



## Construction and application of knowledge graph of *Treatise on Febrile Diseases*

LIU Dongbo, WEI Changfa, XIA Shuaishuai, YAN Junfeng\*

School of Informatics, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

### ARTICLE INFO

#### Article history

Received 06 September 2022

Accepted 27 November 2022

Available online 25 December 2022

#### Keywords

*Treatise on Febrile Diseases* (Shang Han Lun, 《伤寒论》)

Knowledge graph

Ontology

Graph database

Knowledge extraction

Knowledge fusion

### ABSTRACT

**Objective** To establish the knowledge graph of “disease-syndrome-symptom-method-formula” in *Treatise on Febrile Diseases* (Shang Han Lun, 《伤寒论》) for reducing the fuzziness and uncertainty of data, and for laying a foundation for later knowledge reasoning and its application.

**Methods** Under the guidance of experts in the classical formula of traditional Chinese medicine (TCM), the method of “top-down as the main, bottom-up as the auxiliary” was adopted to carry out knowledge extraction, knowledge fusion, and knowledge storage from the five aspects of the disease, syndrome, symptom, method, and formula for the original text of *Treatise on Febrile Diseases*, and so the knowledge graph of *Treatise on Febrile Diseases* was constructed. On this basis, the knowledge structure query and the knowledge relevance query were realized in a visual manner.

**Results** The knowledge graph of “disease-syndrome-symptom-method-formula” in the *Treatise on Febrile Diseases* was constructed, containing 6 469 entities and 10 911 relational triples, on which the query of entities and their relationships can be carried out and the query result can be visualized.

**Conclusion** The knowledge graph of *Treatise on Febrile Diseases* systematically realizes its digitization of the knowledge system, and improves the completeness and accuracy of the knowledge representation, and the connection between “disease-syndrome-symptom-treatment-formula”, which is conducive to the sharing and reuse of knowledge can be obtained in a clear and efficient way.

## 1 Introduction

*Treatise on Febrile Diseases* (Shang Han Lun, 《伤寒论》) is the first classic of traditional Chinese medicine (TCM) with complete theories, methods, formulas, and medicines in Chinese history. However, due to the different annotations of doctors in previous dynasties, the difficulty in understanding the semantics of ancient books and in mastering the theory of syndrome differentiation and other issues, the knowledge system of *Treatise on Febrile Diseases* has encountered resistance in terms of knowledge

representation, sharing, and reuse, which is not conducive to the inheritance and promotion of *Treatise on Febrile Diseases* in the contemporary era.

With the explosive growth of knowledge and the continuous development of information science, the concept of knowledge graph is introduced into the field of knowledge engineering to better describe the relationship between knowledge and realize the storage and sharing of knowledge. A knowledge graph (KG) is a knowledge network that describes objectively existing entities and their relationships in the real world, in which entity

\*Corresponding author: YAN Junfeng, Professor, E-mail: [junfengyan@hnuucm.edu.cn](mailto:junfengyan@hnuucm.edu.cn).

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmcd.2022.12.006

Citation: LIU DB, WEI CF, XIA SS, et al. Construction and application of knowledge graph of *Treatise on Febrile Diseases*. Digital Chinese Medicine, 2022, 5(4): 394-405.

Copyright © 2022 The Authors. Production and hosting by Elsevier B.V. This is an open access article under the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

characteristics are represented in the form of an “Attribute-value Pair (AVP)”, and the relationships between entities are represented by edges in the network [1]. It can form a common understanding of domain knowledge, and realize knowledge representation, sharing, and reuse. In recent years, some scholars have adopted different methods and focused on making some research achievements in the fields of TCM clinical knowledge graph, TCM case records knowledge graph, prescription knowledge graph, and construction methodology of TCM knowledge graph [2, 3]. However, these knowledge graphs still have much room for improvement in scale, standardization, formalization, and systematicness, etc. [4]. Like the existing knowledge graphs related to *Treatise on Febrile Diseases* only involve the graph construction of a certain meridian disease or decoction [5, 6], or the completeness and accuracy of knowledge, and its links need to be improved [7]. In addition, although the knowledge system structure of *Treatise on Febrile Diseases* has been sorted out clearly, the natural language processing technology is still unable to effectively handle the original *Treatise on Febrile Diseases* in ancient Chinese and the annotations of medical experts in previous dynasties, and to effectively extract and integrate knowledge.

This study adopts the construction method of “from top to bottom as the main, and from bottom to top as the auxiliary” to sort out the concept classification and its relationship of *Treatise on Febrile Diseases* from the perspective of information processing in five aspects: disease, syndrome, symptom, method, and formula. Under the guidance of experts in TCM, the concept, entity, and relationship of these five aspects were extracted by combining manual and machine methods to complete the construction of the knowledge graph of *Treatise on Febrile Diseases*. On this basis, a visual query of domain knowledge is realized to systematically and comprehensively represent and store the knowledge system of *Treatise on Febrile Diseases*, and improve the completeness and accuracy of knowledge representation.

