

SOFTWARE

Open Access



AMFormulaS: an intelligent retrieval system for traditional Chinese medicine formulas

Yidi Cui^{1,2†}, Bo Gao^{3†}, Lihong Liu³, Jing Liu³ and Yan Zhu^{3*} 

From International Conference on Health Big Data and Artificial Intelligence 2020 Guangzhou, China. 29 October - 1 November 2020

Abstract

Background: Formula is an important means of traditional Chinese medicine (TCM) to treat diseases and has great research significance. There are many formula databases, but accessing rich information efficiently is difficult due to the small-scale data and lack of intelligent search engine.

Methods: We selected 38,000 formulas from a semi-structured database, and then segmented text, extracted information, and standardized terms. After that, we constructed a structured formula database based on ontology and an intelligent retrieval engine by calculating the weight of decoction pieces of formulas.

Results: The intelligent retrieval system named AMFormulaS (means Ancient and Modern Formula system) was constructed based on the structured database, ontology, and intelligent retrieval engine, so the retrieval and statistical analysis of formulas and decoction pieces were realized.

Conclusions: AMFormulaS is a large-scale intelligent retrieval system which includes a mass of formula data, efficient information extraction system and search engine. AMFormulaS could provide users with efficient retrieval and comprehensive data support. At the same time, the statistical analysis of the system can enlighten scientific research ideas and support patent review as well as new drug research and development.

Keywords: Traditional chinese medicine, Formula, Database, Retrieval system

Background

Traditional Chinese medicine (TCM) is a science that studies human life, health, and disease as well as a summary of the valuable experience of the Chinese nation in long-term survival and practice. Doctors of TCM treat patients based on syndrome differentiation by looking, listening, questioning, and feeling the pulse, then give TCM prescriptions to acquire the therapeutic effect. So,

formula is an important therapeutic concept in TCM and always the research hotspot. There are many research directions of formulas, for example, the research about theory of formulas compatibility [1–4], the research about dosimetry of formulas [5, 6]; and the research about formula and disease [7–9].

However, knowledge of formulas is mainly recorded in diverse TCM books which results in the difficulties of retrieval and acquirement. Consequently, integrating formula information as well as constructing database can greatly improve the efficiency of retrieval, bring convenience of knowledge acquisition and utilization for researchers and clinical doctors. Nowadays, there are

*Correspondence: zhuyan166@126.com

[†]Yidi Cui and Bo Gao have contributed equally to this work and should be considered co-first authors

³ Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China
Full list of author information is available at the end of the article



some formula databases, Guo et al. constructed a formula knowledge graph which presented the knowledge by the way of node-relation-node, and knowledge of the graph included traditional Chinese medicine, dosage, traditional Chinese Medicine processing, efficacy, and so on [10], Min He et al. constructed a traditional Chinese medicine database in the form of node and property, knowledge of the database included Chinese medicines, original plants, bioactive components, and the function of search and display were provided [11]. Both of the two databases store and display formula information based on the node-edge-node. However, the form of node-edge-node could only show the most important information by some terms rather than complete sentences, which might consult information loss; Shen et al. structured the listed proprietary Chinese medicine data and built the Chinese patent medicine database which integrated patent medicine information, but the patent medicine information was not sufficient for academic and clinical research [12]; Ruichao Xue et al. constructed a traditional Chinese medicine integrative database to integrate the traditional Chinese medicine and western medicine which included some TCM knowledge, like formula, TCM drugs, and herbal ingredients. This database mainly focused on the herb molecular mechanism analysis, and didn't meet the needs of other formula research [13].

The above databases have some basic functions, like information retrieval and knowledge display. There are some limitations that need improvement, such as being small-scale and inaccessible to the original information. Meanwhile, with the development of computer technology, users tend to choose rich, strong correlation retrieval results in practical operation. To improve the efficiency of retrieval and utilization as well as realize the knowledge mining and discovery of formulas, we proposed an intelligent retrieval system, named AMFormulaS (means Ancient and Modern Formula system, 古今方药系统 in Chinese), which was based on a database containing a large number of formula and the relevant information of formula, like name of formula, composition, dosage of Chinese medicine and so on. The system also can efficiently extract formula information from the text data to extend the size of the database. In the meantime, we also proposed a method of weight calculation on formula drugs to improve the retrieval efficiency which compose the core part of the intelligent search engine.

Methods

In the study, we firstly constructed an automatic standardization system that embedded the word segmentation packages and term dictionary. The semi-structured data was processed into structured and standardized formula records. A structured formula database was designed by

incorporating the ontology modeling method. Meanwhile, we designed and implemented an algorithm of weight calculation on formula to improve the retrieval efficiency. Lastly, the formula intelligent retrieval system AMFormulaS was realized (See the pipeline in Fig. 1).

Data sources

In this study, the data of AMFormulaS was selected and collected from the formula database maintained by the Institute of Information of Chinese medicine, Chinese academy of traditional Chinese medicine. It is a semi-structured database including 85,989 ancient and modern Chinese medicine formulas from more than 710 ancient books and modern literature. Considering the difference between ancient and modern medication habits, we took the modern medication habits as the standard reference and selected 38,000 formulas whose medicinal source of the component could be found or ascertained at present.

Data processing

Word segmentation

Formula database involves a large amount of data, so it is necessary to use computer technology to improve the efficiency of data processing. Considering the huge workload and our existing researches, we decided to adopt the information extract solution of word segmentation algorithm integrated with large-scale terminology.

