Artificial Intelligence in Medicine 77 (2017) 48–52

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Knowledge graph for TCM health preservation: Design, construction, and applications

Tong Yu*, Jinghua Li*, Qi Yu, Ye Tian, Xiaofeng Shun, Lili Xu, Ling Zhu, Hongjie Gao

Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medicine Sciences, 16 Dongzhimen South Alley, Dongcheng District, Beijing, 100700, PR China

ARTICLE INFO

Article history:
Received 8 March 2017
Received in revised form 30 March 2017
Accepted 8 April 2017

Keywords:
Knowledge graph
Ontology
Knowledge services
Traditional Chinese medicine
Health preservation

ABSTRACT

Traditional Chinese Medicine (TCM) is one of the important non-material cultural heritages of the Chinese nation. It is an important development strategy of Chinese medicine to collect, analyzes, and manages the knowledge assets of TCM health care. As a novel and massive knowledge management technology, knowledge graph provides an ideal technical means to solve the problem of "Knowledge Island" in the field of traditional Chinese medicine. In this study, we construct a large-scale knowledge graph, which integrates terms, documents, databases and other knowledge resources. This knowledge graph can facilitate various knowledge services such as knowledge visualization, knowledge retrieval, and knowledge recommendation, and helps the sharing, interpretation, and utilization of TCM health care knowledge.

1. Introduction

Traditional Chinese Medicine (TCM) is an important part of the Chinese nation's outstanding culture, and it has a long history of development. Ancient physicians attach great importance to health and longevity. They have accumulated rich experience in health care throughout clinical activities and daily-life practices, and also created a system of theory and methods of health care. It is an important strategy for TCM practitioners to systematically sort out and preserve TCM health care knowledge. With the people's living standards are increasing, Chinese people began to pay more attention to quality of life, so the health knowledge has become an increasingly strong demand.

In recent years, the application of information technology in the field of TCM has made great progress. TCM practitioners constructed a large number of terminology systems, document libraries and databases, leading to the formation of a relatively complete system of Chinese medicine scientific data [1]. These digital resources contain a wealth of knowledge, which can support the development of knowledge systems in health care. However, the related knowledge resources are scattered and disconnected, like a set of "knowledge islands". The lack of data integration led to low utilization of resources and hinders the capacity of services. There are still gaps between the people's needs for health knowledge and the quality of services.

Knowledge Graph is an emerging technology for massive knowledge management and intelligent services in the big data era [2,3]. A "knowledge graph" is a huge knowledge system built on the "Semantic Network". It can capture and present the intricate relationships among the domain concepts, and connect the fragmented pieces of knowledge in various information systems. Knowledge graphs are very useful in knowledge retrieval, question-answering, knowledge recommendation, knowledge visualization and other applications. Knowledge graph provides an ideal technical means to solve the problem of "knowledge island" in the field of traditional Chinese medicine, and it is helpful to realize the integration of knowledge resources and enhance the knowledge service ability.

As shown in Fig. 1, knowledge graph is a graph-based knowledge representation and organization method. It uses a set of "subject-predicate-object" triples to represent the various entities and their relationships in a domain. Knowledge graph can also be seen as a huge network, in which nodes represent domain entities and arcs represents semantic relationships between entities. The core component of knowledge graph is "semantic network". The knowledge graph adds more knowledge content into the framework of semantic network, such as the information of the domain entity and the link of related literature resources. These knowledge resources come from digital resources such as databases, document library, and data files, which may be dispersed in different information sys-

* Corresponding authors.
E-mail addresses: yutongoracle@hotmail.com (T. Yu), zingarlee@hotmail.com (J. Li).

