Construction and Research of Pediatric Pulmonary Disease Diagnosis and Treatment Experience Knowledge Graph Based on Professor Wang Lie's Experience

Qingyu Xie and Wei Su[✉]

School of Medical Information, Changchun University of Chinese Medicine,
Changchun 130117, Jilin, China
suwei@ccucm.edu.cn

Abstract. To construct a knowledge graph of Professor Wang Lie, a Master of Traditional Chinese Medicine(TCM), on the diagnostics and treatment of pediatric pulmonary diseases, by providing a foundation for the inheritance of his academic thoughts and clinical experience in TCM. In this study, we focused on Professor Wang's diagnostics and treatment experience in pediatric pulmonary diseases. By utilizing unstructured text data and integrating his academic thoughts and clinical experience, a knowledge graph was built using the Neo4j graph database. Unstructured text experience data meeting the requirements was selected, entered into the Excel® table to establish the Professor Wang's database for the diagnosis and treatment of pediatric pulmonary diseases, and standardized processing of data was performed. In the Neo4j graph database, the schema-level graph comprised 20 entity concept labels, 29 entity nodes, 28 entity relationships, and 22 types of entity relationships. The data-level graph included 20 entity concept labels, 870 entity nodes, 22 types of entity relationships, and 1469 entity relationships. This enabled the visualization of Professor Wang's diagnostics, treatment principles, prescription strategies, and medication patterns for pediatric pulmonary diseases. By constructing a knowledge graph of Professor's diagnostics and treatment for pediatric pulmonary diseases based on the Neo4j graph database, Knowledge extraction, integration, and representation were attained. This work also lays the foundation for optimizing TCM treatment plans for pediatric pulmonary diseases and building intelligent diagnosis and expert systems for TCM-based treatment in the pediatric population.

Keywords: Pediatric pulmonary diseases · Knowledge graph · Neo4j graph database · Professor Wang Lie

1 Introduction

Pediatric lungs are delicate and not fully developed, making children susceptible to various pulmonary diseases when attacked by external pathogenic factors. The lungs function like a canopy and shelter the internal organs and are the first line of defense

against invading pathogens. The lungs govern the respiration, have an opening in the nose, and manifest through body hair. Pathogenic influences can easily invade the lungs without obstruction. External pathogenic factors, known as the "six excesses", primarily affect the lungs and result in, common pulmonary conditions in children, such as cold, cough, bronchitis, recurrent respiratory infections, and asthma. Therefore, pulmonary diseases are very widespread conditions in the field of pediatrics [1]. Professor Wang Lie, a Master of Traditional Chinese Medicine (TCM), has dedicated his entire career to the research of pediatric lung diseases and respiratory disorders. He possesses unique perspectives and a wealth of experience in diagnosing, treating, and prescribing herbal formulas; in addition to utilizing medicinal herbs to treat pediatric pulmonary diseases [2]. Over the years, Professor Wang has accumulated a wealth of clinical data and academic experience in the field of pediatric TCM treatment. In light of the need to inherit and innovate upon the academic thoughts and clinical experience of esteemed TCM practitioners, the introduction of modern information technology becomes crucial to achieve the digitization, intelligence, and transmission of this valuable TCM diagnostic and treatment experience.

The concept of knowledge graph was introduced by Google in 2012. It is a structured representation method that describes concepts, entities, and their relationships in the objective world in a structured form. Knowledge graphs utilize graph structures to model, identify, and infer complex associations between entities. They have found extensive applications in various domains such as intelligent question answering and big data analytics [3–6]. In the research field, knowledge graphs have gained increasing attention among researchers. Yuanyuan Cao [7] constructed an intelligent question answering system based on a knowledge graph for campus network services. The intelligent question answering system realized the convenience and intelligentization of network information services for teachers and students on campus, and improved the service efficiency of campus informatization. Jinlei Liu [8] utilized the Neo4j graph database to build a knowledge graph for coronary heart disease, applying knowledge graph technology to visualize the process of syndrome differentiation and treatment, TCM diagnosis, and medication prescription. This approach allows TCM practitioners to intuitively represent the relationship between diagnosis, treatment processes, and data, thereby providing methods for standardizing and formalizing the diagnosis and treatment of coronary heart disease in TCM.