## 2 Construction methods and results

### 2.1 Overall plan of construction

The knowledge graph can be logically divided into a data layer and a pattern layer. In the data layer, knowledge is stored in the graph data structure in the unit of fact. There are mainly two storage methods of the knowledge: Resource Description Framework (RDF) storage and graph database [8]. A graph database is widely used due to its efficient ability to query associated data. The pattern layer is located on the data layer and stores refined knowledge, in which ontology base is usually used to standardize the expression of facts in the data layer.

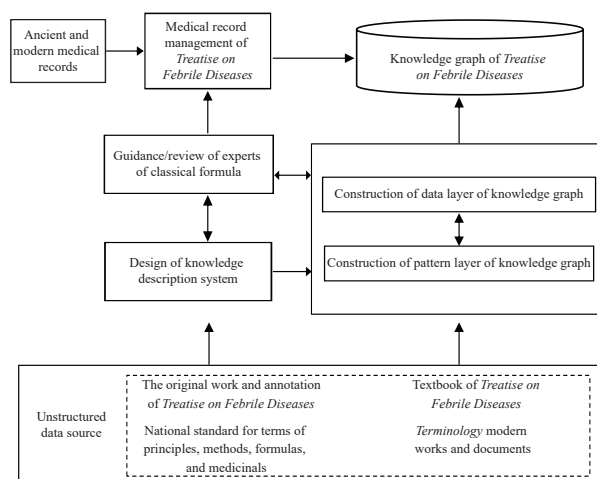
The knowledge graph is mainly constructed in two ways: top-down and bottom-up. The top-down

construction method refers to the building of the corresponding ontology and data pattern first, completing the construction of the pattern layer, and adding entities to the knowledge base to complete the construction of the data layer. The bottom-up construction method refers to the extracting of entities from the open link data, selecting entities with high confidence to join the knowledge base, and constructing the upper ontology pattern [9].

This paper constructed the knowledge graph of *Treatise on Febrile Diseases* via a “from top to bottom and from bottom to top” approach. With the *Annotations of Treatise on Febrile Diseases* (Shang Han Lun Jiao Zhu, 《伤寒论校注》) (LIU Duzhou, published by the People’s Medical Publishing House) [10], the *Selected Readings of Treatise on Febrile Diseases* (Shang Han Lun Xuan Du, 《伤寒论选读》) (The Fourth Edition in the New Century, textbook of the 13th Five Year Plan for Higher Education in the National Traditional Chinese Medicine Industry) [11] as the main data source, under the guidance of experts of classical formula, the concepts, entities, and their interrelationships of five aspects, including disease, syndrome, disease, method, and formula, were extracted by combining manual and machine methods, and then the knowledge graph of *Treatise on Febrile Diseases* was constructed. In addition, the case records of six meridians diseases and syndromes of ancient and modern classics, and the later knowledge reasoning results (which need to be reviewed by experts of classical formula) were added to the existing graph to update the knowledge graph. The overall scheme is shown in Figure 1.

### 2.2 Pattern layer construction

This paper constructed the domain ontology of *Treatise on Febrile Diseases* to realize the construction of its knowledge graph pattern layer. Ontology is a clear formal specification of a shared conceptual model [12]. Ontology



**Figure 1** Overall plan for the construction of the knowledge graph of *Treatise on Febrile Diseases*

construction includes concepts (semantic types) and semantic relations.

With taking the national standards *Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine* (revised in 2020), *Clinic Terminology of Traditional Chinese Medical Diagnosis and Treatment* (revised in 2020), and the Semantic Network Framework of Traditional Chinese Medicine Language System (ISO/TS 17938-2014) [13] as references, this paper constructed a conceptual hierarchy tree according to the superior-subordinate relationship (Is\_a relationship), obtaining 25 representative concept nodes and 27 semantic relationships. The main constraint rules between concepts and semantic relations are shown in Table 1.

### 2.3 Data layer construction

**2.3.1 Data preprocessing** To construct the data layer of the knowledge graph of *Treatise on Febrile Diseases*, knowledge extraction was based on the semantic types and semantic relationships established in the knowledge graph pattern layer *Treatise on Febrile Diseases*. Data sources mainly included the middle 10 chapters of *Treatise on Febrile Diseases* (Chapters 5 – 14, a total of 398 original articles, mainly discussing six meridians diseases and syndromes and its changes, which are the key learning and research part of physicians in previous dynasties, and also the source of the curriculum content of *Treatise on Febrile Diseases* in contemporary TCM colleges), ISO technical specifications, Chinese national standards, contemporary classic annotations of *Treatise on Febrile Diseases*, and professional textbooks. Because different data sources use different names for the same term, it is necessary to standardize the names of terms such as disease, syndrome, symptom, therapeutic methods, therapy, and formula for entities obtained through knowledge extraction to achieve knowledge fusion.

(i) Knowledge extraction of the original text of *Treatise on Febrile Diseases*

In the clinical diagnosis and treatment of TCM, not only comprehensive analysis of pulse and symptoms is needed, but also the original syndrome and the treatment history are important bases for correct diagnosis. Therefore, for the 10 chapters in the middle of *Treatise on Febrile Diseases* that involve disease differentiation and syndrome differentiation, this paper extracted knowledge from the original syndrome, treatment history, current syndrome, etiology and pathogenesis, results of syndrome differentiation (or disease differentiation), and treatment plans. Meanwhile, if more than one diagnostic conclusion were involved in a provision, the original text needed to be split into multiple scriptures, so that only one diagnostic conclusion was included in the split scripture.