The ancient Chinese medicine text is distinctive in grammar and expression as well as professional terms. Therefore, the standard of word segmentation needed to be investigated and determined initially. In one of our previous studies [14], we constructed a corpus for training the algorithm of word segmentation. According to the classification method of *traditional Chinese medicine philology* [15], we firstly selected 30 TCM ancient books of Qing Dynasty involving 10 categories: Materia medica, formulas, febrile diseases, internal medicine, surgery, gynecology, pediatrics, facial features, acupuncture and massage, medical cases respectively, and manually selected 150 pieces of rough corpus which contained 1705 sentences and 88,889 words from these books to train the model. Then the selected corpus was tagged manually by referencing TCM teaching material for higher education students [16], TCM reference books [17], and TCM related standard [18, 19] as terminology sources. After that, we preliminarily summarized the standard of word segmentation in TCM ancient books, that was taking the existing facts and semantic changes as the primary principle and considering the principles of part-of-speech grammar and semantic type in the meantime. There were 17 semantic types of text which were segmented based on the principle including physiology, symptom, syndrome, pathological factors, pathological

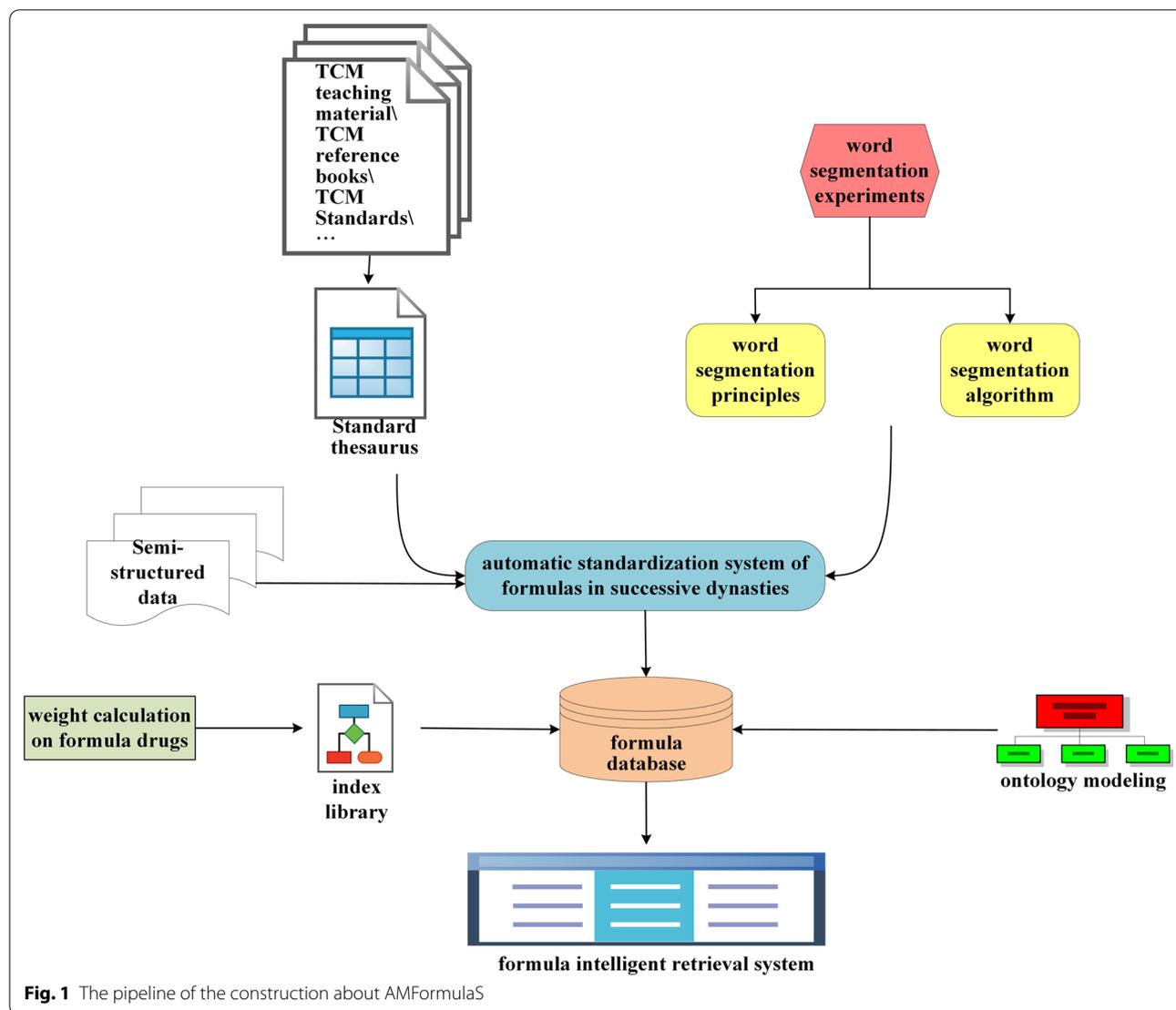


Fig. 1 The pipeline of the construction about AMFormulaS

products, efficacy, method of treatment, channel meridian and acupuncture points, four diagnostic methods, traditional Chinese drug, prescription, nature and flavor, toxicity, processing, contraindications, decoction method, and proprietary words in Chinese medicine.

