Research on master data recognition of TCM based on analytic hierarchy process

Yuhui Sang¹, Shuanggui Tian¹

¹Hubei University of Chinese Medicine, Wuhan, China

Corresponding author and e-mail: Shuanggui Tian, sunnytian610@hbtcm.edu.cn

Abstract. With the rapid development of information technology in the field of medical and health, the requirement for information sharing was higher and higher. TCM plays an irreplaceable role in promoting health development. The information of TCM was represented by the construction, application and data sharing of information system, and faced complicated problems. Data environment, standardized data management requirements and data standards were very important. In the field of TCM, the main line was about “TCM information integration standardization, TCM standards into information”, while the main data was taken as the basis of data standardization. This paper through the investigation and summary of the existing information system and data items, according to theoretical research of main data recognition by scholars, discussed the main data identification method by using analytic hierarchy process, and extracted the main data, so as to improve quality of data management and optimize data management environment.

1. Introduction

The construction of TCM hospital in China is still in the development stage. The construction of hospital information system is an important part of hospital information construction. There are some new problems in the construction and application of hospital information system. The non-standard use of data directly affects information exchange and sharing within the hospital and between different hospitals. As the basis of resource sharing in TCM hospitals, standardization has been studied by more and more scholars. It is expected to formulate perfect field norms and provide industry standards as soon as possible, so as to avoid situation of feudal separatism and solve problem of difficult unified management. Master data (MD), as a kind of slowly changing data, is a key business term that can describe core business entities of an enterprise and share across systems [1]. The identification of master data in TCM hospital is the main content of standard making, and it is also a complex project. It is difficult to accurately identify master data only by personal experience and subjective judgment. Based on the existing information resources of TCM hospitals, comprehensively use statistics, mathematics and other methods to extract information suitable for master data. On the basis of inheriting existing research results, carry out exploratory scientific research, seek common ground while reserving differences, and constantly innovate.
2. Research contents and methods

2.1. Research contents
In this study, the information system of a tertiary TCM hospital and data involved in system are selected as research objects, including data characteristics such as field code, field type, whether it is primary key, field description, application system name, etc. Based on extraction principle of master data, the impact indicators are analyzed and the corresponding weights are provided to determine master data of TCM hospital and study impact of master data on TCM. This paper discusses importance of hospital domain standardization, and prospects development trend of TCM hospital master data identification method, summarizes shortcomings of existing scientific research and corresponding solutions.

2.2. Research methods
This study uses theory and method of qualitative analysis and quantitative analysis, closely combines opinions from experts in the field, analyzes and determines the main data by means of AHP and Delphi methods.

3. Research steps

3.1. Conceptual model construction of master data identification for TCM hospital

3.1.1. Data topic domain partitioning. The division of subject domain must integrate understanding of business personnel and system designers. Too rough and too detailed division is not conducive to data governance of hospital information system. This research is based on the TCM hospital information system currently used in Guang'anmen Hospital. The number of system is increasing and the business work is constantly refined, including 46 systems, such as electronic medical record subsystem and clinical pathway subsystem. In this study, the TCM hospital information subsystem is coded, as shown in Table 1, and the data used in each subsystem are sorted out. At the same time, combined with expert guidance, the process and business units involved are classified. It is concluded that the hospital information system can be divided into six major subject areas: participants, resources, medical treatment, classification, finance, teaching and research.