The exploration and inheritance of the academic thoughts and clinical experience of esteemed TCM practitioners have always been highly regarded in the field of TCM. Professor Wang has accumulated an impressive amount of clinical data and academic experience in the diagnosis and treatment of pediatric pulmonary diseases, particularly by establishing a comprehensive system of principles, methods, formulas, and herbal medicines. In this study, we have utilized Professor Wang's diagnostic and treatment knowledge for pediatric pulmonary diseases as the data source. We designed and constructed the pattern layer and data layer of a knowledge graph, by utilizing the Neo4j graph database to store the knowledge graph. This enables the dynamic visualization display and semantic retrieval functionalities, hence assisting young TCM practitioners to learn Professor Wang's diagnostic and treatment experience for pediatric pulmonary diseases. It supports the inheritance of esteemed TCM practitioners academic thoughts

and clinical knowledge, provides new tools to understand the diagnostic characteristics of renowned TCM practitioners and for better clinical decision-making, and lays the foundation for the deep integration of artificial intelligence and TCM.

2   Materials and Methods

2.1   Data Sources

The data in this study were unstructured text materials that have been publicly published. Relevant literature was electronically retrieved from databases such as Chinese Journal Net, Wanfang Data and CNKI Biological and Medical Literature. Literature and works that could not be obtained through databases were manually retrieved from the library of Changchun University of Chinese Medicine. Literature papers with Professor Wang's clinical experience as the theme were screened, with "Professor Wang Lie" as the keyword or subject retrieval term, and related literature materials were retrieved manually. As a result, 860 papers were included, and repeated or unqualified papers were excluded, 152 papers were reserved.

2.2   Inclusion Criteria

1) Research literature published from January 2003 to May 2023, focusing on the clinical diagnostic and treatment experience of renowned TCM master Professor Wang.
2) Clinical experience literature specifically related to pediatric pulmonary diseases.
3) The clinical experience literature should include comprehensive records, such as, disease name, symptoms, etiology and pathogenesis, disease nature and location, tongue appearance, pulse condition, treatment methods, and prescribed medications.

2.3   Exclusion Criteria

1) Diseases with diagnoses that combine other visceral diseases (pulmonary diseases combine heart and liver diseases)
2) Review articles, conference reports, personal reports.

2.4   Standardized Processing of Data

Referring to the National Prescribed Textbook of Chinese Medical Industry in the "13th Five-Year Plan" Chinese Medicines [9], the Chinese medicine names, flavors, and meridians that appear in the prescriptions were standardized in a unified manner, such as specifying "Fritillaria" as "Bulbus Fritillaria cirrhosa D.Don", "Perilla" as "Perilla frutescens", and "Alkekengi officinarum var. Franchetii (Mast.) R.J.Wang" as "Calyx seu Fructus Physalis".

2.5 Knowledge Extraction

The process of knowledge extraction includes entity extraction, relation extraction and attribute extraction, and knowledge extraction is presented in the form of SPO triples [10]. For example, what is the best way to extract knowledge regarding "Pericarpium Citri Reticulatae treats cough". "Pericarpium Citri Reticulatae" is the name of a Chinese medicine, which can be regarded as an independent entity, and the attribute is Chinese medicine. "Cough" can also be regarded as an entity, and the attribute is a disease. Then the relationship between Pericarpium Citri Reticulatae and cough is the treatment relationship. The representation form of SPO triples is [Pericarpium Citri Reticulatae] <treatment relationship> [cough]. Therefore, in this study, the extraction of entity concepts starts with manual annotation. Relevant entities, relationship attributes, and their relationships were extracted from the clinical experience of Professor Wang in diagnosing and treating pediatric pulmonary diseases. The standards set by National Administration of Traditional Chinese Medicine for terminologies in the TCM industry were used for references in extraction of entity concepts. It combines Professor Wang's academic thoughts and diagnostic-treatment pathways for pediatric pulmonary diseases. After discussions with experts from Professor Wang's research team, a total of 20 entity concept labels were determined, including people, book titles, theories, diseases, contents, symptoms, stages, types, causes, pathogenesis, disease nature, disease location, therapeutic methods, prescriptions, Chinese herbal medicines, effects, alternative names, characteristics, medicinal properties, and meridians. Furthermore, 22 semantic relationship types were identified, namely writing, founding, guiding, content is, symptom is, including, stage is, has, cause is, pathogenesis is, disease nature is, disease location is, therapeutic method is, treats, consists of, effect is, classified as, derived from, alternative name is, characteristic is, medicinal property is, and meridian affiliation is. The relationships between these entity concepts are specified in Table 1.