After manual labeling the training set of word segmentation, we trained a model based on the algorithm of capsule network [20]. Compared with other algorithms, the algorithm of capsule network showed a good performance for word segmentation in ancient traditional Chinese medicine literature, so the capsule network model was used for word segmentation.

Information extraction and standardization

Due to the heterogeneity in the data structures and lack of standards for formula information, we built a system

named automatic standardization system of formulas in successive dynasties to extract and standardize the information of formulas under the guidance of the above-mentioned word segmentation standard and algorithm. The system firstly realized the identification, extraction, and standardization of formula, then the processed data was submitted to the formula database after manual verification. The extracted and standardized content includes name, composition, source, formation year of formulas, the dose of decoction pieces, the processing method of natural crude Chinese medicine, etc.[21]. For example, the formula of Tiefen pellet (铁粉丸 in Chinese) comes from the book *You you xin shu* (《幼幼新书》 in Chinese) written in the Song Dynasty. The system could transform the text into structured data. Firstly, the system recognized the information of the formula in the form

of text, like name of the formula, traditional Chinese medicine, dosage, and then extracted and standardized this information. For instance, one of the components is “Shehuang(蛇黄 in Chinese)” in the original records, the system identified it and normalized to “Shehanshi(蛇含石 in Chinese)”; the dosage and the measuring unit also could be normalized (as shown in Fig. 2).

Design of formula database

There is a variety of information about formulas, therefore, all contents of the formula database should be completed and well-organized including formula name, source of formula, formation date, author, and composition of formula, etc. Employed our formal research [22, 23], on the ontology-based modeling, the concept, relation, and property were analyzed and determined, then schema of formulas database was designed based on the conceptual modeling method of ontology and the authoritative references of TCM, such as Pharmacopoeia of the People’s Republic of China (part 1) [24], Coding Rules and Codes of Traditional Chinese Medicine [25], Chinese materia medica [26], Dictionary of traditional Chinese medicine [27], etc. The entities of the ontology model contain the information of formula (name, source, author, the subordinate departments, effect, and nature, flavor and channel tropism), information about Chinese medicine (medicinal name, medicinal sources, effect, Chinese patent medicine, decoction pieces, and effect

and nature, flavor and channel tropism), the core concept graph of formula database is shown in Fig. 3.

Implementation of the intelligent retrieval system

There exist some formula databases or retrieval systems that integrate some formula information, most of which only support full-text retrieval or retrieval by keyword. Yet, formulas are composed based on the TCM theory named Monarch, Minister, Assistant and Guide (君臣佐使 in Chinese). In the context of drug retrieval, users prefer to get the results which the search term of composition herb plays an important role. For example, by inputting ‘processed licorice (炙甘草 in Chinese);’ the users usually expect the result including formulas in which processed licorice play the Monarch role. Hence, a method of weight calculation on formula drugs was proposed in this research which made results be sorted by the importance of decoction pieces in formula or formation time of formula [28]. In this research, three factors were included to calculate the weight of the drug composition in the formula:

- (1) Whether the decoction piece is part of the name of the prescription which is processed by string matching;
- (2) The relative dose of traditional Chinese medicine. The relative dose was calculated as:



Fig. 2 The information extraction and standardization of Tiefen pellet

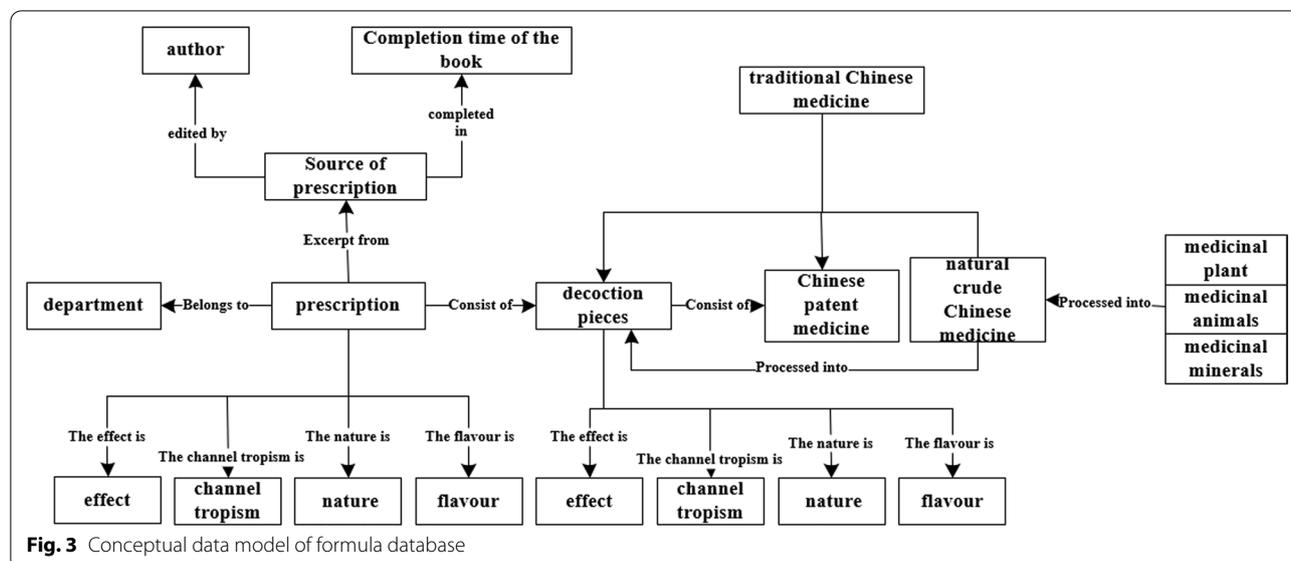


Fig. 3 Conceptual data model of formula database

$$f(o|S) = \frac{1}{n} \sum_{i=1}^n G_h(t, t_i) = \frac{1}{n\sqrt{2\pi}} \sum_{i=1}^n e^{-\frac{d(t,t_i)^2}{2h^2}} \tag{1}$$

where $d(t, t_i)^2$ represents the dose-distance between tuples t, t_i , $G_h(t, t_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d(t,t_i)^2}{2h^2}}$ is the Gaussian kernel function, n represents the number of different doses (or dose intervals) in T.