| Order | System name                                           | Coding | Order | System name                                           | Coding |
|-------|-------------------------------------------------------|--------|-------|-------------------------------------------------------|--------|
| 1     | Chinese medicine electronic medical record subsystem  | C101   | 24    | Outpatient (emergency) triage system                  | H203   |
| 2     | Outpatient (emergency) doctor workstation subsystem   | C102   | 25    | Outpatient (emergency) diagnosis pricing system       | H204   |
| 3     | Resident workstation subsystem                         | C103   | 26    | Inpatient management subsystem                        | H205   |
| 4     | Experience inheritance subsystem of famous veteran TCM| C104   | 27    | Inpatient charge subsystem                             | H206   |
| 5     | Surgical and anesthetic management subsystems          | C105   | 28    | Cost accounting and performance management subsystem  | H207   |
| 6     | TCM auxiliary diagnosis and treatment subsystem        | C106   | 29    | Material management subsystem                         | H301   |
| 7     | Health examination management subsystem                | C107   | 30    | Equipment management subsystem                        | H302   |
3.1.2. Data conceptual model construction. This research is based on the information system of Guang'anmen Hospital, Chinese Academy of TCM. The system has perfect functions, smooth business connection between different subsystems, and covers many fields of hospital work. Therefore, when analyzing data conceptual model, we can refer to the business process of the existing system, analyze requirements, combine with the understanding of the information system, collect and summarize system process, abstract and summarize real things, and produce a conceptual model independent of the computer system and DBMS. The global conceptual model is drawn based on the topic domain, as shown in Figure 1.

![Figure 1. Global conceptual model.](image-url)
3.1.3. Data identification entity determination. Take medical subject field as further research object. See Figure 2 for relationship between subject field, subject and entity. In the medical subject domain, the classification of subjects and entities is shown in Table 2 below:

![Figure 2. Relationship between the subject field, subject and entity.](image)

**Table 2. The subject and entity comparison relationship in the medical subject domain.**

| Subject                          | Entity                                               |
|---------------------------------|------------------------------------------------------|
| Doctor-patient information      | Doctor, Patient, Doctor-patient relationship          |
| Medical service                 | Medical items, Medical sites, Medical expenses, Medical operations |
| Medical product                 | Doctor's advice, Medical documents                   |

3.2. Identification principle of master data in TCM hospital

Taking the basic concept of master data as standard and referring to the existing research results at home and abroad, the identification principles of master data are summarized. With continuous promotion and application of master data management (MDM), more and more software manufacturers begin to engage in master data related work, and their understanding and definition of master data are not the same. Based on the definition of master data by various companies and referring to the existing relevant literature, this study describes master data with "slow changing data is the key business term that can be shared across systems". Combined with expert investigation method, this paper collects interpretation and evaluation of master data from experts, and concludes that the following indicators should be referred to when identifying the master data: data sharing degree, business demand degree, long life cycle, unique identification, independent existence, inseparability, stability and data privacy.

3.3. Construction of master data identification index in TCM hospital

When identifying master data of hospital information system, based on the identification principle determined in the previous step, it is necessary to further evaluate the weight of identification principle in the determination of master data. In the selection of method to determine weight, priority is given to the analytic hierarchy process and precedence chart. Because there are eight indexes to determine weight, the number is small, and the importance of individual indexes is equal, the proportion scale of AHP is more accurate than priority diagram method, which uses "0" and "1" to distinguish, and it can quantify the qualitative indexes in experts' minds more accurately. This academic research adopts the method of expert investigation. Experts are asked to evaluate eight principles: data sharing degree, business demand degree, long life cycle, unique identification, independent existence, indivisibility, stability and data privacy. They are assigned as $x_1 \sim x_8$ in turn, and compared according to the relative importance, make judgments. Experts' judgments on these principles $(a_{ij})$ are expressed by numerical values. The judgment matrix is obtained by summarizing the experts' opinions.
The 1 ~ 9 scale of Saaty is introduced to assign relative importance of each evaluation principle [3]. In this study, experts are invited to make pairwise comparison, which can reduce number of comparison on the lack of subjectivity of expert evaluation, minimize impact of individual judgment errors on overall ranking, and avoid systematic judgment errors. The weight of expert is \( \mathbf{f}_i, (i = 1, 2, ..., n) \). The score is \( x_1 \), which gives high weight to experts with academic influence and experience in this field. The calculation formula is as follows:

\[
G_n = \sum_{i=1}^{n} \prod_{j=1}^{n} x_i^f = \sum_{i=1}^{n} f_i \sqrt[2]{x_1 x_2 f_2 ... x_n f_n}
\]