http://dx.doi.org/10.1016/j.artmed.2017.04.001
0933-3657/

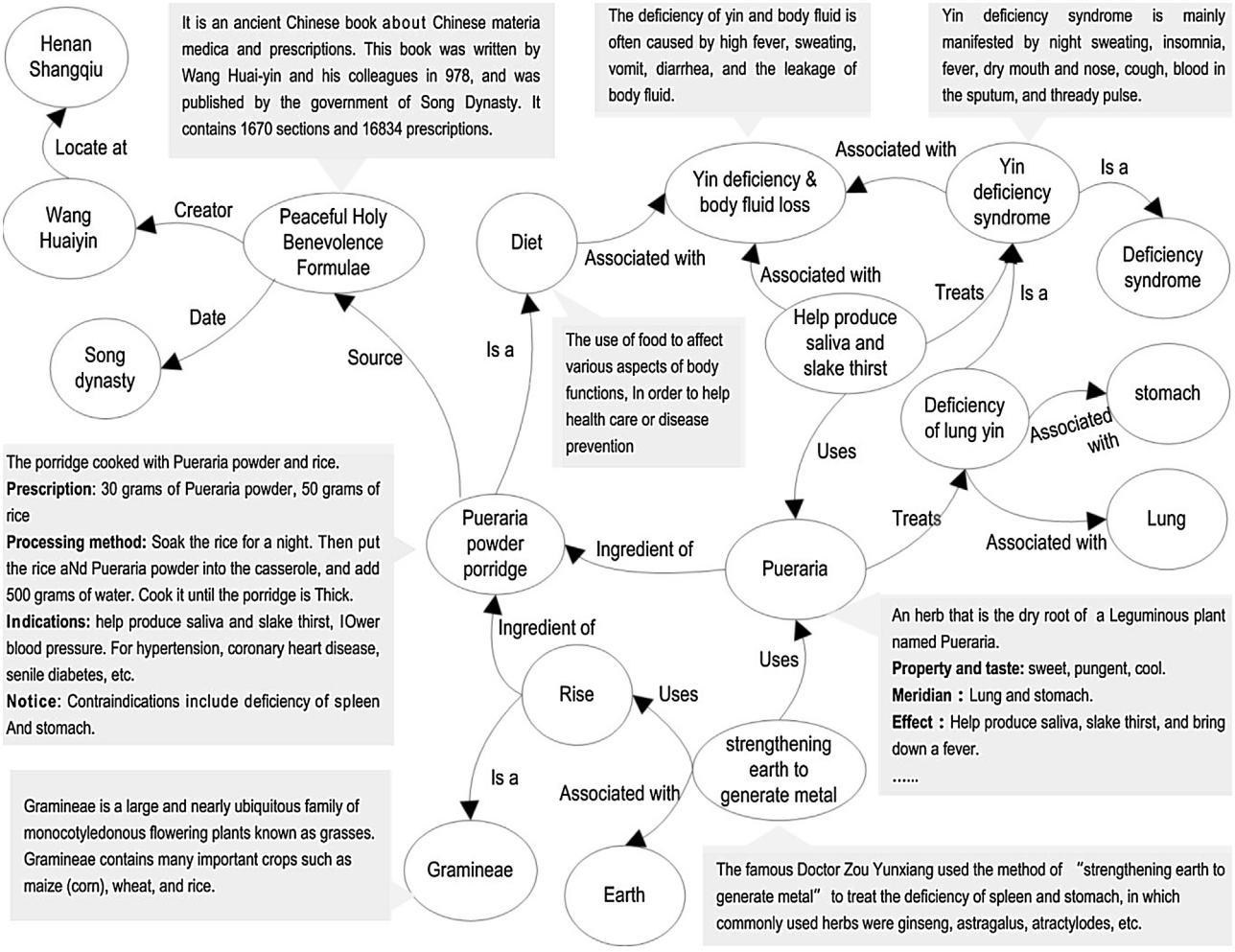

Fig. 1. An illustrative example of knowledge graph for TCM health care.

tems and organizations. The knowledge graph uses the semantic network as the skeleton to collect and systematically organize the scattered domain knowledge and realize the functions of knowledge retrieval, knowledge query, knowledge browsing, knowledge display and knowledge visualization.