- (3) Whether decoction pieces are commonly used. The weight was calculated as:

$$w(t) = \log(n/f(t)) \tag{2}$$

where $w(t)$ is the weight of decoction piece t , n is the number of all different decoction pieces in set S , $f(t)$ is the number of formulations containing the specified decoction pieces t .

- (4) Multiple linear regression was used to calculate the optimal parameters:

$$y = w_0 + x_1w_1 + \dots + x_iw_i + \dots + x_nw_n \tag{3}$$

where x_1 : Whether the drug is commonly used, that is, the occurrence frequency of the drug; x_2 : Whether the drug appears in the drug name; x_3 : The ratio of the dose used to the general dose of the drug.

Based on a training data set of 400 records (part of the experimental results shown in Table1), the obtained training parameters were:

$$w_1 = 52.8231, w_2 = 0.8773, w_3 = 0.0470; w_0 = 3.0705$$

The standard deviation between the predicted result and the labeled result was 0.719, and the error was within the acceptable range. Then the algorithm was applied to the intelligent search engine system of formula.

Besides, other retrieval functionalities also were implemented, including:

Table 1 Part of the experimental results of parameter calculation

Formula ID	Formula	Decoction piece ID	Decoction piece	Annotated weight	X ₁	X ₂	X ₃
2	Daqinjia-o Powder	19,153	Gentiana macroph-ylla	4	0.00064414	1	0.066
6	Damangcao Pow-der	19,381	shikimic	4	0.00052605	1	0.0523
14	Datong Pills	17,578	cinnabar	4	0.00228797	0	0.0224
14	Datong Pills	19,988	clove	3	0.00022269	0	0.0266
14	Datong Pills	17,878	lead powder	3	0.00022269	0	0.038
15	Datong Pills	17,591	dendrob	2	0.00028504	0	0.098
15	Datong Pills	19,972	ginseng	4	0.00248455	0	0.0349
18	Datong Pills	17,546	tatarian aster	3	0.00487358	0	0.1042
19	Datong Pills	17,820	niter	3	0.00156434	0	0.0529

- (1) Full-text retrieval: link to the index base according to the search terms and realize the global retrieval by the keywords.
- (2) Precise retrieval: by different semantic types of search terms to achieve precise retrieval including:
 - (a) by decoction pieces
 - (b) by Chinese crude drug
 - (c) by the creation time of formula
 - (d) by the department of formula
 - (e) by the classification of the formula efficacy
 - (f) by the nature, flavor, and meridian tropism of the formula
 - (g) combination of full text and precise retrieval: by keywords and semantic entries

Results

AMFormulaS was developed based on B/S architecture, Java language, and MySQL5.7, composed by modules of information retrieval of formulas, decoction pieces and Chinese crude drug, statistical analysis, and visualization (the home page of the retrieval system is shown in Fig. 4). On the search results page, users can not only browse the specific information of formulas, the related information of decoction pieces and decoction pieces combination, but also the global statistical information of decoction pieces and formulas in the whole database. Users can search relevant information according to their needs and select the appropriate presentation pages, such as formula retrieve, decoction pieces retrieve, decoction pieces combination retrieve, and Dashboard.

Formula retrieval

By the entered name of the formula, the information of formulas will be displayed on the page including an ID



Fig. 4 The home page of AMFormulaS

苏子煎

基本信息	
ID	43924
组成	紫苏子、杏仁、蜂蜜、生姜汁、生地黄汁
方剂药性	温、平
方剂药味	辛、甘、苦
方剂归经	脾、肺、胃、大肠
方剂功效	润肺、止咳、平喘、润肠
所属科室	
来源	外台秘要

组成			
饮片	剂量	剂量分级	剂量单位
紫苏子	2升		朝代：唐 单位：体积
杏仁	2升		朝代：唐 单位：体积
蜂蜜	2升		朝代：唐 单位：体积
生姜汁	2升		朝代：唐 单位：体积
生地黄汁	2升		朝代：唐 单位：体积

文本信息	
别名	
文本来源	《外台》卷十引《深师方》。
药物组成	苏子2升，生姜（汁）2升，白蜜2升，生地黄（汁）2升，杏仁2升。
加减	
功效	
主治	上气咳嗽。
制备方法	上捣苏子，以地黄、姜汁浇之，绢绞取汁，更捣，以汁浇复绞，如此6-7过，令味尽，去滓，熬杏仁令黄黑，捣令如脂；又以向汁浇之，绢绞取汁，往来6-7过，令味尽，去滓，纳蜜和，置铜器中，于重汤中煎之，令如饴，煎成。
用法用量	每服方寸匕，日3次，夜1次。
用药禁忌	忌芜荑。
临床应用	
药理作用	
各家论述	《千金方衍义》：咳嗽日久而客邪未除，蕴化为热，故用姜汁散邪，地黄滋热，且制入苏子顺气味中，更以杏仁熬黑，如大陷胸丸中法，专涤胸中宿垢，纳蜜熬煎，如饴，缓行搜逐而无骤伤中气之患。
备注	

