

AGILE DEVELOPMENT DRIVEN BY DATA MODEL – THE BEST PRACTICE OF TCM LITERATURE INFORMATION SYSTEM

XUEBO LI¹, YANG ZHOU¹, SHUJUN MA² AND ZHENGUO WANG¹

¹Information Management School
Shandong University of Traditional Chinese Medicine
No. 465, Daxue Road, Changqing District, Jinan 250355, P. R. China
lixuebo@sducm.edu.cn

²Department of Human Resource
Shandong Liming Polytechnic Vocational College
No. 389, Jiwei Road, Jinan 250116, P. R. China
Mashujun001@163.com

Received February 2016; accepted May 2016

ABSTRACT. *A large number of Chinese ancient books provide important guidance for clinical practice. Traditional Chinese Medicine (TCM) information system can help us to speed up the transformation of TCM knowledge from the books to the clinical practice. Traditional waterfall model driven by system function needs the whole documents of development at the beginning. If the users make any change in the period of development, it will bring so many extra work, so this kind of development model cannot fit the development without clear demand. Agile development pattern driven by data model can help us construct function models rapidly under the condition that the function demands cannot be defined exactly. Through continuous iteration, we can complete the construction of literature information system with high efficiency and quality.*

Keywords: Traditional Chinese medicine, Information system, Agile development

1. Introduction. Traditional Chinese Medicine (TCM) literature is the most valuable achievement in the development of TCM, which supports the decisions in clinical practice. The development of information technology has greatly enriched the ways of people acquiring information. The traditional literature information recorded in paper leads to poor using efficiency in scientific and clinical research. In order to give more accurate information of TCM literature efficiently, we should build the TCM literature information service system. A key question to develop information system with high efficiency and quality is choosing a scientific development pattern according to the characteristics and functions requirement of the information system [1].

People have already made some practices in the information system building of traditional Chinese medicine literature. Wu et al. of Shanghai University of TCM have made an information system to search and analyze the TCM cases [2]. Zhang et al. of Shandong University of TCM have built a database based on Visual FoxPro 6.0 platform, and the institute of information on TCM of CACMS has developed a database system to search information of TCM literature content [3]. Almost all the study and development of TCM literature information system focus on the functions of division and description, and use the development method such as waterfall model driven by system functions and corresponding documents to build system, and even no paper tries to find a system development model especially for the characteristics of TCM.

Waterfall model is one of the most reliable and common models in the development of information systems, which is characterized by the documents as the basis. The use of documents can help developers control every link effectively, and ensure the stability and accuracy of development process. As the development pattern of the traditional “business

process” system, we need specify function of the business process needed in the development and analysis [4]. Waterfall model driven by function development pattern is suitable for the following development scenarios: when the development of the software is about the fields that have been defined clearly; when there are reusable architectures or models; when the system needs to be very safe and have extreme performance requirements. Otherwise, in the information system development of TCM literature, the TCM literature information has the characteristics of “large amount of data”, “unstructured data” and “complex data”. In the early period of system construction, the functional requirements are difficult to be defined. It is impossible to get the full functions of the system before this system runs. If the development of the literature information system is carried out using the functionally driven waterfall pattern, the following questions will appear. (1) It is not easy for developer to go back a step and if something wrong happened in the design phase, things can get very complicated in the implementation phase. (2) People are not very clear of what they exactly want from this system. Any changes on this information system may cause a lot of confusion. (3) Small changes or errors that arise in the completed system may cause a lot of problems. (4) Until the final stage of the development cycle is complete, a working model of the software does not lie in the hands of the client. Thus, he is hardly in a position to inform the developers, if what has been designed is not exactly what he had asked for.