(2)

The expert evaluation matrix is obtained as follows:

\[
A = \begin{bmatrix}
1 & 1/2 & 2 & 1/2 & 4 & 1/3 & 1/3 & 3 \\
2 & 1 & 2 & 3 & 6 & 3 & 4 & 6 \\
1/2 & 1/2 & 1 & 1/3 & 3 & 1 & 1/2 & 3 \\
2 & 1/3 & 3 & 1 & 1/5 & 1/2 & 1/4 & 6 \\
1/3 & 1/6 & 1/2 & 1/3 & 1 & 1/2 & 1 & 2 \\
4 & 1/3 & 1 & 2 & 2 & 1 & 3 & 5 \\
1/4 & 1/4 & 1/3 & 4 & 1 & 1/6 & 1 & 2 \\
1/3 & 1/6 & 1/3 & 1/6 & 1 & 1/5 & 1/2 & 1
\end{bmatrix}
\]

(3)

The eigenvalue method is used to calculate weight vector, eigenvectors (\( \mathbf{\omega} \)) and the largest eigenvalue (\( \tau \)) are obtained by precise calculation as follows:

\[
\mathbf{\omega} = (0.110, 0.295, 0.098, 0.126, 0.056, 0.177, 0.103, 0.036), \tau = 8.74
\]

(4)

CI=0.106 was calculated. The order of recognition principle matrix is eight, and the corresponding CI critical value is 0.169 [4], so it has good consistency. The weight indexes of correspondence recognition principle in the master data of traditional Chinese medicine hospitals are shown in Table 3 below:

**Table 3. Recognition principle weight index.**

| Identification principle          | Weight index | Identification principle          | Weight index |
|----------------------------------|--------------|----------------------------------|--------------|
| Data sharing degree              | 0.110        | Business demand                  | 0.295        |
| Long life cycle                  | 0.098        | Unique identification            | 0.126        |
| Independent existence            | 0.056        | Inseparability                   | 0.177        |
| Stability                        | 0.103        | Data privacy                     | 0.036        |

High demand is an important indicator of master data identification. With the continuous improvement of hospital information system, the business process is gradually specific, and cross system phenomenon of data is gradually highlighted, involving medical work, financial management, hospital warehouse, teaching and other systems. At this time, the data is widely needed, which is also the master data that we need to manage. Only by properly and effectively managing master data, can we achieve goal of master data identification. To optimize data environment of hospital information system and reflect value of data. Secondly, the recognition principle of weight is inseparability. The business behavior and basic management of hospital are composed of small business entities. These business entities must be minimum and inseparable in order to flexibly combine business processes and improve management efficiency.
3.4. Validation of master data identification results in TCM hospital

Combined with the main data recognition entity, main data recognition principle and main data recognition index weight obtained in previous steps, main data recognition score template of TCM hospital is constructed. The main data recognition principle is used as column vector of scoring template, and the main data recognition entity is used as the row vector of scoring template. If there is an association between recognition entity and recognition principle, it is assigned 1, otherwise it is assigned 0. Combined with the weight of main data recognition index, each recognition principle is multiplied by corresponding weight to score entities in the hospital information system. And through soliciting experts' opinions, set the score threshold as 0.75 [5], when the master data recognition score is greater than the threshold, it can be managed as master data. Taking the business entities in medical subject field as an example, the score of each entity item is calculated by using score template, as shown in Table 4 below:

| Table 4. Entity score. |
|------------------------|
| Entity | Score | Entity | Score |
|--------|-------|--------|-------|
| Doctor | 0.924 | Patient | 0.918 |
| Doctor-patient relationship | 0.245 | Medical project | 0.863 |
| Medical parts | 0.577 | Medical expense | 0.512 |
| Medical operation | 0.134 | Medical advice | 0.102 |
| Medical bills | 0.279 | |

