Research on Regional Medical Heterogeneous Data Integration Technology

Chunfei Zhang1*, Yuchun Zhang1 and Yunna Duan1
1 Public Computer Education and Research Center, Jilin University, Changchun, 130012, P.R.China
*cfzhang@jlu.edu.cn

Abstract. The integration of regional medical heterogeneous information is a problem that must be solved in the process of modern medical digitization. Based on the analysis of the characteristics of the existing medical information system, combined with the existing data integration and exchange technology, this paper proposes a heterogeneous data integration scheme for regional medical care. It also focuses on the data processing in the process of regional medical data integration and exchange, and describes in detail the process of medical data extraction, medical data conversion and medical data loading. The goal is to integrate existing heterogeneous medical information systems and eliminate information barriers to improve the effective utilization of medical information.

1. Introduction
Heterogeneous information is a very common phenomenon in today’s society. Among medical institutions, there is also a heterogeneous information phenomenon in medical information. To solve heterogeneous information, it is necessary to share information resources and interactive integration of information resources. The sharing of resources is a prerequisite. There are two main situations for sharing information resources. The first is to use a unified information management and storage system. It is also required that each subsystem must use the same system operation method. The disadvantage of this approach is that the cost is too high and difficult to achieve. The second is to build an information resource sharing platform and use the platform as an intermediary. This approach also facilitates the integration process of resources. The integration of information resources must first realize the information sharing of data among medical institutions, that is, horizontal integration. On the basis of this sharing, the important level information is distinguished, and the more stable data are used as the core data to process the integration, that is, vertical integration. In order to achieve interoperability of information between medical systems, it is necessary to achieve two-way integration of medical heterogeneous information.

Under the existing conditions, there are two main ways of data interaction between medical institution systems. One is to adopt a point-to-point approach, that is, to develop a specific interface for different systems that need to exchange data, as shown in Figure 1. The second is to use an integrated approach to achieve data interaction between medical information systems by building a medical data exchange platform, as shown in Figure 2.
It can be seen from the figure that the application of data interaction in point-to-point mode has a mesh structure. The advantage of this method is that it is relatively simple to implement and does not need to change the existing system, but the application needs to be used for data interaction. The more you come, the more the number of interfaces you need to develop increases exponentially and the entire system becomes more complex. The integration method is not easy to implement, but once implemented, it will greatly improve the scalability of the system and greatly reduce the coupling of the system. When an application changes, it will not affect other applications in the system.

In summary, the regional medical heterogeneous data integration scheme adopted in this paper is an integrated model for integrating medical information systems by constructing an information platform. The aim is to maximize the preservation of the original medical information subsystem and to achieve the integration and sharing of medical information.

2. Data processing in regional medical data integration and exchange
The regional medical data integration and exchange system has many businesses and complex functions, and it is also the key to regional medical heterogeneous data integration. The following focuses on data extraction, data conversion and data loading functions in regional medical data integration and exchange.
Data processing is used to extract, transform, and load heterogeneous data, and process the source data into standardized data that conforms to regional medical data exchange standards. The flow of data processing is shown in Figure 3.

![Data processing flow diagram](image)

Figure 3. Data processing flow

2.1. **Data extraction**
The data extraction first loads the extracted interface class according to the configuration information, processes the source data line by line or item by item, and then calls the corresponding conversion interface class to check and process the extracted information, and finally calls the loading interface class to summarize the data and convert it into actual needs. Types are such as files or databases.

The data extraction process is shown in Figure 4.

![Data extraction diagram](image)

Figure 4. Data extraction diagram

2.2. **Data conversion**
After the data extraction is completed, the data is cleaned and converted. The main contents are as follows.
- Data supplementation. That is to say, the data is filled in for the empty data and missing data and the unprocessable mark is made.
- Data replacement. Replace the invalid data with valid data.
- Format normalization. Convert the data format extracted from the source data into a target data format that is easy to enter the warehouse.
• Primary and foreign key constraint. By creating a primary foreign key constraint, the illegal data is replaced or the error file is reprocessed.
• Data merge. Usually it implemented by table association, large-small table association with lookup, large-large table intersection with join.
• Data split. Data was split according to certain rules.
The data conversion process is shown in Figure 5.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig5.png}
\caption{Data conversion diagram}
\end{figure}

2.3. Data loading
After the above-mentioned cleaning and conversion, the original data becomes a data conforming to the standard, and data loading can be performed. For the cleaned data, you can choose the incremental load mode or the full load mode. The data loading process is shown in Figure 6.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig6.png}
\caption{Data loading diagram}
\end{figure}
3. Conclusion
The purpose of regional medical heterogeneous data integration is to eliminate medical information barriers, to maximize the foundation of the original medical information subsystem, and to integrate and share medical information. This paper focuses on the data processing process in the process of regional medical heterogeneous data integration and exchange, hoping to provide some reference for similar research.

Acknowledgments
This paper was supported by Science and Technology Project of Department of Education in Jilin under Grant No. JJKH20170806KJ.

References
[1] Chen, X., Liu, C. (2018) Research Progress on Basic Data Sets of Electronic Medical Record Information of Health Information Platform in Traditional Chinese Medicine Clinic (TCM Pavilion). Medical Information, 31: 7-9.
[2] Sheng, Z.H., Xia, K.J. (2018) Design of Medical-Data Interaction Platform Based on Cloud Services. China Medical Devices, 11: 132-137.
[3] Zhang, Y. (2018) Design of large-scale medical information sharing platform under cloud computing. Modern Electronics Technique, 41:179-182.
[4] Zhao, L. (2018) Design and implementation of regional medical information platform based on cloud computing. Electronic Technology & Software Engineering. 9: 151-152.
[5] Shi, S.Z., Cao, W.L. (2016) Construction and application of the regional medical information platform. China Medial Equipment. 9: 104-108.
[6] Xiao, M., Chen, P. (2017) Development of the Chongqin regional medical imaging information and big data analysis platform. China biotechnology, 37: 28-36.