(a) The indivisibility of the original *Treatise on Febrile Diseases*

Take Line 12 in the original *Treatise on Febrile Diseases* as an example, “Greater Yang wind stroke syndrome is characterized by floating pulse and weak ulnar pulse. Those with a floating pulse will have a fever, while those with a weak ulnar pulse will sweat. The patient is stingy but aversive to cold, with intermittent bad wind, fever as if covered by fur, accompanied by snorting and retching, which can be treated with Guizhi Tang (桂枝汤)”. The original text has only one diagnosis result, so it does not need to be split. This paper will extract knowledge from the original line in the manner of Table 2.

It should be noted that each knowledge content cell in the knowledge table has a specific semantic type (unless the cell content is “none”). For example, the knowledge content corresponding to the original syndrome and the current syndrome is composed of entities of disease, syndrome, and symptom. To facilitate the automation of subsequent entity and relationship extraction, each term in the knowledge type such as original syndrome, treatment history, current syndrome, and treatment scheme is also included in the knowledge table, indicating its semantic type.

The greater Yang wind-invasion syndrome (太阳中风证) in the above case represents a concept of syndrome, including multiple entities (i.e. specific syndromes). To distinguish these syndromes, and provide a definite syndrome differentiation result for each medical case, this paper follows the provisions of the textbook *Selected Readings on Febrile Diseases* (The Fourth Edition in The New Century), and names these specific syndromes according to the decoction syndrome, that is, provides the syndrome differentiation result in the way of “syndrome concept + formula syndrome” (for example, the greater Yang wind-invasion syndrome + Guizhi Tang syndrome).

As for the regulation of “scheme code”, words such as “govern” “should” “then give” “may give” “cannot be given” and “cannot” appear when a formula is used in the original text of *Treatise on Febrile Diseases*, which represent four different levels of meaning (the first three meanings represent availability, the degree decreases gradually, and the fourth meaning means prohibition). Therefore, “scheme code” is used in the knowledge type in Table 2 to represent the meanings of these four different levels. Therefore, when the original text appears “××× Decoction governs”, the value of this field is 1. If it is “should”, the field value is 2. If it is “then give” “may give”, etc., the value of this field is 3. If it is “cannot be given” or “cannot”, the value of this field is 4.

b) The original text of *Treatise on Febrile Diseases* can be split

Take Line 15 in the original text of *Treatise on Febrile Diseases* as an example, “When in greater Yang disease, after precipitation, Qi surges upward, [one] may give

**Table 1** Restriction rules of the knowledge system of *Treatise on Febrile Diseases*

No.	Head concept	Semantic relation	Tail concept
1	Disease	Complicates	Disease
2	Disease	Transmission and transformation	Disease
3	Syndrome	Co-occurs with	Disease
4	Symptom	Manifestation of	Disease
5	Symptom	Primary symptom	Disease
6	Symptom	Accompanied symptom	Disease
7	Symptom	Pulse symptom	Disease
8	Therapeutic methods	Treats	Disease
9	Formula	Treats	Disease
10	Symptom	Manifestation of	Syndrome
11	Symptom	Primary symptom	Syndrome
12	Symptom	Accompanied symptom	Syndrome
13	Symptom	Pulse symptom	Syndrome
14	Symptom	Necessary symptom	Syndrome
15	Etiology and pathogenesis	Causes	Syndrome
16	Therapeutic methods	Treats	Syndrome
17	Therapeutic methods	Mistreats	Syndrome
18	Syndrome	Forbids	Therapeutic methods
19	Treatment	Treats	Syndrome
20	Formula	Treats	Syndrome
21	Formula	Mistreats	Syndrome
22	Syndrome	Forbids	Formula
23	Symptom	Co-occurs with	Symptom
24	Symptom	Precedes	Symptom
25	Symptom	Meaning expansion	Symptom
26	Therapeutic methods	Uses	Formula
27	Formula	Consists of	Herbs
28	Formula	Affects	Acupuncture point
29	Formula	Affects	Meridian and collateral
30	Formula	Affects	Viscera
31	Formula	Affects	Basic substance in the human body
32	Food	Treats	Syndrome
33	Viscera	Inter_exterior and interior with	Viscera
34	Six meridians	Inter_exterior and interior with	Six meridians
35	Person	Location of	Geographic area
36	Examination	Contains	Content
37	Content	Manifestation of	Syndrome
38	Disease, syndrome, symptom	Original disease and syndrome	TCM case
39	Therapeutic methods, treatment, formula	Treatment history	TCM case
40	Disease, syndrome, symptom	Current disease and syndrome	TCM case
41	Syndrome	Syndrome identification result	TCM case
42	Therapeutic methods, treatment, formula	Treatment plan	TCM case
43	Concept	Is a	Concept
44	Concept	Equivalent to	Concept

