

A Storage Retrieval Method of Medical Imaging Based on XML Description

Lei Xia^{1, a}, Liyong Wan^{2, b}

¹Information Technology and Training Center, Nanchang Institute of science & Technology, Nanchang, 330108, China;

²School of Information, Nanchang Institute of science & Technology, Nanchang, 330108, China.

^acsuzju@163.com, ^bwanliyong@163.com

Keywords: XML; storage retrieval; medical image; non-structure data.

Abstract. With the development of medical image technology, the medical image has played a more and more important role on medicine. There is more and more the number of medical images in hospital. How to manage a large numbers of medical images will be one of the major problems for many hospitals. This paper proposes a storage retrieval method of medical image based on XML description. A storage retrieval system for supporting XML description has been implemented, and some corresponding experiments are tested.

1. Introduction

The medical image can show clearly and accurately the internal structure of human body. And the clinician can has knowledge of the patients' condition with these medical images. Therefore, it has an irreplaceable role on the diagnosis and treatment. The medical images were stored with the film forms. The storage form is suitable for storing, copy and transfer. With the further development of the modern medical image and image-processing technology, Dwyer from Kansas University proposed the concept of the medical images through network in the 1980s [1]. In the years since then, digital images resource can be interchanged and queried, and forms the Digital Imaging and Communication in Medicine (DICOM) [2]. The DICOM has ensured standard interface from imaging equipment to system in medical system, and entered a stage of all-around popularization. In recent years, the number of medical images has become more and more large [3]. The University Hospital of Geneva had produced 12,000 medical images as early as 2002. For large and medium-sized hospitals, the medical images of about 1 to 6 TB will be produced each year in our country. All hospitals will be produced imaging data of about 2PB in our country. According to predicting of IBM, medical imaging data will be occupied 30% of the national storage in the next 5 to 10 years [4, 5]. So how to query valuable information from these medical imaging data will emerge a key challenge for medical science worker and .researcher.

At the moment, the retrieval methods of medical imaging have three ways. The first way is to traditional retrieval way based on keyword. Tamura proposed the imaging retrieval technology based on artificial comment text, and summarized overall the research. The second way is content-based image retrieval, which acquired the information from images through some visual features of color, texture, shape and area [6]. The third way is medical imaging retrieval based on semantic, which can solve the semantic gap question between the visual feature and man's understands. However, these retrieval methods have some problems, such as low retrieval precision, single retrieval and poor expression, etc. For these problems, this paper proposes a storage retrieval of medical imaging, the storage retrieval system for supporting XML description has been implemented, and some corresponding experiments are tested.

2. The Retrieval Method of Medical Imaging Based on XML

2.1 The Storage Model of Medical Image Based on XML.

There is no uniform structure standard for the medical imaging data of CT, PET, MRI and MU, which is stored with raw format in the content. The computer can't understand and deal with them directly. Therefore, for computer, these data information is no valuable. In order to deal with different type of medical imaging data and facilitate data processing, this paper implements the management and operation for medical imaging data content with descriptive information. This paper thinks that medical imaging data includes metadata, key thumbnail image and raw data. They are become three factors of medical imaging storage. Among them, XML metadata includes some descriptions of basic attribute, semantic feature and bottom feature of medical imaging; The key thumbnail image represents visual expression of medical imaging content; The raw data represents medical imaging data of the raw file. This paper proposes a storage model of the three factors is shown in Fig. 1. The storage model of the three factors in this paper is described the medical imaging data from above three factors.