Fig. 5 The retrieval results of Suzi Decoction

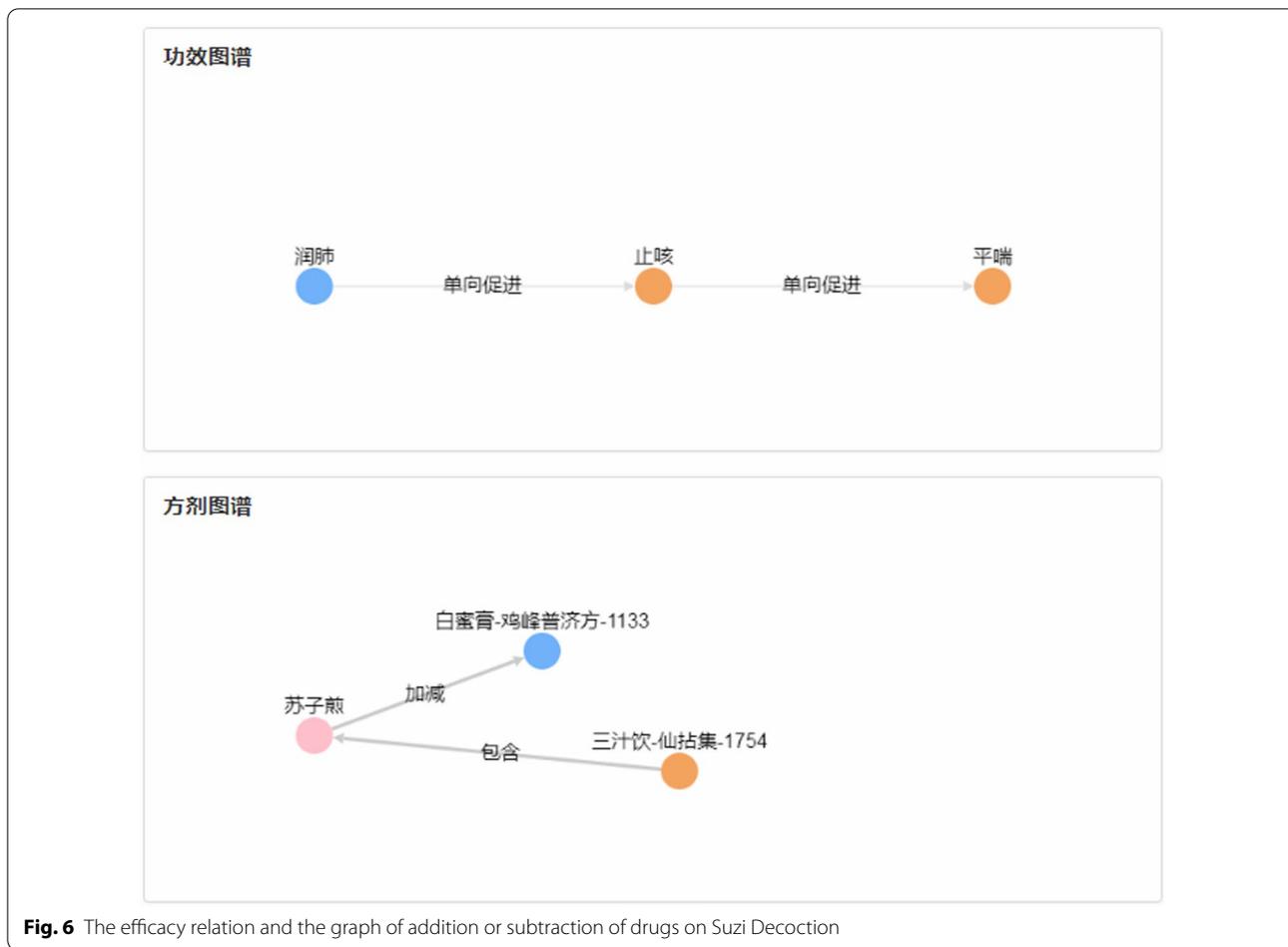


Fig. 6 The efficacy relation and the graph of addition or subtraction of drugs on Suzi Decoction

of the formula, composition, efficacy, nature, flavor and channel tropism, department, source, formation time as well as the original text information of the formula. A formula is made up of decoction pieces, the addition or subtraction of drugs lead to continuous changes of formula, like name, efficacy. Take Suzi Decoction for example, the retrieval results are shown in Fig. 5. The efficacy relation and graph about the addition or subtraction of composition drugs are shown in Fig. 6.

Decoction pieces retrieval

By the inputted name of a decoction piece, the system will search and return the basic information, time distribution of formulas containing the decoction piece, and the use frequency about the diverse dosage of the decoction piece. Take ginseng for example, the retrieval results are shown in Fig. 7.

Retrieval of decoction pieces combination

The system supports the query of the combination of decoction pieces. By the entered name of decoction pieces, the system can display the basic information of the decoction pieces combination, like clinical application, indications, action classification, efficacy, compatibility of the combination, etc. (as shown in Fig. 8).