The content of the TCM knowledge graph is explained with an example as shown in Fig. 1. This graph contains TCM concepts of various types, such as principles, human organs, diseases, herbs, methods, techniques, famous doctors, and classics. For example, the concept "Earth" comes from the "five phase theory", which is one of the philosophical theories of medical practice in ancient China, concerning the five phases: wood, fire, earth, metal and water, and their movements and changes [7]. The concept "Strengthening Earth to Generate Metal" is a therapeutic principle according to the "five phase theory", which uses herbal therapies (e.g., Pueraria) that reinforce the spleen (belongs to earth) in order to benefit the lung (belongs to metal). Accordingly, the graph connects the concept "Pueraria" and the concept "Deficiency of lung yin" with an edge labeled "Treats", which means Pueraria can be used to treat the deficiency of lung yin (which means the yin fluid that moistens the lung). Also, the graph establishes an "Ingredient of" relation between "Pueraria" and "Pueraria Powder Porridge", which is the porridge cooked with Pueraria and rise. In addition, the graph connects the porridge with its source as an ancient Chinese book "Peaceful Holy Benevolence Formulae", connects the book with its creator Wang Huaiyin who was a famous doctor, and so on. In summary, the TCM knowledge graph is an intuitive and scalable knowledge model that can encapsulates essential TCM knowledge through concepts, relations, and annotations.

In recent years, knowledge graph has become a hot topic in knowledge management, knowledge service and other fields. Internet companies have introduced knowledge graphs to improve the quality of services, such as *Google Knowledge Graph*, *Baidu Knowledge Graph*, and *Sogou Knowledge Cube*. TCM health care has its own characteristics and needs, and it is necessary to study the construction method of TCM knowledge graph and solve the key technical problems, in order to realize the rapid growth of knowledge graph and fully realize its application value. In view of this, this project takes the conceptual system of traditional Chinese medicine as the focus of study, and integrates related structural knowledge resources into a large-scale knowledge graph. This knowledge graph can be embedded in the literature search, knowledge retrieval and other systems to provide knowledge services such as knowledge navigation, integration, and visualization. It promotes knowledge sharing, dissemination and utilization of TCM health care, and has the value of reference and service for TCM practitioners as well as ordinary people.

2. Construction of knowledge graph

The purpose of this study is to construct the TCM health care knowledge graph by using the data resources accumulated in the

