprocessing. Pandas is a powerful data processing library that can easily filter, organize, and transform data.

(3) Knowledge graph construction. There are many tools for building a knowledge graph, including Neo4j, RDFLib, Protégé, etc. These tools can construct knowledge graphs based on entities and relationships, while also providing graph query functionality. This article uses Python's networks library, where nodes represent entities and edges represent relationships in the graph. The relationships between entities will be connected using edges.

(4) Visualization of knowledge graph. In order to better understand and display the knowledge graph, this article uses the Gephi visualization tool. This tool can present a knowledge graph in an intuitive manner, helping us better understand and interpret data.

(5) Natural language processing. The data contains a large amount of natural language text. For word frequency statistics and word cloud graph generation, this article uses four tools: Pandas, Counter, and Jieba, and WordCloud. Firstly, preprocess the text, select the text to be analyzed, perform necessary text cleaning, such as removing punctuation marks, special characters, etc., and perform word segmentation on the text, dividing continuous Chinese text into individual words or phrases for analysis. Secondly, the frequency of words is counted to determine the frequency of each word or phrase appearing in the text. Afterwards, choose a font. As not all fonts support Chinese characters, you need to choose a font that contains the required Chinese character set. Finally, use WordCloud tool to generate word cloud maps and word frequency tables.

Table 1. Word frequency of traditional medicine intangible cultural heritage knowledge (word frequency ≥ 6)

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>medicine</td>
<td>118</td>
</tr>
<tr>
<td>traditional</td>
<td>61</td>
</tr>
<tr>
<td>chinese</td>
<td>59</td>
</tr>
<tr>
<td>tibetan</td>
<td>36</td>
</tr>
<tr>
<td>method</td>
<td>32</td>
</tr>
<tr>
<td>tcm</td>
<td>26</td>
</tr>
<tr>
<td>preparation</td>
<td>24</td>
</tr>
<tr>
<td>technique)</td>
<td>23</td>
</tr>
<tr>
<td>diagnosis</td>
<td>23</td>
</tr>
<tr>
<td>therapy</td>
<td>22</td>
</tr>
</tbody>
</table>
3.3. **Knowledge Graph Design**

3.3.1. **Entities and Relationships**

(1) Define entities and relationships. Identifying entities and relationships in data is a key part of knowledge graph design. This usually requires understanding the domain and context of the data. For example, in the data, we identified "inheritor", "project name", and "applying region or unit" as entities, as well as "responsible" and "belonging" as relationships.

![Figure 4. Entity and Relationship (Taking Tibetan Medicine in Qinghai Province as an Example)](image)

(2) Modeling entities and relationships. After determining the entities and relationships, we need to decide how to represent them in the diagram. In our example, we chose the directed graph model, where nodes represent entities and edges represent relationships.

(3) Data import. The next step is to import the data into the graph. This usually involves parsing the data source and adding nodes and edges to the graph.

(4) Data cleaning and processing. In most practical cases, the imported data needs further cleaning and processing, such as removing duplicates, handling missing values, standardizing data, etc.

(5) Visualization of knowledge graph. The final step is the visualization of the knowledge graph, which can help understand and analyze the data. The visualization of the knowledge graph is generated using networkx and matplotlib libraries, and Gephi is used to achieve a clearer and more aesthetically pleasing layout of the knowledge graph.

3.3.2. **Process Design**

(1) Requirement analysis. Build a knowledge graph based on the provided CSV file. According to the requirements, we need to convert the data in the file into a knowledge graph and export the knowledge graph to a Gephi readable format.

(2) Data collection. We collected and organized two CSV files from the "China Intangible Cultural Heritage Network", providing one CSV file containing data on "inheritor", "project name", and "declared region or unit", and one CSV file containing data on "project name" and "declared region or unit".

(3) Data preprocessing. We read the CSV file and viewed its contents. At this stage, if there are duplicates, missing values, or inconsistent formats in the data, we need to perform data cleaning.
(4) Entity recognition and relationship extraction. In the first CSV, we identified three entities ("inheritor", "project name", "reporting region or unit") and two relationships ("responsible" and "belonging"). In the second CSV, we identified two entities ("project name", "reporting region or unit") and one relationship ("belonging to").

(5) Knowledge graph construction. We used the NetworkX library to create a graphical object and add entities and relationships to the diagram.

(6) Visualization of knowledge graph. Gephi was used to draw a knowledge graph for easy understanding and presentation.

(7) Knowledge graph storage. We used Python to export the knowledge graph as a Gephi file, and in Gephi, we modified and beautified the relationships, node colors, and edge colors of the knowledge graph nodes.

3.3.3. Structural Construction

(1) Entity construction. Firstly, three types of entities were identified from the data: inheritors, project names, and declared regions or units. Each entity is considered a node in the graph.

(2) Relationship building. After determining the entities, two types of relationships between entities were established: the relationship between the inheritor and the project, denoted as "responsible"; The relationship between the inheritor and the declared region or unit is denoted as 'belonging'. Each relationship is considered as an edge in the graph, and the direction of the edge represents the direction of the relationship.