Figure 9 shows rich information about relations between the formulas and the combination of ginseng (人参 in Chinese) and largehead atractylodes rhizome (白朮 in Chinese). Such as: (1) ginseng and largehead atractylodes rhizome both appear in 3,047 formulas; (2) ginseng, largehead atractylodes rhizome and Indian bread (茯苓 in Chinese) appeared in 1,665 formulas, and (3) ginseng, largehead atractylodes rhizome and dried tangerine peel (陈皮 in Chinese) appeared in 1,143 formulas.

Besides, the retrieved formulas containing these decoction pieces can be looked up and sorted according to the importance of the combinations of these decoction pieces in formulas. For instance, by inputting “ginseng” and “largehead atractylodes rhizome”, there are 3,047



formulas that can be retrieved. After intelligent sorting, the first formula shown to users is "Renshenbaizhu soup" (as shown in Fig. 10). In the same light, when searching "ginseng", "largehead atractylodes rhizome" and "Indian bread", the first formula is "Sanwu soup" (as shown in Fig. 11).

Dashboard

The dashboard shows the statistics about all the Chinese medicine decoction pieces and formulas in the database, such as the top 10 decoction pieces and efficacy of formulas appeared in this database, the statistics about nature,

flavor and channel tropism of formula and the number of formulas formed in every dynasty (as shown in Fig. 12).

Discussion

AMFormulas aims at sorting formula information and providing intelligent retrieval services for medical staff, researchers, and students, at the same time, providing data support for the generation of class formulas, screening of classic formulas, data mining of formulas, as well as new drug research and development. Based on integrating the formula data, the system makes a multi-dimensional statistical analysis of the formation time

人参-白术

基本属性	
临床应用	人参：治劳伤虚损，食少，倦怠，反胃吐食，大便滑泄，虚咳喘促，自汗暴脱，惊悸，健忘，眩晕头痛，阳痿，尿频，消渴，妇女崩漏，小儿慢惊，及久虚不复，一切气血津液不足之证。白术：治脾胃气弱，不思饮食，倦怠少气，虚胀，泄泻，痰饮，水肿，黄疸，湿痹，小便不利，头晕，自汗，胎气不安。
主治	1、久病虚弱。2、胎动不安。习惯流产。
作用分类	补气药
功效	人参：大补元气，固脱生津，安神。白术：补脾，益胃，燥湿，和中。
各家论述	《本草新编》：“人参，宜同诸药共用，始易成功。如健脾也，必加茯苓、白术。”
归经	人参：归脾、肺经。白术：归脾、胃经。
性味	人参：甘微苦、温。白术：苦、甘、温。
注意事项	
用法用量	人参：6-15g。白术：6-15g。
药对出处	人参配白术，见于《太平惠民和剂局方》四君子汤。
配伍机制	人参甘温，大补元气，尤善补脏腑之气。白术苦温，健脾燥湿，尤善补脾气。二药合用，互补优势，补气健脾。培元安胎功力更强。

Fig. 8 The basic information about the combination of ginseng and largehead atractylodes rhizome

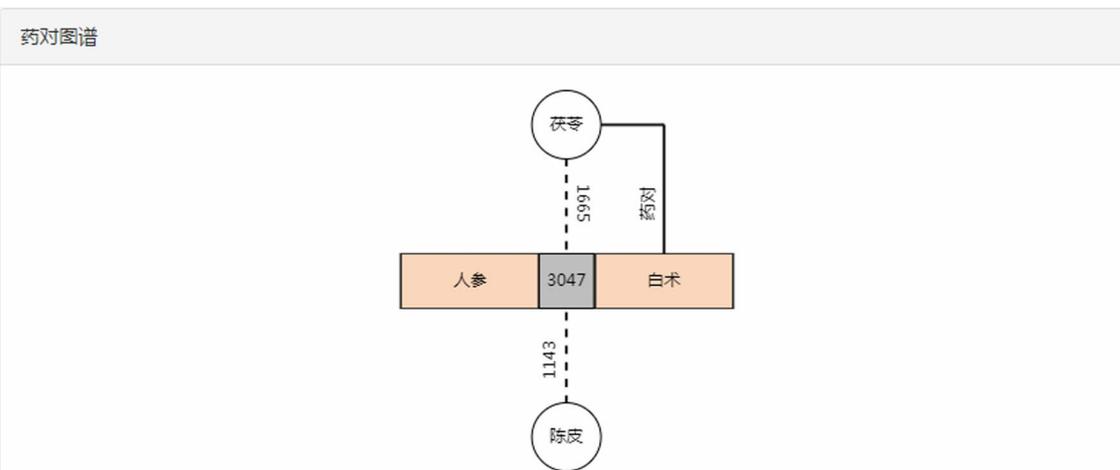


Fig. 9 The compatibility and frequency of ginseng and largehead atractylodes rhizome

of formula, medication frequency, dosage of formulas and drugs in the past dynasties. The data and analytical results about the time, medication habits, dosage of traditional Chinese medicine to help enlighten many new

research directions. Meanwhile, the system can provide comprehensive and accurate intelligent query services for patent application and protection.