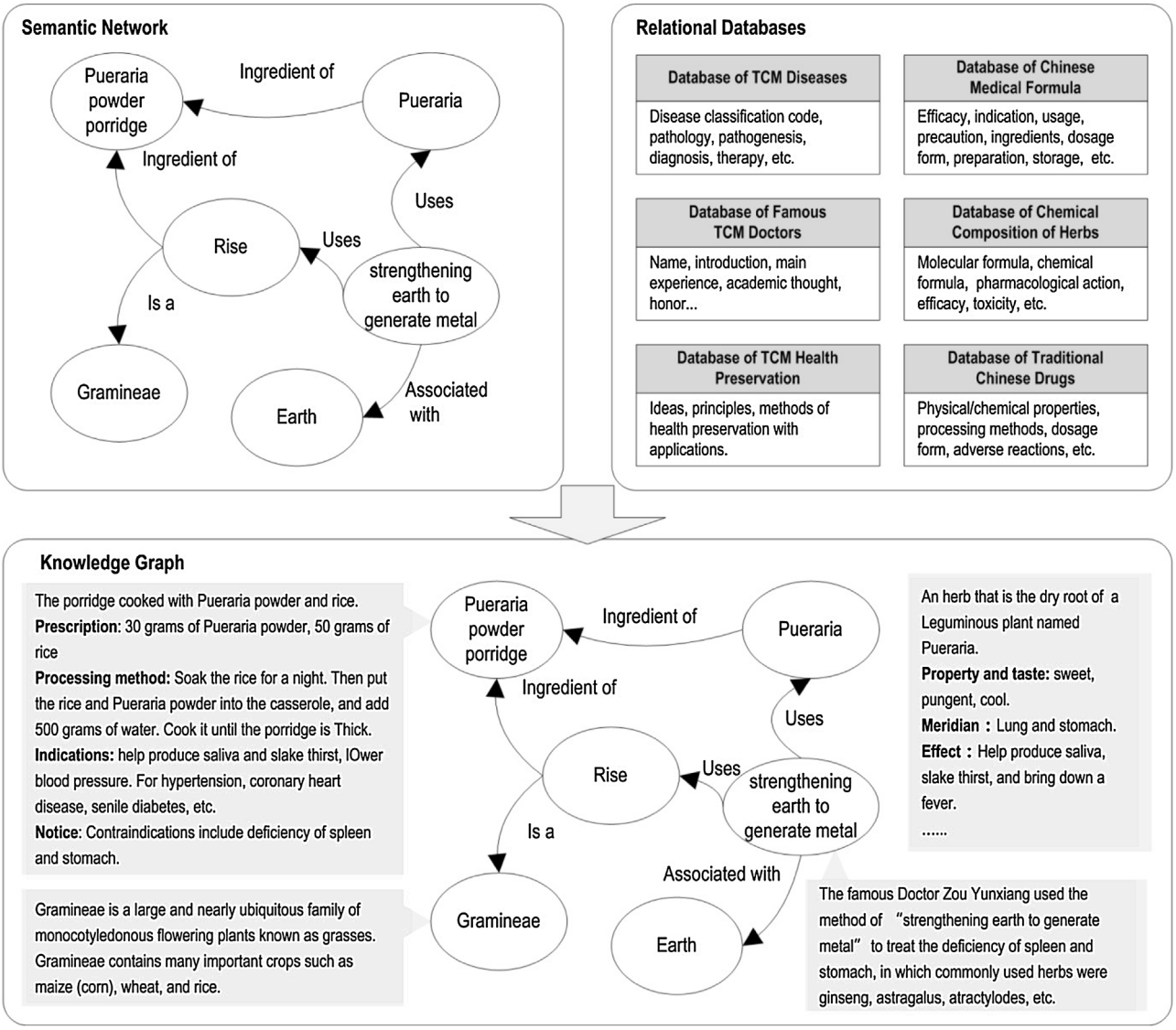

Fig. 2. Construction of knowledge graph for TCM health care.

field of traditional Chinese medicine. As shown in Fig. 2, the general idea of this study is: the existing data resources are integrated to form a knowledge graph with semantic network as the skeleton, so as to realize the systemic organization of terminology, literature, database and other related knowledge resources and to support intelligent knowledge services.

The "TCM health care ontology" is the basis for the construction of traditional Chinese medicine health knowledge graph. In the previous study, we use the methods of ontology learning and manual compilation to construct an ontology by referring to relevant papers, dictionaries, textbooks and reference books. This ontology describes TCM health care concepts and expresses the concept system of this field accurately. It is divided into two parts: "top-level ontology" and "thesaurus". "Top-level ontology" is essentially a semantic network framework. It consists of two parts: (1) a hierarchical system of semantic types; (2) a hierarchical system of semantic relations. The top-level ontology classifies the domain concepts and knowledge resources reasonably and defines the correlation between domain concepts. The "Thesaurus" is a structured collection of a wide range of concepts and terminology. This ontology basically covered the concept system and provided a relatively complete framework for constructing TCM health knowledge graph.

On the other hand, the existing TCM databases provide the data resources for populating the knowledge graph. TCM practitioners selected and rigorously examined knowledge from professional literature with clear source, and recorded it in the database. They have established a large number of databases about diseases, prescriptions and chemical constituents of Chinese herbal medicine, which constitutes a relatively complete database system. For example, Fan et al. [4] built a "TCM health care database" through digitization of TCM classics, which effectively supports ancient books preserving and knowledge mining. In another example, Yu et al. [5] constructed a specialized database that contains rich and accurate knowledge about the ideas, principles, and methods of TCM health care.

We constructed the TCM health care knowledge graph through ontology-based database integration. First, we populated the content of databases into the knowledge graph, and implemented the query. Second, we extend the knowledge graph by acquiring knowledge from text in a semi-automatic process: A text-mining tool was used to perform semantic analysis on texts and extract entities