**Table 2** Example 1 of knowledge extraction from the original text of *Treatise on Febrile Diseases*

No.	Knowledge type	Knowledge content
1	Original disease and syndrome	None
2	Treatment history	None
3	Current disease and syndrome	Floating and slow pulse (symptom of pulse), aversion to cold (main symptom), aversion to wind (main symptom), fever (main symptom), rhinorrhea (accompanied symptom), retching (accompanied symptom)
4	Etiology and pathogenesis	Wind-cold attacking exterior, insecurity of defense Yang, outward discharge of nutrient-Yin, construction-defense disharmony
5	Result of syndrome differentiation	Greater Yang wind-invasion syndrome + Guizhi Tang syndrome (太阳中风证 + 桂枝汤证)
6	Treatment plan	Sweating method (therapeutic methods), medical treatment (therapy), Guizhi Tang (formula)
7	Scheme code	1

Guizhi Tang, according to the previously mentioned method. If [there is] no upsurge, [one] cannot give this formula". Since the original line has different treatment methods under two kinds of syndromes, first divide the original line into two scriptures numbered "15-1" and "15-2", and then extract knowledge in the way of Table 3 and Table 4.

The scripture numbered "15-1" reads: "When in greater Yang disease, after precipitation, Qi surges upward, [one] may give Guizhi Tang". The method of knowledge extraction is shown in Table 3.

The scripture numbered "15-2" reads: "When in Greater Yang disease, after precipitation, if [there is] no

upsurge, [one] cannot give Guizhi Tang". The method of knowledge extraction is shown in Table 4.

Since the scheme code in Table 4 is "4", it means that the treatment scheme in this table is prohibited for the current disease and syndrome. Therefore, based on the original content, the syndrome differentiation results in this table only need to specify the concept of syndrome. It is not necessary to give the syndrome differentiation results in the way of the "syndrome concept + formula syndrome" as in the case of other scheme codes.

(ii) Terminology specification

Concerning the standardization of disease names and syndrome names, 15 disease names and 299 syndrome

**Table 3** Example 2 of knowledge extraction from the original text of *Treatise on Febrile Diseases*

No.	Knowledge type	Knowledge content
1	Original disease and syndrome	Greater Yang disease (disease)
2	Treatment history	Purgative method (therapeutic methods + mistreats)
3	Current disease and syndrome	Consciously feel Qi ascending counterflow (primary symptom)
4	Etiology and pathogenesis	Misuse of the purgative method shows that the exterior pathogen is not resolved and the internal Qi is injured, but the positive Qi is sufficient
5	Result of syndrome differentiation	Greater Yang wind-invasion syndrome + Guizhi Tang syndrome
6	Treatment plan	Sweating method (therapeutic methods), Guizhi Tang (formula)
7	Scheme code	3

**Table 4** Example 3 of knowledge extraction from the original text of *Treatise on Febrile Diseases*

No.	Knowledge type	Knowledge content
1	Original disease and syndrome	Greater Yang disease (disease)
2	Treatment history	Purgative method (therapeutic methods + mistreats)
3	Current disease and syndrome	No sense of feeling Qi ascending counterflow (primary symptom)
4	Etiology and pathogenesis	After falling down by mistake, the healthy Qi is severely frustrated, unable to resist the evil, with the inward invasion of exterior pathogen the greater Yang
5	Result of syndrome differentiation	Deteriorated syndrome of greater Yang disease (太阳病变证)
6	Treatment plan	Sweating method (therapeutic methods), Guizhi Tang (formula)
7	Scheme code	4



names (including those in the form of “greater Yang wind-invasion syndrome + Guizhi Tang syndrome”) were extracted from the original text of *Treatise on Febrile Diseases* by reference to the national standards *Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine* (revised in 2020), *Clinic Terminology of Traditional Chinese Medical Diagnosis and Treatment* (revised in 2020), *Selected Readings of Treatise on Febrile Diseases* (The Fourth Edition in The New Century), and *Annotations of Treatise on Febrile Diseases*.

Concerning the standardization of symptom names, due to the lack of unified national standards at present, various opinions are prevalent. Therefore, the preferred term for symptoms is set according to the *Norms for Terms of Common Symptoms in Clinical Chinese Medicine* (revised) [14]. Only when there is no symptom name involved in this work, the term in the *Selected Readings of Treatise on Febrile Diseases* (The Fourth Edition in the New Century) will prevail.

Concerning the standardization of the names of therapeutic methods and therapies, the classification and naming of therapeutic methods in the existing national standard *Clinic Terminology of Traditional Chinese Medical Diagnosis and Treatment — Part 3: Therapeutic Methods* (revised in 2020) are quite different from those in the original *Treatise on Febrile Diseases*. Therefore, to reflect the classification structure and naming method of therapeutic methods and therapies in the original *Treatise on Febrile Diseases*, this paper adopts the therapeutic methods, classification structure, and naming methods in *Selected Readings of Treatise on Febrile Diseases* (The Fourth Edition in the New Century) and *Annotations of Treatise on Febrile Diseases*.

Concerning the standardization of formula names, since the names of the formulas in *Treatise on Febrile Diseases* involved in the textbooks of *Chinese Medical Formulas* in modern colleges and universities have not changed, and do not include all the formulas in *Treatise on Febrile Diseases*, this paper takes the names of the formulas in the original text of *Treatise on Febrile Diseases* as the naming criteria. At the same time, according to the classification method of the *New Treatise on Febrile Diseases* [15] written by LIU Duzhou, all the formulas are divided into 18 categories.