2. Development Driven by Data Model. The core of TCM literature information system is the literature data, and its main function is to provide data retrieval and analysis services, so it is one of data intensive applications [5]. That is, the function of the system is determined by the use of data from the system users. Therefore, the first step in the construction of the system should be a complete and accurate data analysis. This paper used the object-oriented data analysis method. Firstly, we used the concept of object to find the entities that exist in literature of TCM, and built the whole collection of the entities, such as document entity, Chinese medicine entity, formula entity, medicine record entity, medical case entity, and medical note entity. Secondly, we analyzed the attributes of the entities. These attributes information should be only related to specified entity, and they are objective characters of the entity, such as that the document entity should contain documents name, author, book’s age and other information, that formula entity should include name, source, usage, efficacy, indications, use of taboo and other attributes, and that Chinese medicine entity should contain medicine name, processing, dosage, and usage note. Finally, we analyzed the attributes of one entity related to other entities, and found out the relationship between entities, such as that the formula entity contains more than one Chinese medicine, and that a document entity contains a plurality of medicine entities.

Figure 1 shows the entities, attributes and relationships of the Chinese medicine, medicine record, document and formula. All the database tables about TCM information are divided into three main parts: “CHINESE_MEDICINE”, “DOCUMENT” and “FORMULA”. In the “CHINESE_MEDICINE” part, the core entity (table) is CHINESE_MEDI-CINE, the ID of which is the foreign key of the entity (table) CHINESE_MEDICINE_PHARMACOLOGY, CHINESE_MEDICINE.RECORD and CHINESE MEDICINE COMPONENT. In the “FORMULA” part, the core entity (table) is FORMULA, the ID of which is the foreign key of the entity (table) FORMULA_ILLUSTRATION, FORMULA_PHARMACOLOGY, and FORMULA_COMPONENT. The part of “DOCUMENT” has one entity named DOCUMENT, which is where the Chinese medicine record and formula exist. The CHINESE_MEDICINE part and the FORMULA part are connected with the entity (table) FORMULA.CONSTITUTE. According to Figure 1, we found which data the system users are interested in and which query they will made, and made the definition of complete functional requirements based on the data model.