(3) Knowledge graph construction. We used the NetworkX library to create a graphical object and then added entities and relationships to the graph. In terms of visualization, the Fruchterman Reingold layout algorithm was used in Gephi. This is a force oriented layout algorithm that simulates physical forces such as gravity and repulsion to layout nodes in the graph. In this layout, nodes are regarded as charged particles, edges are considered springs. By simulating the interactions between these particles, the algorithm attempts to find a layout that keeps all springs as long as possible at their ideal length, thereby minimizing the energy of the entire system.

3.4. Knowledge Storage and Graph Construction

The knowledge graph is stored in a database (SQL Server), where the connections between sources, targets, and colors can be viewed.

3.5. Knowledge Graph Query and Visualization

Figure 5. Query diagram of the relationship between the inheritor project name and the application area or unit (retaining core nodes)
Figure 6. Query diagram of the relationship between the inheritor project name and the application area or unit (non core node)

(1) Node (point): represents an entity. In the graph, there are three types of entities in this article: "inheritor", "project name", and "applying region or unit". Each entity is labeled with a different color. For example, "inheritor" is represented in light purple, "project name" is represented in orange, and "applying region or unit" is represented in light green.

(2) Edge (line): represents the relationship between entities. In this graph, there are two types of relationships in this article: "responsible" and "belonging". For example, the pink edge pointing from "inheritor" to "project name" represents a "responsible" relationship, indicating that this person is responsible for the project. The green edge pointing from "inheritor" to "declared region or unit" represents the "belonging" relationship, indicating that this person belongs to this region or unit.

Using the Fruchterman Reingold layout in Gephi will cause closely related nodes to gather together and sparsely related nodes to stay away from each other. Therefore, some regions in the graph may have high node density, indicating a dense relationship between these entities. On the contrary, the blank areas between nodes represent sparse relationships between these entities.

4. Conclusion

In recent years, the exploration and research of traditional medical intangible cultural heritage projects have received much attention. In order to better document and promote traditional medical culture, this article uses Python technology to construct a framework of knowledge graph, presenting knowledge in a graphical manner, allowing people to more intuitively understand and explore the knowledge system of traditional medical intangible cultural heritage projects. Through visualization forms such as graphs and relationship diagrams, the connections and hierarchical structures between different knowledge can be displayed, helping people quickly understand the knowledge organization methods of traditional medical intangible cultural heritage projects.

As an important component of China's intangible cultural heritage, traditional medical intangible cultural heritage projects are of great significance for the inheritance and development of traditional Chinese medicine culture. This article conducts exploratory research on the inheritance and protection of traditional medical projects in intangible cultural heritage using digital technology, and has made some progress. However, the work done is limited, and many areas need to be improved and improved. The main considerations can be from the following aspects:
(1) The knowledge of traditional Chinese medicine intangible cultural heritage projects has semantic ambiguity and diversity. The knowledge of traditional Chinese medicine intangible cultural heritage projects often has some vague concepts and semantics, such as the name of traditional Chinese medicine and the description of its nature, taste, and meridians. The same concept often has different expressions and interpretations, which brings certain difficulties to the construction and application of knowledge graphs. Therefore, it is necessary to research and develop semantic representation and inference methods suitable for traditional Chinese medicine intangible cultural heritage projects, in order to improve the accuracy and reliability of the knowledge graph.

(2) The knowledge graph construction of traditional medical intangible cultural heritage is not perfect enough. At present, there is a lack of comprehensive and accurate data resources for the construction of traditional medical intangible cultural heritage knowledge maps based on Python. The knowledge of traditional Chinese medicine intangible cultural heritage projects involves multiple aspects such as prescriptions, medicinal herbs, disease diagnosis and treatment, etc. The relevant knowledge of traditional Chinese medicine is missing in the process of inheritance, such as not timely recording in oral transmission, loss or damage of compiled books, which leads to certain difficulties and limitations in collecting and organizing relevant data. Therefore, in future research, researchers need to further organize and explore relevant data resources to enrich and improve the content and quality of the traditional medical intangible cultural heritage knowledge map.

(3) The knowledge of traditional Chinese medicine intangible cultural heritage projects is dynamically changing. The knowledge of traditional Chinese medicine intangible cultural heritage projects not only includes ancient classic literature and traditional experience, but also includes the latest achievements in modern medical research. These knowledge are constantly updating and evolving, and need to be updated in a timely manner into the knowledge graph. The current methods for constructing knowledge graphs often rely on manual annotation and construction, which have problems such as low efficiency and slow updates. Therefore, it is necessary to combine natural language processing and machine learning technologies to achieve automated extraction and updating of knowledge for traditional Chinese medicine intangible cultural heritage projects.

The construction of knowledge maps for traditional Chinese medicine intangible cultural heritage projects still faces many challenges and shortcomings, and further research and exploration are needed. By introducing methods such as cross domain knowledge graph construction, automated information extraction and update, semantic representation and inference technology, the representation and application effect of traditional Chinese medicine intangible cultural heritage project knowledge graph can be improved, and the inheritance and development of traditional Chinese medicine culture can be promoted.

Acknowledgments

This work is supported by 2022 Anhui University of Finance and Economics National College Student Innovation and Entrepreneurship Training Program Project (No.: 202210378039).

References


[10] Han Fanfan. Research on the Construction and Semantic Relationship Discovery of Intangible Cultural Heritage Knowledge Graph Based on Graph Database - Taking Intangible Cultural Heritage along the Beijing Hangzhou Grand Canal as an Example [D]. Baoding: Hebei University, 2021.