In addition, if there are multiple synonyms under the above provisions, they should be stored in the synonym list. For example, “floating Yang and weak Yin” and “wind-invasion exterior deficiency syndrome” mentioned in the original Line 12 of the original work should be stored in the synonym list.

**2.3.2 Knowledge extraction** Knowledge extraction is the primary task of knowledge graph construction, to extract knowledge elements such as entities, attributes, and relationships from structured data, semi-structured data, and

unstructured data, which provides a knowledge base for the construction of the knowledge graph [16].

According to the knowledge extraction method of the original text of *Treatise on Febrile Diseases* described in this paper, we have converted the knowledge of the 10 chapters in the middle of the original work of *Treatise on Febrile Diseases* related to disease differentiation, syndrome differentiation, and treatment into the form of Table 2 and Table 3. Because the original lines of *Treatise on Febrile Diseases* are refined and rich in semantics, the context between the lines is close, and the original line involving the treatment of formulas usually includes a medical case, that is, the original syndrome, the treatment history, the current syndrome, the results of syndrome differentiation (or disease differentiation), and the treatment scheme, it is difficult to achieve the automatic extraction of knowledge on the premise of ensuring the accuracy of the extraction of knowledge of ancient Chinese medicine. Therefore, to accurately express the original knowledge of *Treatise on Febrile Diseases* under the guidance of experts of TCM classical formula with rich clinical experience and profound theoretical research, we completed the transformation of the above knowledge forms by referring to the notes of the planned teaching materials *Selected Readings of Treatise on Febrile Diseases* (The Fourth Edition in The New Century), the *Annotations of Treatise on Febrile Diseases*, and *Yi Zong Jin Jian* (《医宗金鉴》) [17] (the contents of the first 17 volumes, *Revised Zhongjing Quanshu, Annotations to Treatise on Febrile Diseases* 《订正仲景全书·伤寒论注》). A total of 318 tables were obtained.

Because the content of each cell in these knowledge tables of *Treatise on Febrile Diseases* has a clear range, and the types of synonyms and symptoms are also marked with special characters, we adopted a synchronous extraction method of entities and relationships based on dictionaries and rules. In the process of sorting out the knowledge system of *Treatise on Febrile Diseases*, we have collected relevant words from five aspects, namely, disease, syndrome, symptom, method, and formula, and added them to the dictionary as standard terms. In addition, we have also introduced the national standard *Basic Theory Nomenclature of Traditional Chinese Medicine* (2006) into the dictionary [18]. In each cell of the knowledge table, we used a series of special symbols (parentheses, commas, stop signs, plus signs, and slashes in Chinese and English) to achieve Chinese word segmentation by using the NLP-ICTCLAS Chinese word segmentation system [19], and completed the entity extraction of disease, syndrome, symptom, method, and formula.

In the process of constructing the knowledge graph model layer of *Treatise on Febrile Diseases*, we have defined 27 semantic relations and 44 constraint rules. Based on these provisions, we assigned specific entity types to each cell of the knowledge table of *Treatise on*

*Febrile Diseases*, as well as specific relationship types between cells. The specific meanings are shown in Table 5 and 6.

Based on the constructed terminological dictionary of *Treatise on Febrile Diseases* and the above knowledge

table rules, we extracted 3 142 triplets from 318 original knowledge tables of *Treatise on Febrile Diseases* and completed the relationship extraction of five aspects in *Treatise on Febrile Diseases*, such as disease, syndrome, symptom, method, and formula.

**Table 5** Representation of entities and attributes in knowledge tables of *Treatise on Febrile Diseases*

No.	Knowledge type	Knowledge content
1	<i>Original disease and syndrome</i>	Cell01: Composed of diseases, syndromes, and symptoms
2	<i>Treatment history</i>	Cell02: Composed of methods and formulas
3	<i>Current disease and syndrome</i>	Cell03: Composed of diseases, syndromes, and symptoms
4	<i>Etiology and pathogenesis</i>	Cell04: Composed of ordinary text as an attribute of the medical case type entity
5	<i>Result of syndrome differentiation or disease differentiation</i>	Cell05: Composed of diseases and syndromes
6	<i>Treatment plan</i>	Cell06: Composed of methods and formulas
7	<i>Scheme code</i>	Cell07: An integer between 1 and 4, representing the credibility of the medical case, as an attribute of the medical case type entity

Italicized words are fixed and refer to specific meanings. Cell 01 – cell 07 refer to five aspects of knowledge: disease, syndrome, symptom, method, and formula.