Fig. 10 The retrieval results about the combination of ginseng and largehead atractylodes rhizome



Fig. 11 The retrieval results about the combination of ginseng, largehead atractylodes rhizome

Considering the knowledge of formulas involves a wide scope and enormous quantity, more formulas need be included in the future. The current version of

AMFormulaS only aims at verification for system design and retrieval algorithm. As the scale of database and users growing, more tests and updates on performance

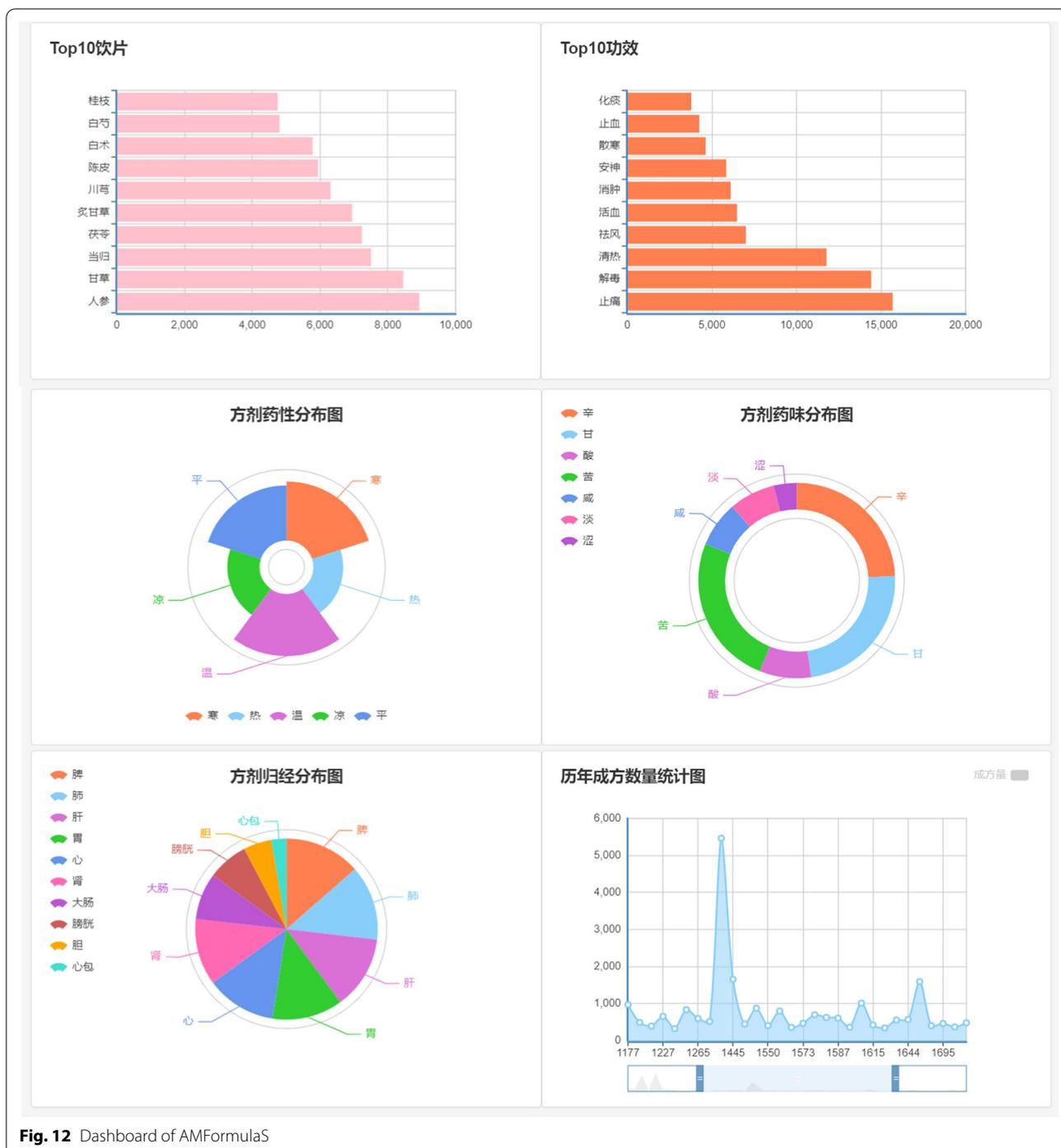


Fig. 12 Dashboard of AMFormulaS

will carried out to meet users' needs of more accurate retrieval engine, high-quality data, and other services.

Conclusions

In this study, a total of 38,000 formulas were structured and standardized through information extraction methods, then imported into the structured formula database.

A novel intelligent formula retrieval system, AMFormulas, was built capable of multi-dimensional retrieval, and statistical analysis of formula information. The system collected, standardized, and integrated a large amount of formula information, including the original text of formulas. It not only realizes efficient retrieval and statistical analysis but also enables the users to access the original data source.

Abbreviations

AMFormulaS: Ancient and Modern Formula System; TCM: Traditional Chinese medicine.

Acknowledgements

We thank Mingzhe Li for help with critical proofreading of the manuscript.

About this supplement

This article has been published as part of BMC Medical Informatics and Decision Making Volume 21, Supplement 2 2021: Health Big Data and Artificial Intelligence. The full contents of the supplement are available at <https://bmcmmedinformdecismak.biomedcentral.com/articles/supplements/volume-21-supplement-2>

Authors' contributions

YZ designed this study, YC and BG processed the data, LL and JL reviewed the data. Meanwhile, YC and BG wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by National Key R&D Program of China (2019YFC1710400; 2019YFC1710401). The work was also partially supported by the Fundamental Research Funds for the Central public welfare research institutes (ZZ13-YQ-126; ZZ13-YQ-127) and Beijing Natural Science Foundation (7174328). The publication charges of this study come from Fundamental Research Funds for the Central public welfare research institutes (ZZ13-YQ-127).