Table 1. Professor Wang Lie's knowledge graph relationship summary table

Head	Tail	Relation
people	book_title	writes
people	theory	founds
theory	disease	guides
theory	content	content_is
disease	disease	includes
disease	stage	stage_is
disease	type	has
disease	symptom	symptom_is

Table 1. (*continued*)

Head	Tail	Relation
disease	cause	cause_is
disease	pathogenesis	pathogenesis_is
disease	disease_nature	disease_nature_is
disease	disease_location	disease_location_is
disease	therapetic_method	therapeutic_method_is
type	causes	cause_is
type	symptom	symptom_is
prescription	disease	treats
prescription	herbal	consists_of
herbal	type	treats
herbal	disease	treats
herbal	effect	effect_is
herbal	herbal	classifies
herbal	book_title	derived_from
herbal	alternative_name	alternative_name_is
disease	alternative_name	alternative_name_is
herbal	characteristic	characteristic_is
herbal	medicinal_property	medicinal_property _is
herbal	meridian	meridian_affiliation_is

2.6   Knowledge Graph Construction Method

Neo4j is an open-source, high-performance, graph database based on Java. Unlike traditional relational data model graph databases, it stores structured text data in a graph network and visually displays the relationship associations between complex data [11, 12]. The main method is to use nodes and edges to correlate the relationships between complex data. The nodes in each graph network represent entities or concepts, and edges represent the relationships among entities and between entities and concepts [13, 14]. Therefore, this study used the Neo4j graph database to store the knowledge graph. Based on the extracted data, it was converted into a CSV file, The code has been written in Python3.6 Py2neo to import the database to draw the knowledge graph.

3   Results

3.1   Pattern Layer Graph

After referring to the TCM terminology standards established by National Administration of Traditional Chinese Medicine and consulting experts from the Professor Wang's research team, a total of 20 entity concept labels, 29 entity nodes, 28 entity relationships, and 22 entity relationship types were finalized in the schema-level graph construction. Based on the predefined entity concepts and semantic relationships, the knowledge graph schema layer for Professor Wang's expertise in pediatric pulmonary diseases was constructed (Fig. 1).


Fig. 1. Knowledge Graph Model Layer of Professor Wang Lie's Diagnosis and Treatment of Pediatric Pulmonary Diseases

3.2   Data Layer Graph

A total of 1,470 clinical experience and treatment data regarding Professor Wang's expertise in pediatric pulmonary diseases were collected for this study. These data were transformed into a CSV file and imported into the Neo4j graph database using Python. The construction was performed based on the 20 entity concept labels and 22 entity relationship types, resulting in 870 entity nodes and 1,469 entity relationships (Table 2). By executing the Cypher query "MATCH (n:*) RETURN n," the knowledge graph representing Professor Wang's experience in diagnosing and treating pediatric pulmonary diseases could be obtained (Fig. 2).

4   Application of Professor Wang's Knowledge Graph for the Diagnosis and Treatment of Pediatric Pulmonary Diseases

Cypher language is a powerful query language of Neo4j [14]. With the help of the match statement to query-specified nodes or relationships and the where statement to set query conditions, the entities and relationships between entities in diagnostic experience can be queried [15–17], providing references for the utilization of Professor Wang's knowledge resources for the diagnosis and treatment of pediatric pulmonary diseases.
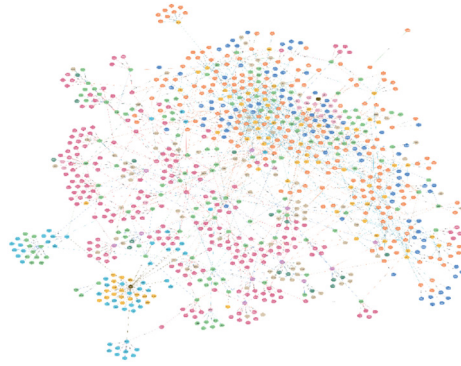
Fig. 2. Professor Wang's Experience Knowledge Graph on the Treatment of Pediatric Pulmonary Diseases