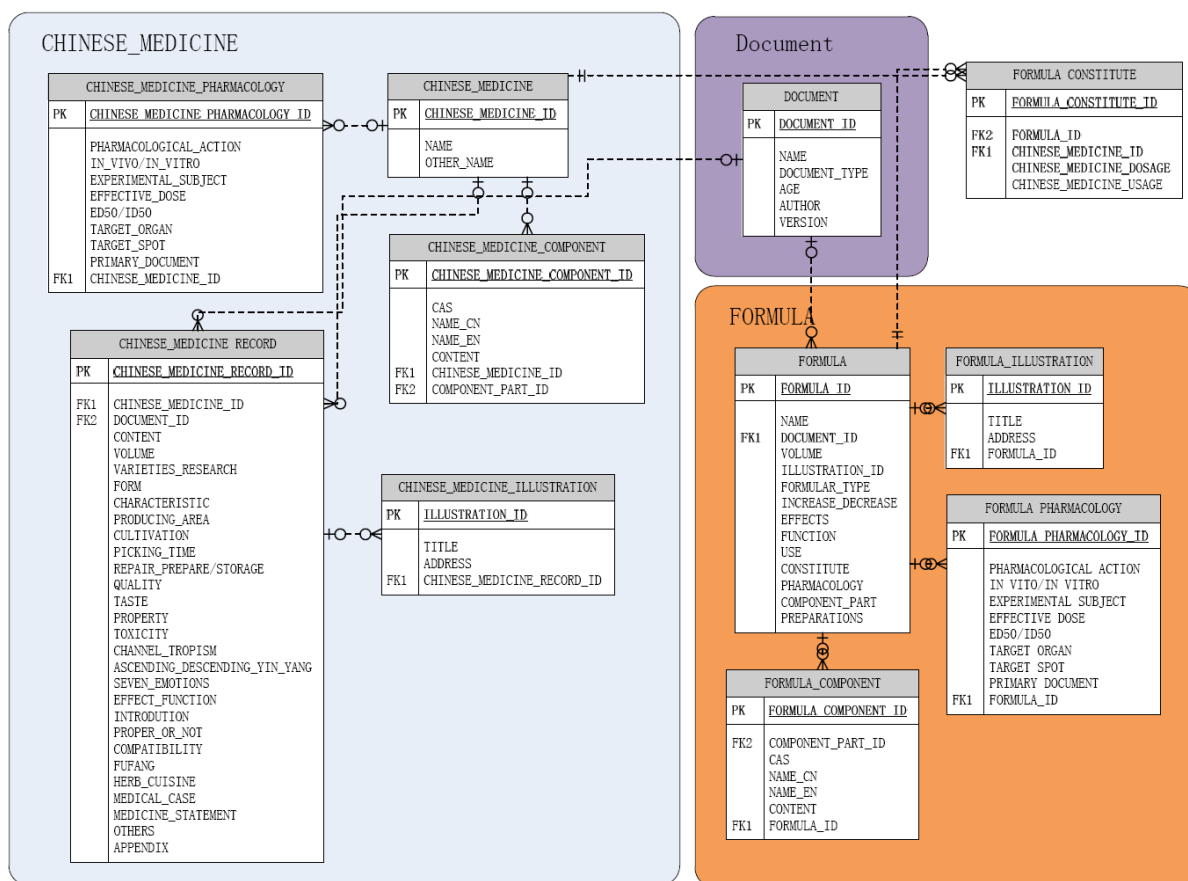


FIGURE 1. Data model

3. Agile Development. Agile development is a development pattern, which focuses on the developers, the iteration and the steps. It is adaptive not predictive, people-oriented not process-oriented [6]. In contrast with the waterfall pattern, the agile development pattern meets the development requirements of the system well [7]. The function from analyzing of data model is the main user demand rather than all, and is uncertain, which is called the “rough demand”. Therefore, in making system development planning and demand analysis, we should not over pursue the integrity and specification of documents, only need to make a simple report after analyzing the “rough demand”. According to the contents of the report we collected typical data, analyzed the relationship between the data attributes and built data E-R model. Based on the definition of the model, we had collected more data in a wide range. At this time, due to the large workload, this work often needs to be completed in the multi terminal workstations.

The practice shows that the use of Microsoft Excel as data collection tool is more convenient than the use of data collection software specially designed. It should maintain a certain degree of redundancy, so as to ensure the integrity of the original data. When data collection is completed, the data collection is formed, and we had made the definition of system function based on the definition of the data [8]. We used agile development language ROR (Ruby On Rails) to build system prototype, imported the test data into the prototype to iterate the development and made the system perfectly step by step. Finally, we used Microsoft VBA programming interface to import all the data into the completed system from Excel.

Figure 2 shows the entire flows of agile development. In every step of the flow, we used the corresponding tools to do the right work. In the step of rough demand collection and data structure definition, we used E-R model to express the relationships of the entities. In the step of source data collection and system function definition, we used MS Excel and