**Table 6** List of relationships between knowledge cells in *Treatise on Febrile Diseases*

Head entity cell		Relation type	Tail entity cell	
Number	Entity type		Number	Entity type
Cell02	Treatment	Treats	Cell01	Disease
Cell02	Treatment	Treats	Cell01	Syndrome
Cell02	Formula	Treats	Cell01	Disease
Cell02	Formula	Treats	Cell01	Syndrome
Cell02	Treatment	Mistreats	Cell01	Disease
Cell02	Treatment	Mistreats	Cell01	Syndrome
Cell02	Formula	Mistreats	Cell01	Disease
Cell02	Formula	Mistreats	Cell01	Syndrome
Cell06	Treatment	Treats	Cell05	Disease
Cell06	Treatment	Treats	Cell05	Syndrome
Cell06	Formula	Treats	Cell05	Disease
Cell06	Formula	Treats	Cell05	Syndrome
Cell06	Treatment	Uses	Cell06	Formula
Cell01	Syndrome	Co-occurs with	Cell01	Disease
Cell05	Syndrome	Co-occurs with	Cell05	Disease
Cell01	Disease	Transmission and transformation	Cell03	Disease
Cell03	Symptom	Manifestation of	Cell05	Disease
Cell03	Symptom	Primary symptom	Cell05	Disease
Cell03	Symptom	Accompanied symptom	Cell05	Disease
Cell03	Symptom	Pulse symptom	Cell05	Disease
Cell03	Symptom	Necessary symptom	Cell05	Disease
Cell03	Symptom	Manifestation of	Cell05	Syndrome
Cell03	Symptom	Primary symptom	Cell05	Syndrome
Cell03	Symptom	Accompanied symptom	Cell05	Syndrome
Cell03	Symptom	Pulse symptom	Cell05	Syndrome
Cell03	Symptom	Necessary symptom	Cell05	Syndrome

**2.3.3 Knowledge fusion** Since the knowledge obtained through knowledge extraction may have conflicts (polysemy) or overlaps (synonymy), it is necessary to use knowledge fusion technology to process the knowledge, so as to improve the quality of the knowledge graph and enrich the stock of knowledge [20]. In the knowledge graph of *Treatise on Febrile Diseases* constructed in this paper, the knowledge mainly comes from the original text of *Treatise on Febrile Diseases* and the case records of famous Shang Han experts. Since the original knowledge of *Treatise on Febrile Diseases* is the core of the whole graph, which determines the quality of the graph construction, and medical knowledge itself also has high requirements for accuracy, we adopted a manual approach to complete entity alignment for the triple set of original knowledge of *Treatise on Febrile Diseases* obtained in the knowledge extraction stage.

For the knowledge from famous Shang Han experts' case records, we used the method of combining rules and statistics to achieve entity alignment, and submitted similar entities to TCM classical formula experts for review to ensure the quality of knowledge fusion. Specifically, we first attempted to map entity names in triples to canonical entities using the thesaurus. If it was successful, the entity alignment process was completed. Otherwise, a statistical method was used to calculate entity similarity to achieve entity alignment, including character similarity calculation and semantic similarity calculation. We measured the character similarity by calculating the Jaccard Distance between two entity strings, as shown in Formula (1).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

Where,  $| \cdot |$  represents the number of characters,  $|A \cap B|$  the character intersection of entity A and entity B, and  $|A \cup B|$  the character union of entity A and entity B.

The basic principle of TCM in understanding and treating diseases is syndrome differentiation and treatment. The process of syndrome differentiation is the process of distinguishing symptoms to obtain the name of a syndrome (or disease). The process of treatment is the process of establishing a treatment based on the name of a syndrome and prescribing a formula accordingly. Therefore, we aligned entities in famous Shang Han experts' medical records in the order of symptoms, syndromes, diseases, methods, and formulas. For this reason, the attribute names of the entities to be processed and other entity names associated with the knowledge table were used as vectors, and unique heat coding was adopted. The same kind of entities corresponding to the original articles of *Treatise on Febrile Diseases* were also processed in the same way with case records as the knowledge unit, and then their cosine similarity was calculated. The calculation method is shown in Formula (2).

$$sim_s(A, B) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (2)$$

Where  $(x_1, x_2, \dots, x_n)$  denotes the vector representation of entity A, and  $(y_1, y_2, \dots, y_n)$  the vector representation of entity B.

The calculation method of entity similarity was thus obtained, as shown in Formula (3), and the entity with the highest similarity and greater than the threshold value of 0.8 was submitted as a candidate alignment entity to the TCM classical formula experts for review. If the audit is passed, the attributes and relationships of the two entities will be merged, and a synonymous relationship will be established between the entity and the normative entity in the knowledge graph. Otherwise, the entity will be added to the graph as a new entity.

$$sim(A, B) = 0.2 \times J(A, B) + 0.8 \times sim_s(A, B) \quad (3)$$

Where  $J(A, B)$  denotes the character similarity between entities A and B, and  $sim_s(A, B)$  the semantic similarity between entities A and B.

**2.3.4 Knowledge storage** Because there are a lot of relationships between entities in the knowledge graph built in this paper, and entities and relationships in the graph need to be frequently queried, this paper uses the Neo4j graph database [21] as the tool for knowledge storage, which is based on attribute graph model, and has efficient data storage structure and visualization capability. In addition, this paper employs the open-source front-end graphical tool Echarts [22] to show the query results. The network diagram of "disease, syndrome, symptom, method, and formula" is displayed in a graphical way.

Based on the semantic types and relationships established in the knowledge graph pattern layer of *Treatise on Febrile Diseases*, as well as data preprocessing, knowledge extraction, and knowledge fusion, we have constructed the "disease, syndrome, symptom, method, and formula" knowledge graph of *Treatise on Febrile Diseases*, which includes 6 469 entities and 10 911 relationship triads, concerning the 10 chapters in the middle of the original *Treatise on Febrile Diseases* involving disease identification, syndrome differentiation, and treatment, and 328 famous Shang Han experts' case records. Table 7 and 8 show the statistical results of the entities and relationships stored by Neo4j, respectively. Figure 2 shows the effect picture stored in Neo4j, with different colors representing entities with different labels.