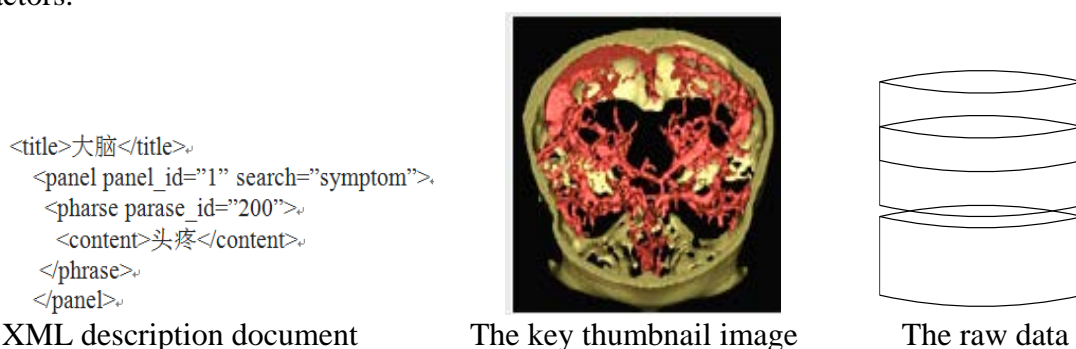


Fig.1 Storage model of three factors

Definition 1. The three factors of storage model data. The data structure of storage model is composed of the only ID, XML metadata, key thumbnail image and raw data. The description of the four tuples is: Trifactors= (ID, Metadata, KeyThumbnail, Rawdata), where ID is primary key of the storage model of database, which only identifying combination of the three factors; The Metadata is metadata description of medical imaging based on XML, which includes basic attribute, semantic feature and bottom feature of medical imaging; The KeyThumbnail is a general introduction to key content of raw medical imaging, which is the image format of JPEG. The Rowdata represents the raw data, and each node represents a raw data of medical imaging. The data structure of three factors model can describe and identify roundly the raw medical imaging, so as to provide the theoretical basis of generous medical imaging data.

2.2 The Retrieval Technology of Medical Image Based on XML.

The XML metadata is queried with XQuery technology in the three factors model. The XQuery is a kind of query language of describing XML data source, which has the characteristic of accuracy, flexibility and usability. Since there is one to one mapping relationship between XML metadata and raw medical imaging file. When finding out the corresponding XML document with XQuery technology and the storage retrieval model can locates the position of the raw medical imaging data, so as to implement the query function of the raw medical imaging.

In order to implement the XML document query, the XQuery parser needs to be constructed firstly. This paper implements XQuery parser, its role is to extract XML document and Xpath expression by matching XQuery statements. For XQuery expression statements, how to identify and extract XML document and Xpath expression are key factors, and the problem need to be resolved. During the parsing processing, the XQuery statements need to be analyzed with lexical analysis and semantic analysis. When XQuery parser is constructed, the function of XML query operation can be implemented.

3. Study of Storage Retrieval System of Medical Imaging Based on XML Description

3.1 The System Design.

The retrieval system mainly includes two modules of the medical imaging storage and medical imaging retrieval. The storage model includes description stage of three factors, that is to say, the data persistence stage. The description stage of three factors include three steps of XML description of the raw medical imaging, the key thumbnail image of extracting medical imaging and classification file up. The XML query manager is a bridge linking the XML query engine and storage model of three factors, which implements the data exchange for the XML query engine and storage model of three factors.

The XML query engine includes also three parts of XQuery parse, XML query optimization and XML parallel query. In data persistence stage, we store the open source XML document with Native XML database of eXist. For the key thumbnail image and medical imaging data, we store them with open source Nosql database of Sones GraphDB. The raw medical imaging information needs to semantic processing: firstly, the medical imaging information need be deal with XML description; next, the key thumbnail image needs be extracted from information of XML document description; finally, these key thumbnail images will be stored by classifying. In addition, XML data will be stored XML database, and the key thumbnail images will be stored the thumbnail image database. The retrieval is XML retrieval based on XQuery query, which includes three query stages of XQuery parse, XML query optimization and XML parallel query.

The retrieval of this paper is based on description, query and retrieval of XML raw information, then the XML management becomes the core of the retrieval system. So we add a XML meta information manager into the storage retrieval module, so as to manage the XML document of system. The XML information manager can implement data exchange between medical imaging of storage model and XML document data of XML query engine.

The overall framework of the retrieval system is shown in Fig. 2.