Table 2. Label and Relationship Type Count Table

| Label	Nodes Number	Relationship Type	Count
people	1	writes	17
book_title	18	founds	26
theory	28	guides	12
disease	119	content_is	14
content	14	symptom_is	271
symptom	202	includes	29
stage	11	Stage_is	19
type	47	has	43
cause	35	classifies	4
pathogenesis	16	causeis	35
disease_nature	1	Pathogenesis_is	16
disease_location	4	disease_nature_is	1
therapeutic_method	63	disease_location_is	16
prescription	56	therapeutic_method_is	70
herbal	177	treats	230
effect	97	consists_of	494
alternative_name	18	effect_is	120
characteristic	1	derived_from	1
medicinal_property	11	aiternative_name	18
meridian	10	characteristic_is	1
		medicinal_property_is	13
		meridian_affiliation_is	29

Through the use of Cypher language's MATCH query, it is possible to visualize the semantic relationships [18] within the graph. To provide a clearer and more intuitive representation of the query graph, each entity concept label is represented by nodes with different colors. For example, to understand Professor Wang's insights, treatments, and herbal formulas for treating "Pediatric Asthma in Remission," the following query can be executed: MATCH (n: Disease) where n. name = "Pediatric Asthma in Remission" RETURN n. This query will return a graph, as shown in Fig. 3. From the graph, it can be observed that "Pediatric Asthma in Remission" falls under the category of pediatric asthma. The symptoms associated with this condition include productive cough, shortness of breath, and fatigue. The pathogenesis involves lung-spleen-kidney deficiency and phlegm stagnation, with the affected areas being the lungs, spleen, and kidneys. The treatment approach focuses on clearing heat, purging the lungs, stopping the cough, resolving phlegm, tonifying the spleen, and nourishing the kidneys. Within the context of pediatric asthma in remission, there are two pattern types identified: namely "Pediatric Asthma with Spleen Deficiency and Phlegm Accumulation" and "Pediatric Asthma with Lung Heat Stagnation." According to Professor Wang's academic thoughts and treatment system developed over many years, during the remission stage of an asthma attack, when the pathogenic factors have subsided and the body is in a state of deficiency, qi deficiency and internal dampness may lead to the accumulation of phlegm, resulting in the primary symptom of excessive cough with phlegm. Treatment in this stage primarily focuses on supplementing deficiencies and resolving phlegm. Therefore, for the pattern of lung heat stagnation, the prescription "Xie Fei Fang" (Clearing Heat and Purging the Lungs Formula) is used, which consists of Scutellaria baicalensis Georgi, Calyx seu Fructus Physalis, Chelidonium majus, Fritillaria cirrhosa D.Don, Eriobotrya japonica Thunb, Stemona japonica, Pinellia rhizome praeparatum cum alumine, and Fructus trichosanthis. For the pattern of spleen deficiency with phlegm accumulation, the prescription "Hua Tan Fang" (Resolving Phlegm Formula) is used, which consists of Euryale ferox, Pinellia rhizome praeparatum cum alumine, Exocarpium Citri Rubrum, Wolfiporia cocos, Platycodon grandiflorus, Fritillaria cirrhosa D.Don, Fructus trichosanthis, Radix Glehniae. With the constructed knowledge graph, the corresponding information about the disease etiology, pathogenesis, treatment methods, and prescriptions can be obtained by executing the appropriate query statements include the disease name.