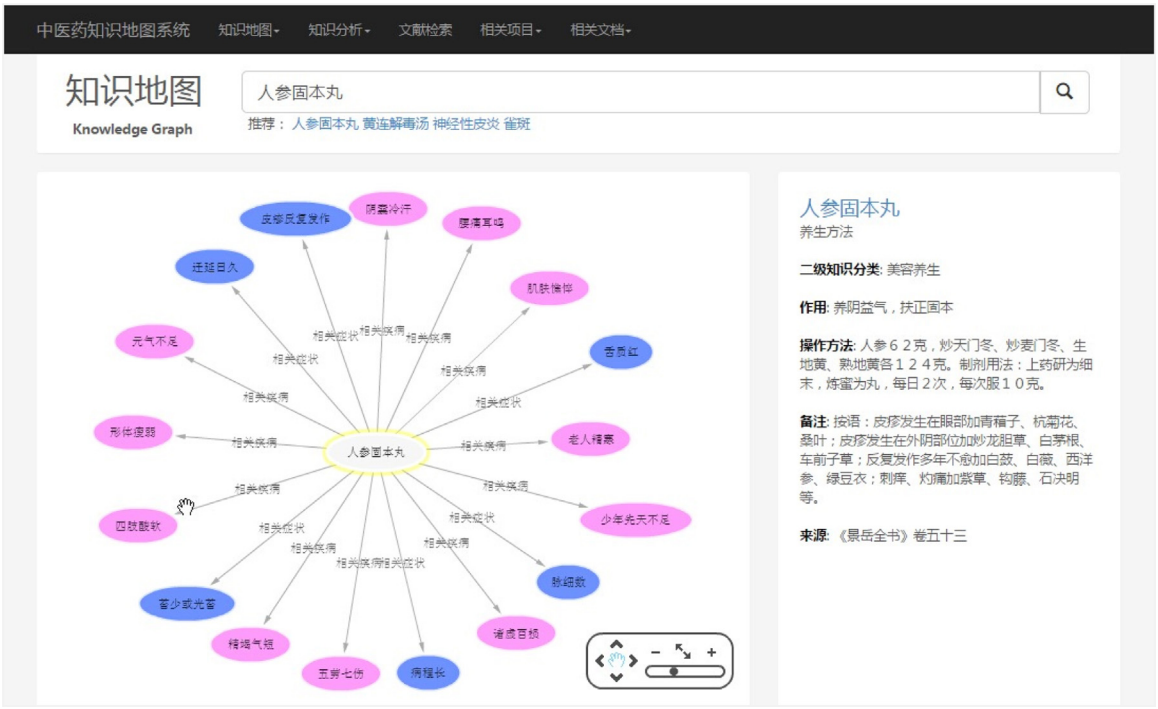

Fig. 3. Knowledge visualization for TCM health care (e.g., Ginseng Guben pills).

and relations, which were added into the knowledge graph after expert validation. In addition, we developed an editing tool that can help domain experts to browse, examine, and edit the knowledge graph, which helps to ensure the accuracy and completeness of the knowledge graph.

3. Applications

After the establishment of the knowledge graph, the method of searching, browsing and visualizing based on the knowledge graph was developed. The knowledge graph was also embedded into the TCM knowledge service platform, which provides authoritative, accurate, comprehensive knowledge of traditional Chinese medicine health care for Internet users [6]. The Web address of the system is: http://www.tcmkb.cn/kg/. The applications of the knowledge graph are described as follows.

- Knowledge visualization

As shown in Fig. 3, the knowledge graph can graphically show the association between domain concepts and provide brief conceptual information. Users can interactively browse the domain concept and select one of the concepts to further explore the information or construct the query. The knowledge graph can establish the connection between "knowledge islands", and enhance the connectivity of the knowledge resources in the TCM domain. It can help users to browse the TCM knowledge at the concept level, and discover potential connections between concepts, so as to better control the complexity of TCM health knowledge system.

- Knowledge retrieval

Knowledge graph can significantly improve the effectiveness of knowledge retrieval system, making it more comprehensive, accurate and intelligent. Knowledge graph makes the knowledge retrieval model transition from "text-centered" to "things-centered". Based on the Knowledge graph, a "knowledge card" can be embedded in the retrieval system: the system will recognize the related entities from the user input and present the knowledge about the entity, including name, type, introduction, description, related documents and semantic relations. The Knowledge graph is also the core component of the question-answering system, which makes the system understand the user's query intention and directly returns the accurate answer instead of a large number of search results.