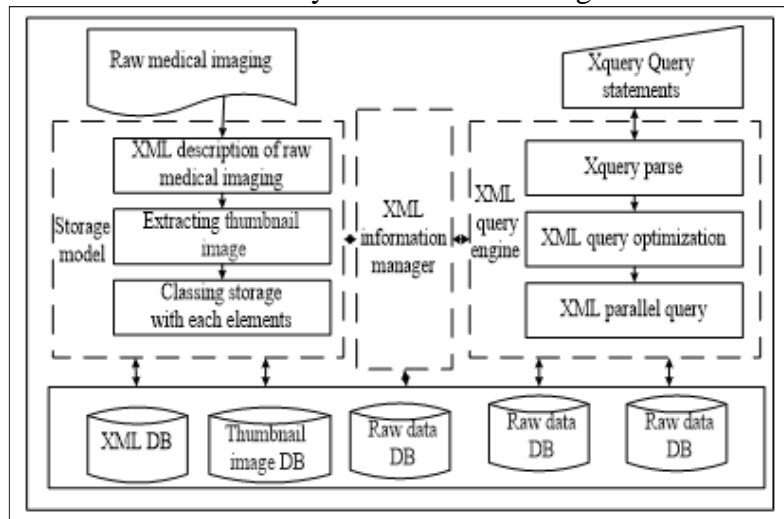


Fig.2 The overall framework of the retrieval system

3.2 The Experimental Analysis.

We implement the storage retrieval system of medical imaging based on XML description. Some corresponding medical XML document is imported into the storage retrieval system. The size of XML dataset is 5M to 12M. The experimental result data is shown in table 1.

The experimental result shows that if the size of XML document data is no more than 20M, and then the XQuery query can acquire retrieval result from corresponding XML document data and takes less time. While if the size of XML data is more than 40M, and then it takes about 1s to a acquire retrieval result from corresponding XML data.

Table.1 Experimental result

Size of XML document	Size of medical imaging	The consuming time (s)	The precision correlation (%)
5M	32G	5	90
10M	60G	8	88
20M	8G	18	85
40M	240 G	30	81
60M	400G	58	79
80M	600G	70	75
120M	855G	76	70

4. Summary

This paper proposes a storage retrieval method of medical image based on XML description. A model for three factors is constructed in the storage retrieval method, which can implement the description of the medical imaging. According to semantic description of the medical imaging, the storage retrieval system can implement the query of medical imaging. Finally, we design and develop a storage retrieval prototype system of medical imaging, which can supports XML description. In addition, the storage retrieval prototype system is tested in various conditions. The experimental result shows the storage retrieval method is feasible and valid. But for the moment, this storage retrieval system for query speed and precision need to further improvement. Therefore, the next step in our research is to focus on optimizing the working of query performance.

References

- [1]. LI H Y. The Time Model Optimization of Trained Text-to-speech System. *Journal of Computer and Application*, Vol. 1(2012).No 1:14-15.
- [2]. LI W. Digital Imaging and Communications in Medicine (DICOM).National Electrical Manufacturers Association, Vol.2 (2007) No 3:1321-1323.
- [3]. HENNING Muller, Nicolas Michoux, David Bandon, et al. A Review of Content-based on Image Retrieval Systems in Medical Applications-clinical Benefit and Future Directions. *International Journal of Medical Informatics*, Vol.7 (2009) No 1:11-16.
- [4]. Haralick R.M. Shanmugam K, Dinstein. Textural Features for Image Classification. *Systems, Man and Cybernetics IEEE Transaction on*. Vol. 3(2010) No 6: 610-613.
- [5]. Simeulders A.W.M., Worrying M., Santini S., et al. Content-based Imaging Retrieval at the End of the Early Years. *Journal of Pattern Analysis and Machine Intelligence, IEEE Transaction on*. Vol. 22(2002) No.12:1349-1353.
- [6]. Jones K.S, Walker S., Robertson S.E. A Probabilistic Model of Information Retrieval: Development and Comparative Experiments. *Information Processing and Management*, Vol. 36(2000) No.6:809-840.