Fig. 3. Knowledge Graph of Asthma Remission Period in Children

By executing the query "MATCH (n:Symptom) where n.name = "cough" RETURN n", a graph with represents the symptom group related to the symptom "cough" can be obtained as shown in Fig. 4. The associated symptoms include cough in relation to pediatric asthma, pediatric bronchitis, pediatric pneumonia, recurrent respiratory infections in children, pediatric nasal congestion, pertussis in children, common cold in children, and pediatric cough. Based on the formed knowledge graph by prior queries, more relationships can be revealed and explored by clicking on any entity label in the graph. For example, clicking on the entity label for pediatric nasal congestion will lead to a visual representation centered around pediatric nasal congestion in the graph. The symptoms manifested in this case include sudden or recurrent sneezing, nasal itching, nasal congestion, clear and watery nasal discharge, coughing, wheezing, eye-related symptoms, mouth-breathing, and associated sleep snoring. Clicking further on the entity label for the remission period of pediatric nasal congestion will provide a visual display of related information such as etiology, pathogenesis, treatment methods, prescriptions, and composition of herbal formulas specific to the remission period of pediatric nasal congestion.
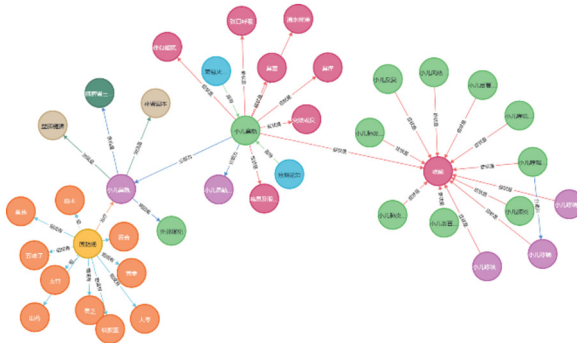
Fig. 4. Knowledge Graph of Cough

Professor Wang is skilled in treatment of various diseases using Chelidonium majus, such as fever, cough, wheezing, and asthma in children. He has developed numerous renowned formulas for treating pediatric lung-related conditions, including Xie Fei Fang (Clearing Lung Formula) and Ping Fei Fang (Balancing Lung Formula), in which Chelidonium majus was commonly used. To understand Professor Wang's perception of Chelidonium majus, the following query "MATCH can be executed: (n:Herb) where n.name = " Chelidonium majus " RETURN n". The results shown in Fig. 5 indicate that Chelidonium majus belongs to the Papaveraceae family and has various folk names, including Picrorhizae rhizoma, Rhizoma Coptidis, Munronia henryi Harms, and Gelsemium elegans. It has a slightly bitter and warm nature and belongs to the lung and spleen meridians. Chelidonium majus is known for its effects in relieving cough and wheezing, alleviating pain, promoting diuresis, killing bacteria, and resolving toxins. Chelidonium majus is included in formulas such as Xie Fei Fang, Ping Fei Fang, Huan Xiao Fang, Xiao Er Xiao Ke Chuan Fang, Bi Xiao Fang, and Xiao Er Bai Bei Zhi Ke Ling. Professor Wang pioneered the use of Chelidonium majus in treating coughs, starting with its application

for pertussis (whooping cough), which yielded good therapeutic effects. Later, Chelidonium majus was also used to treat wheezing. Since Professor Wang first introduced it to pediatric clinical practice, it has been effectively used in the treatment of pediatric pertussis, bronchitis, asthma, pneumonia, and other conditions.

Fig. 5. Chelidonium majus Knowledge Graph

5 Conclusion

Professor Wang is a third-generation national master of TCM and and he has been a doctor of TCM for pediatric disease more than 60 years. He has extensive clinical experience, particularly in the prevention and treatment of pediatric lung-related diseases. Professor Wang has proposed numerous famous theories, including theories on cough, three-stage treatment of asthma, among others, totaling over 10 kinds of theories. Unprecedented in the history of TCM pediatrics are specific TCM medicines for pediatric asthma, such as Xiao Er Xiao Ke Chuan, Xiao Er Fei Re Ping, and Xiao Er Yi Qi Gu Ben Jiao Nang, and et al. The wealth of academic ideas and clinical knowledge of Professor Wang is mostly in the form of unstructured text, comprising scattered, unordered, and independent concepts or viewpoints. Therefore, there is an urgent need to integrate and summarize this knowledge into a structured format, to provide a pathway for inheriting and promoting the academic ideas and clinical experience of esteemed senior TCM practitioners.