The top three entities of master data identification were doctors, patients and medical project, and the scores were all higher than the set score threshold of 0.75. In 46 hospital information systems, the doctor's job number exists in 46 systems, the patient's certificate number exists in 45 systems, and the medical item number exists in 43 systems. The detailed distribution is shown in Table 5. The identified master data accords with concept, principle, and method is also suitable for identification process of master data in the other five subject domains.

| Table 5. Main entity application system. |
|----------------------------------------|
| Entity | Primary key | Application system | Number of systems |
|--------|-------------|---------------------|------------------|
| Doctor | Doctor ID | C101,C102,C103,C104,C105,C106,C107,C108,C201,C202,C203,C204,C205 | 46 |
|        |            | C301,C302,C303,C304,C401,C402,C403,H101,H201,H202,H203,H204,H205,H206,H207,H301,H302,H303,H304,H305,H306,H401,H402,H403,H404,H405,H406,H407,H408,H409,H410,H411,H412 |
| Doctor-patient relationship | Patient ID | C101,C102,C103,C104,C105,C106,C107,C108,C201,C202,C203,C204,C205 | 45 |
|        |            | C301,C302,C303,C304,C401,C402,C403,H201,H202,H203,H204,H205,H206,H207,H301,H302,H303,H304,H305,H306,H401,H402,H403,H404,H405,H406,H407,H408,H409,H410,H411,H413 |
| Medical parts | Medical project number | C101,C102,C103,C104,C105,C106,C107,C108,C201,C202,C203,C204,C205 | 43 |
|        |            | C301,C302,C303,C304,C401,C402,C403,H101,H201,H202,H203,H204,H205,H206,H207,H301,H302,H303,H304,H305,H306,H401,H402,H403,H404,H405,H406,H407,H408,H409,H410,H411,H414 |
| Medical operation | Doctor ID | C101,C102,C103,C104,C105,C106,C107,C108,C201,C202,C203,C204,C205 | 46 |
|        |            | C301,C302,C303,C304,C401,C402,C403,H101,H201,H202,H203,H204,H205,H206,H207,H301,H302,H303,H304,H305,H306,H401,H402,H403,H404,H405,H406,H407,H408,H409,H410,H411,H412 |
4. Conclusions
Because of its complex business process and a large number of updated data, it was very important to introduce standardized information system in TCM hospital. The data environment was the most important part of governance. This study used mathematical and statistical methods to identify the main data of the information system of a third-class TCM hospital. Combined with characteristics of data environment, the paper put forward corresponding suggestions. The recognition principle and identification index served as reference significance for other hospital information system master data identification, at the same time, it can improve work quality of data maintenance personnel, and help eliminate data barriers between business departments and information systems in the process of hospital information, so as to facilitate data sharing and improve value of data service. The process of identifying master data based on AHP relied on cooperation of expert experience and business personnel. However, when selecting indicators, this process could be affected by subjective opinions of experts. The current academic research still used qualitative and quantitative methods to avoid influence of expert subjectivity on the results as far as possible, so as to make results more accurate and verified. The semi-automatic recognition can be realized by software to improve recognition efficiency and facilitate promotion of recognition methods in other fields. In the process of identification, attention should be paid to the semantic correlation of data between hospitals and various systems. Based on semantic and corresponding statistical model, the importance of main data was evaluated, and then the main data of TCM hospital was determined.

References
[1] Zhang Wei. Research on ontology based enterprise master data management [D]. Shanghai Jiaotong University. 2011
[2] Long Wenjia. Analysis of the application of computer database system in information management [D]. Zhixing College, Hubei University. 2017
[3] Miao Shuanghu. Research on performance evaluation of secondary departments of Xinxiang Medical College Based on analytic hierarchy process [J]. Journal of Xinxiang Medical College. 2012, 29 (6): 475-477
[4] Research on the construction of statistical index system [J]. Li Moxuan. Management observation. 2017 (09)
[5] Jiang Wenchun, Wen Xianxiu, Gu Fengjiao. Application of Delphi method and analytic hierarchy process to construct quality evaluation standard system of safe drug use management in ward [J]. China health management, 2019 (03): 194-197