- Knowledge Recommendation

Based on the knowledge graph, we built the intelligent recommendation system which provides useful knowledge and effective advice based on user's personal health information. A contextual knowledge model was designed to establish the relationships between user personal health information and TCM knowledge system. For example, establish "functionally related to" relationships between "constitutions" and "health care methods", "specialized in" relationships between "experts" and "diseases", "treats" relationships between "prescriptions" and "diseases" and so on. According to these associations and user personal information (such as gender, age, constitution, disease, syndromes, symptoms, etc.), personalized recommendations can be made. For example: A user who is obese can be recommended the advise to prevent diabetes, and a user who is irritable can be recommended "clear fire" medicine. By combining the knowledge graph with expert scoring and user click-through rates, we can establish a scientific and reasonable knowledge recommendation mechanism to help users formulate useful diet, life and exercise planning according to their own characteristics, so as to achieve disease prevention and to improve the quality of life.

In summary, the knowledge graph technique was used to classify, organize and correlate TCM health care knowledge. It has played a key role in the knowledge service platform which provides real-time, accurate, personalized knowledge services for the

people. The practical applications show that the knowledge graph can highlight the core concept of the domain, and quickly present the structure of the knowledge system. Compared with other methods such as reading literature, the time for knowledge search and knowledge acquisition can be saved.

4. Conclusions

Knowledge graph is an emerging technology in the field of Internet, and is rapidly becoming a hot research topic in the field of knowledge management. This study aims to introduce the knowledge graph technology into the field of TCM health preservation. We establish a new knowledge base system of TCM health preserving Knowledge graph, and developed the corresponding technical system of knowledge organization, knowledge acquisition and knowledge services. Knowledge graph provides a novel method for the systematic organization and deep analysis of TCM health knowledge system, and has broad application prospects in knowledge management, knowledge service, education, and training in the field of TCM health care.

Acknowledgements

This work is supported by “the Fundamental Research Funds for the Central public welfare research institutes (NO. ZZ090305, NO. ZZ070804, NO. ZZ090317)” and “Beijing science and technology development fund of traditional Chinese Medicine (NO. JJ2014-61)”.

References

[1] Meng Cui, Ai-ning Yin, Wei-yu Fan, Yuan-bai Li. Research progress on scientific data construction of traditional chinese medicine. Chin J Inf Traditional Chin Med 2006;13(11):104–5 [Article in Chinese] http://xxzz.cintcm.com/oa/pdfdow.aspx?Sid=200611056 [accessed 2017-3-29].
[2] Wang Hao-fen. Large scale knowledge graph technology. Chin Soc Comput Sci 2014;10(3):64–8 [Article in Chinese] http://www.36dsj.com/archives/39306 [accessed 2017-3-29].
[3] Yu Tong, Liu Jing, Jia Li-rong, Zhang Zhulv, Yang Shuo, Liu Lihong, et al. Study on the construction of large scale knowledge graph of traditional chinese medicine. Chin J Digital Med 2010;10(3):80–2 [Article in Chinese] http://jcdm.org.cn/Magazine/PDFShow.aspx?ID=141076 [accessed 2017-3-29].
[4] Fan Wei-yu, Qiu Jian, Fu Yong-chi, Li Bin, Zhang Weina, Liu Pei-sheng, et al. Study on the database of classics of traditional Chinese medicine health preservation. Int J Chin Mater Med 2008;30(4):250–2 [Article in Chinese] http://d.wanfangdata.com.cn/Periodical/gwyx-zyzyfc200804004 [accessed 2017-3-29].
[5] Yu Qi, Li Jing-hua, Gao Hong-jie, Xie Qi, Zhou Hong-wei, Yu Tong. Construction of Chinese medicine health database. Chin J Digital Med 2015;10(8):73–5 [Article in Chinese] http://jcdm.org.cn/Magazine/PDFShow.aspx?ID=141342 [accessed 2017-3-29].
[6] Yu Tong, Li Jing-hua, Yu Qi, Li Zong-you, Wang Ying-hui, Gao Hong-jie, et al. Study on the construction of traditional chinese medicine health knowledge service platform. The 3rd Chinese Medicine Information Conference, Ordos, July 2016 2016 [Article in Chinese] http://www.tcmkb.cn/docs/TCMHealthPreservationPlatform.pdf [accessed 2017-3-29].
[7] WHO Regional Office for the Western Pacific. WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region. http://www.wpro.who.int/publications/PUB_9789290612483/en/ [accessed 2017-3-30].