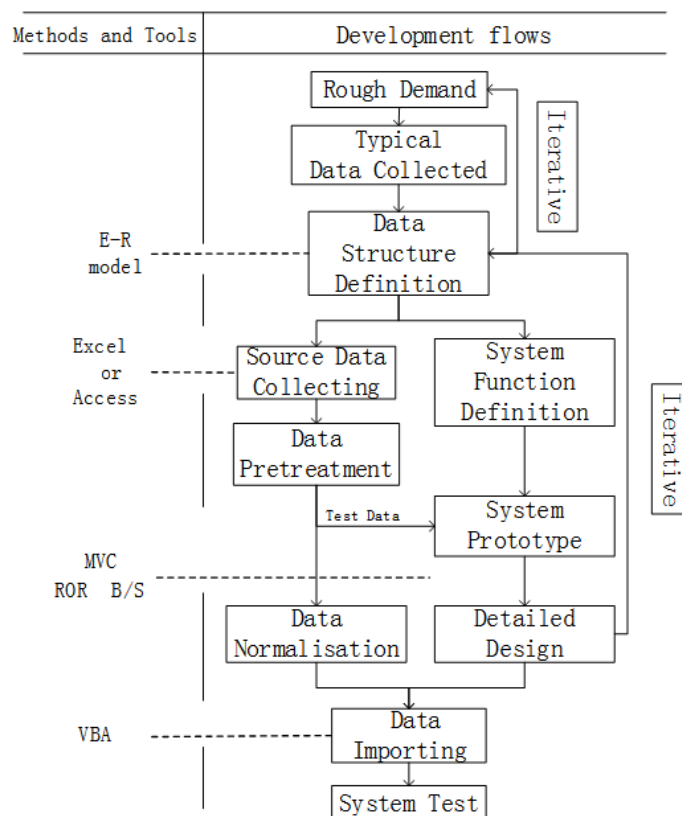


FIGURE 2. Agile development flows

Access to complete the data pretreatment. In the step of system construction (building prototype, detailing design), we used agile development tools of ROR and MVC model to finish this work. MS VBA was used to import the data into the information system. During this process, we used iterative to perfect the system until we finished it.

4. Conclusions. Through the research about TCM literature information system construction in this paper, we built the main models of TCM literature information system, and illustrated the properties of the data models such as document, Chinese medicine, and formula. We defined the related entities and their main properties, and defined the relationships among the entities.

At the beginning of system on TCM treatment of hypertension development, we used waterfall model to build system. Firstly, we want to collect the complete user's demand and finish the demand analysis document, but the user has much difficulty in making the complete system demand, and even they do not know what the finished system looks like. The analyzers have to take about 30 man-months workload to finish the document in this dilemma, and then the programmer took about 80 man-months to complete the system. Otherwise, the user was not satisfied with our work, and he believed that what we completed was not what he wants. So, we turned back to start this work with agile model. We took 3 man-months workload to complete the prototype system without any document, and then communicated with the user completely. We only took 8 man-months workload to finish the first iteration, so we repeated this process over and over until the user thought that what we made together is what he wants. All above iterations took only 40 man-months.

Compared with the previous system development work using traditional waterfall model, using data driven agile model can complete prototype system in the short time.

Continuous communication with the system users, and iterations of the system can reduce the discussion time for the system function demand and rebuild work to ensure

the quality of the whole system. The construction of this agile development method driven by data model can provide reference for the future TCM information system.

With the development of the concept and application of big data, the analysis and data mining of mass TCM literature data become more and more important. In the future research, we should try to build the TCM data model and storage method based on distributed computing, find the deep knowledge and rules among different structured TCM data, and provide the guidance to the clinical research and practice.

Acknowledgment. This work is partially supported by the Taishan Scholar's Research Group of Wang Zhenguo. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] X. Li, Y. Song, X. Fu and Z. Wang, Agile construction of information service system on traditional Chinese medicine treatment of hypertension, *World Science and Technology/Modernization of Traditional Chinese Medicine*, vol.17, no.6, pp.135-137, 2015.
- [2] Z. Wu et al., Introduction of analysis system about Chinese medicine cases, *Shanghai Journal of Traditional Chinese Medicine*, vol.37, no.3, pp.54-56, 2003.
- [3] Q. Zhang et al., Building and analysis of Chinese traditional medicine cases, *Journal of Shandong University of Traditional Chinese Medicine*, vol.29, no.4, pp.298-299, 2005.
- [4] P. Yang, Y. Zhou, D. Guo and Z. Piao, Analysis and comparison of processing models in software engineering sector, *Journal of Jixi University*, vol.8, no.1, pp.86-88, 2008.
- [5] Z. Wang and Y. Liu, Ideas and methods for the construction of ancient scientific and technological literature – Ancient Chinese medicine literature development and utilization, *Information and Document Service*, no.2, pp.32-34, 2005.
- [6] M. Zhang, Analysis and research on the principle and process of agile development, *Information Technology*, no.2, pp.197-198, 2013.
- [7] Y. Liang, The difference between agile development and other traditional development methods, *Electronic World*, no.24, 2012.
- [8] W. Li, Data-driven programming in software development, *Computer Programming Skills & Maintenance*, 2012.