## 3 Applications

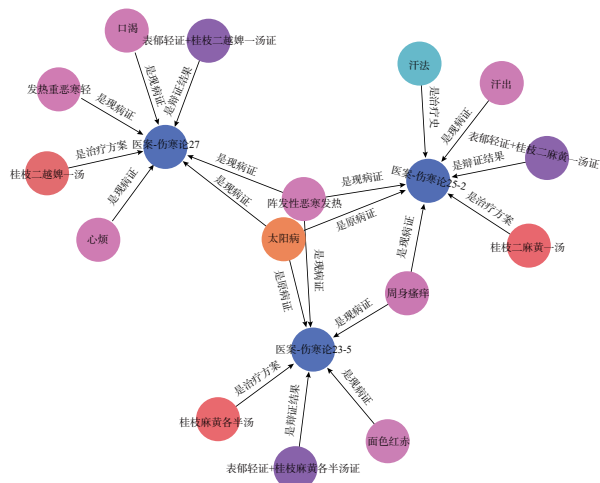
### 3.1 Query of knowledge structure of *Treatise on Febrile Diseases*

By embedding the Cypher query statement provided by Neo4j in the program, users can obtain the sum of









**Figure 5** Knowledge graph of case records with mild type of exterior stagnation syndrome

The purple node represents syndromes, the pink node represents symptoms, the orange node represents diseases, the red node represents formulas, and the dark blue node represents case records. See Table 1 for the relationship between nodes.

## 4 Discussion

The knowledge graph is a structured semantic knowledge base, which is widely used in intelligent search, in-depth question and answering, decision support, etc. Compared with the knowledge graphs in the field of *Treatise on Febrile Diseases* that we found, the knowledge graph constructed in this paper is no longer limited to a certain disease or decoction category, but focuses on the extraction of concepts, entities, and relationships from five aspects: disease, syndrome, symptom, method, and formula. The graph covers 25 concepts and 27 semantic relationships, including 6 469 entities and 10 911 relationship triplets, through the construction method of “from top to bottom as the main, and from bottom to top as the auxiliary”, which can fully display the knowledge structure of *Treatise on Febrile Diseases* and its interrelationships.

In the process of graph construction, we standardized the names of diseases, syndromes, symptoms, treatments, therapies, formulas, and other terms in *Treatise on Febrile Diseases* (see Section 2.3.1), and designed the method of extracting knowledge from the original book according to the knowledge characteristics of the original literature, which can better solve the problem that different data sources use different names for the same term, and improve the accuracy of knowledge. It is also conducive to subsequent knowledge extraction and knowledge fusion, and significantly improves the quality of the graph. In addition, in order to further expand the knowledge in the graph, we have developed a B/S-based medical record management system for *Treatise on Febrile Diseases*, which can import the case records involved in

the original literature and other case records of famous Shang Han experts into the system. For the above two types of case records, different knowledge fusion strategies were adopted (see Section 2.3.3) to improve the system performance while ensuring the accuracy of knowledge.

Due to the professionalism of the knowledge system of *Treatise on Febrile Diseases* and the inadequacy of the current knowledge extraction technology in dealing with Chinese ancient books and literature, this paper mainly adopts the method of “manual annotation + machine extraction + expert review” to extract the concepts, entities, and their interrelationships in the knowledge system of *Treatise on Febrile Diseases*, which makes the efficiency of graph construction to be improved and the entity scale to be enriched. In the next step of research, we will use knowledge extraction and knowledge fusion technology to enrich the clinical experience and case records of contemporary famous Shang Han experts into the existing graph, and study the machine translation technology of well-known TCM literature, to realize the automation of knowledge extraction of ancient works on *Treatise on Febrile Diseases*, and further improve the existing knowledge graph of *Treatise on Febrile Diseases*.

## 5 Conclusion

From the perspective of information processing, this paper combs the conceptual classification and relationship of *Treatise on Febrile Diseases* from five aspects: disease, syndrome, symptoms, method, and formula. Under the guidance of TCM experts, the knowledge extraction, knowledge fusion, and knowledge storage of the original text of *Treatise on Febrile Diseases* are completed by combining manual and machine methods. On this basis, the visual query of domain knowledge is realized, and the completeness and accuracy of knowledge representation are improved, which facilitates knowledge sharing and reuse, and lays the foundation for later knowledge reasoning and its application.

## Fundings

The Open Fund of Hunan University of Traditional Chinese Medicine for the First-Class Discipline of Traditional Chinese Medicine (2018ZYX66), the Science Research Project of Hunan Provincial Department of Education (20C1391), and the Natural Science Foundation of Hunan Province (2020JJ4461).

## Competing interests

The authors declare no conflict of interest.