This study focuses on utilizing Professor Wang's diagnostic and treatment experience in pediatric lung-related diseases as the data source, integrating his academic ideas and clinical experience into a knowledge graph. It explores the principles, evidence, formulas, and herbs used by Professor Wang in the diagnosis and treatment of pediatric lung-related diseases. This will lay the foundation for optimizing TCM treatment plans for pediatric lung-related diseases and the development of intelligent diagnostic and expert treatment systems in TCM for pediatrics illnesses.

References

1. Bai, N.: Preliminary exploration of the academic thought of children's pharmacology and cough theory instraight deduction of children's pharmacology. Clin. Res. Tradit. Chin. Med. **13**(33), 4–5 (2021)
2. Guo, T., Sun, L., Wang, L.: Application of Wang Lie's thought of prevention and treatment of diseases in children with lung diseases. Chin. Med. Guide **18**(17), 134–6+45 (2021)
3. Yu, T., Li, J., Yu, Q., et al.: Construction and application of traditional Chinese medicine health knowledge graph. China Digit. Med. **12**(12), 64–6 (2017)
4. Deng, Y., Zhou, W., Zhang, Z., et al.: Construction of knowledge graph based on medical records of famous old Chinese physicians. J. Hunan Tradit. Chin. Med. **35**(07), 186–187 (2019)
5. Yang, Y., Li, Y., Zhong, X.: A link prediction model for traditional Chinese medicine case knowledge graph. Chin. Med. Inf. **39**(03), 1–6+15 (2022)
6. Niu, Z., Zhang, X., Xu, W., et al.: Research and practice of Chinese medical knowledge learning based on knowledge graph. Comput. Program. Skills Maint. (02), 20–23 (2023)
7. Cao, Y., Su, W.: Research on knowledge graph construction and Q&A application of campus network operation and maintenance service based on Neo4j. Adv. Educ. Humanit. Soc. Sci. Res. **1**(3), 360–366 (2023)
8. Liu, J., Hui, X., Zhang, Z., et al.: Construction of coronary heart disease knowledge graph based on TCM diagnosis and treatment guidelines. Chin. J. Exp. Tradit. Med. Formulae **29**(07), 208–215 (2023)
9. Zhao, Y., Li, Y., Zhang, J.: Construction of traditional Chinese medicine knowledge graph for influenza. Chin. J. Med. Libr. Inf. Sci. **30**(05), 24–30 (2021)
10. Zhong, G.: Chinese Materia Medica, 4th edn. China Press of Traditional Chinese Medicine, Beijing (2016)
11. Cao, H., Xu, J., Dou, F.: Research progress of graph neural network in knowledge graph construction and application. Comput. Era **6**, 35–38 (2020)
12. Wu, D., Zhou, Z., Shang, H.: Analysis of research hotspots and trends of TCM health examination based on knowledge graph. Comput. Era (02), 59–61+5 (2022)
13. Chen, S., Xia, S., Deng, W., et al.: Research on coronary heart disease TCM syndrome differentiation and treatment knowledge graph based on Neo4j. China Med. Guide **18**(21), 138–141 (2021)
14. Tong, Y., Chen, L.: Analysis of hotspots and trends in TCM health examination based on knowledge graph. Chin. Med. Manag. Mag. **31**(08), 194–196 (2023)
15. Li, L., Li, Y., Xu, H., et al.: Research and practice of building intelligent Q&A APP for traditional chinese medicine health preservation based on knowledge graph. J. Med. Inf. **43**(07), 50–54 (2022)
16. Xiao, F., Zhang, S., Hu, Z.: Research on disease prevention and control knowledge graph based on Neo4j. Electron. Technol. Softw. Eng. **22**, 180–182 (2021)
17. Yin, D., Zhou, L., Zhou, Y., et al.: Research on "graph search mode" design of traditional Chinese medicine prescription knowledge graph. Chin. J. Tradit. Chin. Med. Inf. **26**(08), 94–98 (2019)
18. Yang, Y., Li, Y., Shuai, Y., et al.: Construction of knowledge graph based on traditional Chinese medicine case records. J. Med. Inf. **43**(10), 50–54 (2022)