Availability of data and materials

The data that supporting the findings of this study are available from the corresponding author on request.

Ethics approval and consent to participate

Not applicable.

Consent to publication

Not applicable.

Competing interests

The authors declare that there are no competing interests.

Author details

¹ Zhejiang Academy of Traditional Chinese Medicine, Hangzhou 310007, Zhejiang, China. ² Tongde Hospital of Zhejiang Province, Hangzhou 310007, Zhejiang, China. ³ Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China.

Received: 27 January 2021 Accepted: 2 February 2021

Published: 30 July 2021

References

- Pei M, Duan X, Pei X, et al. Research on compatibility chemistry of acid-alkaline pair medicines in formulas of traditional Chinese medicine. *Zhongguo Zhong yao za zhi = Zhongguo zhongyao zazhi = China Journal of Chinese Materia Medica*, 2009, 34(15):1989–93.
- Sun B. Study on the properties theory and compatibility law of the mild-nature traditional Chinese medicine. Jinan: Shandong University of traditional Chinese Medicine; 2010.
- Yu-Hang LI. Discussion on the syndrome-factors and the formula-factors. *China J Tradit Chin Med Pharm*. 2009;24(02):117–21.
- Wang J, Wang Y, Yang G. Methods and modes about the theory of traditional Chinese prescription composition. *China J Chin Mater Medi*. 2005;7:9–12.
- Liu S. Study on the historical Track of clinical dosage in Dacheng Qi Decoction. Beijing: Beijing University of Chinese Medicine; 2016.
- Song Y, Fu Y. A preliminary study on the dosage of medicines in Li Dongyuan's prescriptions. *J Tradit Chin Med*. 2011;62:64–81.
- Ai N. Malignant disease syndromes the literature of traditional Chinese medicine research. Harbin: Heilongjiang University Of Chinese Medicine; 2017.
- Leung WK, Wu JCY, Liang SM, et al. Treatment of diarrhea-predominant irritable bowel syndrome with traditional chinese herbal medicine: a randomized placebo-controlled trial. *Am J Gastroenterol*. 2006;101(7):1574–80.
- Iwasaki K, Kato S, Monma Y, et al. A pilot study of banxia houpu tang, a traditional Chinese medicine, for reducing pneumonia risk in older adults with dementia. *J Am Geriatr Soc*. 2008;55(12):2035–40.
- Guo W. Research and implementation of knowledge mapping of Traditional Chinese Medicine Prescription. Lanzhou: Lanzhou University; 2019.
- He M, Yan X, Zhou J, et al. Traditional Chinese medicine database and application on the web. *J Chem Inf Comput*. 2001;32(2):273–7.
- Shen D. Chinese patent drug database construction and prescription rule research. Beijing: Chinese Academy of Traditional Chinese Medicine; 2014.
- Xue R, Fang Z, Zhang M, Yi Z, Wen C, Shi T. TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic acids research*, 2013, 41(Database issue).
- Fu L, Li S, Li M, et al. Discussion on the standard of word segmentation of ancient Chinese medicine books: taking the medical books of Qing dynasty as an example. *China J Chin Mater Med*. 2018;33(10):454–9.
- Yan J, Gu Z, et al. Traditional Chinese medicine philology. Beijing: China Press of Traditional Chinese Medicine; 2002.
- Zhou Z, Tang D. Traditional Chinese pharmacology. Beijing: China Press of Traditional Chinese Medicine; 2016.
- Wu L, et al. Chinese traditional medicine and materia medica subject headings (last volume). Beijing: Traditional Chinese Medicine Classics Press; 2008. p. 01.
- National standard of the people's Republic of China. Chinese information processing vocabulary Part 01: basic terms(GB12200-1-90). Beijing: Standards Press of China; 1991.
- National standard of the people's Republic of China. GB/T13715–92. Modern Chinese word segmentation standard for information processing. Beijing: Standards Press of China, 1992
- Li S, Li M, Xu Y, et al. Capsules based Chinese word segmentation for ancient Chinese medical books. *IEEE Access*. 2018;6:70874–83.
- Liu L, Zhu Y, Li H, et al. Building of Traditional Chinese Medicine Integrative Data Model(TCMIDM). *China Dig Med*. 2015;10(10):70–2.
- Liu L, Liu J, Jia L, et al. Study of concept description of Chinese medicine ontology. *China Dig Med*. 2016;11(2):90–2.
- Liu L, Zhu Y. Building Chinese medicine conceptual data model based on semantic representation. *World Chin Med*. 2017;12(4):936–9.
- Chinese Pharmacopoeia Commission. Pharmacopoeia of People's Republic of China: Part one. Beijing: China Medical Science Press; 2010.
- General Administration of quality supervision, inspection and Quarantine of the people's Republic of China, Standardization Administration of China. Coding Rules and Codes of Traditional Chinese Medicine :GB/T 31774–2015. Beijing: Standards Press of China, 2015.
- Editorial board of Chinese materia medica of National Administration of traditional Chinese Medicine. *Chinese Materia Medica*. Shanghai: Shanghai Scientific & Technical Publishers, 1999.
- Nanjing Medical University. Dictionary of Traditional Chinese Medicine. SHANGHAI RENMIN CHUBANSHE, 1977.
- CN109801697A_ An evaluation method of the importance of Chinese Herbal Pieces[EB/OL]. <https://zhuanli.tianyancha.com/80a3d65404855cb9ee3ec8fd244b2dd2>, 2020-4-29.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.