## References

- [1] YANG YJ, XU B, HU JW, et al. Accurate and efficient method for constructing domain knowledge graph. *Journal of Software*, 2018, 29(10): 2931-2947.
- [2] ZENG ZL, ZHANG HM, YU T, et al. Review on the research and application of knowledge graph and its key technologies in the field of traditional Chinese medicine. *Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*, 2022, 24(2): 780-788.
- [3] WANG S, LI ZJ, YANG T, et al. Current status and development trend of knowledge graph research in traditional Chinese medicine. *Journal of Nanjing University of Traditional Chinese Medicine*, 2022, 38(3): 272-278.
- [4] AO DM, YANG YF, SUI ZF, et al. Preliminary study on the construction of Chinese medical knowledge graph. *Journal of Chinese Information Processing*, 2019, 33(10): 1-9.
- [5] ZHAO K, WANG HX, SHI N, et al. Study and implementation on knowledge graph of Guizhi Decoction associated formulas based on Neo4j. *World Chinese Medicine*, 2019, 14(10): 2636-2639, 2646.
- [6] LIN ZY, CHEN HT, LIU ZR, et al. Study on excavation of key points of syndrome differentiation of Taiyang Disease in Shang Han Lun based on generalized mechanics model. *Chinese Journal of Information on TCM*, 2022, 29(7): 12-16.
- [7] WANG JW, XIAO L, YAN JF, et al. Research on construction of knowledge graph of *Treatise on Febrile Diseases* based on Neo4j. *Computer & Digital Engineering*, 2021, 49(2): 264-267, 396.
- [8] WANG HF, DING J, HU FK, et al. Survey on large scale enterprise-level knowledge graph practices. *Computer Engineering*, 2020, 46(7): 1-13.
- [9] LIU Q, LI Y, DUAN H, et al. Knowledge graph construction techniques. *Journal of Computer Research and Development*, 2016, 53(3): 582-600.
- [10] LIU DZ. Annotation of *Treatise on Febrile Diseases*. Beijing: People's Health Publishing House, 2013.
- [11] WANG QG. Selected readings of *Treatise on Febrile Diseases*. Beijing: China Traditional Chinese Medicine Press, 2016.
- [12] STUDER R, BENJAMINS VR, FENSEL D. Knowledge engineering: principles and methods. *Data and Knowledge Engineering*, 1998, 25(122): 161-197.
- [13] YU T, CUI M, LI HY, et al. Semantic network framework of traditional Chinese medicine language system: an upper-level ontology for traditional Chinese medicine. *China Digital Medicine*, 2014, 9(1): 44-47.
- [14] LI JB, MA L. Terminology specification for common clinical symptoms of Traditional Chinese Medicine (revised). Beijing: China Pharmaceutical Science and Technology Press, 2015.
- [15] LIU DZ. Newly organized classified-formula of *Treatise on Febrile Diseases*. Beijing: People's Health Publishing House, 2013.
- [16] TIAN L, ZHANG JC, ZHANG JH, et al. Knowledge graph survey: representation, construction, reasoning and knowledge hypergraph theory. *Journal of Computer Applications*, 2021, 41(8): 2161-2186.
- [17] WU Q. *Yi Zong Jin Jian of royal compilation (WU YD version)*. Beijing: People's Health Publishing House, 1998.
- [18] GB/T 20348-2006. Terminology of basic theory of traditional Chinese medicine. Beijing: General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, China National Standardization Administration, 2006. Available from: <https://openstd.samr.gov.cn/bz/gk/gb/newGbInfo?hcno = EFB5E3CEF5147682E9678C7F9DA2CBDE>.
- [19] NLP-ICTCLAS Chinese word segmentation system. Available from: <http://ictclas.nlpir.org/>.
- [20] MA ZG, NI RY, YU KH. Recent advances, key techniques and future challenges of knowledge graph. *Chinese Journal of Engineering*, 2020, 42(10): 1254-1266.
- [21] LV HK, HONG L, MA FC. Constructing knowledge graph for financial equities. *Data Analysis and Knowledge Discovery*, 2020, 4(5): 27-37.
- [22] The Apache Software Foundation. Apache ECharts: an open source visual chart library based on JavaScript. Available from: <https://echarts.apache.org/zh/index.html>.

## 《伤寒论》知识图谱构建与应用

刘东波, 韦昌法, 夏帅帅, 晏峻峰\*

湖南中医药大学信息科学与工程学院, 湖南长沙410208, 中国

**【摘要】目的** 建立《伤寒论》“病-证-症-法-方”知识图谱,降低数据的模糊性和不确定性问题,为后期的知识推理及其应用奠定基础。**方法** 在中医经方专家指导下,采用“自顶向下为主,自底向上为辅”的方式,针对《伤寒论》原文从病、证、症、法、方五个方面进行知识抽取、知识融合和知识存储,完成《伤寒论》知识图谱的构建,在此基础上实现知识结构与关联性查询及其可视化。**结果** 完成《伤寒论》“病-证-症-法-方”知识图谱的构建,共得到6469个实体和10911个关系三元组,可以进行实体及其之间关系的查询,并以可视化的形式展示查询结果。**结论** 《伤寒论》知识图谱较为系统的实现了《伤寒论》知识体系的数字化,提高了知识表示的完备性和准确性,能够清晰、高效的获取“病-证-症-法-方”之间的联系,有利于知识的共享和重用。

**【关键词】** 《伤寒论》;知识图谱;本体;图数据库;知识抽取;知识融合