



# Information Retrieval Technique in Medical Imaging Technology

Azrulhizam Shapi'i<sup>#</sup>, Riza Sulaiman<sup>#</sup>, Mohammad Khatim Hasan<sup>#</sup>, Abdul Yazid Mohd Kassim<sup>\*</sup>

<sup>#</sup> *School of Information Technology, Faculty of Information Science and Technology*

*Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia*

*Tel.: +60389216180, E-mail: azrul@ftsm.ukm.my*

<sup>\*</sup>*Department of Orthopaedics and Traumatology, Universiti Kebangsaan Malaysia Medical Centre*

*Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia*

*E-mail: dryazidk@gmail.com*

**Abstract**— The medical field at present is an example for appropriate innovation in the information technology sector where it can bring benefits to physicians and patients and also the medical tools to prevent disease. Among the examples of innovation in information technology are in the field of medical imaging technology, laser technology, nanotechnology and others. As we know, medical imaging technology is very important because it can help medical experts in conducting research on patients. A medical image is known as DICOM. DICOM format is unique, it is not only contain the image of the patient, but it also contains information related to the patient. This paper will show the development of software to retrieve the information from the digital x-ray image (DICOM). Software developed using Netbeans and library used was pixelmed library. This software is then tested with a number of patient data. Results showed that the software is able to find the information contained in the patient x-ray images accurately.

**Keywords**— DICOM, x-ray, retrieve, information, software

## I. INTRODUCTION

DICOM stands for *Digital and Communications in Medicine*. DICOM is a standard used for storage, printing and transfer of information in medical imaging [1]. It includes the definition of file format and network communication protocol. Communication protocol is an application protocol that uses the concept of TCP / IP to communicate between systems. DICOM files can be transferred between two entities that can receive images and patient data in digital format. National Electrical Manufacturers Association( NEMA) has the copyright to this standard.

DICOM allows the integration of scanners, servers, workstations, printers and network hardware from various manufacturers into photograph and communication system [2]. DICOM has been widely used in hospitals throughout Malaysia. DICOM data format is different from others because it compiles all the information together into one set. For example, an X-Ray usually contains the patient ID in it. An object usually contains attributes include items such as patient's name, date of x-ray taken and the most important attribute is the pixel image data. Pixel data is very important

because it will be used to compare the size of the digital x-ray with the actual x-ray [3].

DICOM files contain not only the X-ray images in digital form, but it also has data such as patient name, type of image, pixel size and x-ray [4]. Each DICOM file has a header that stores the information. The DICOM standard allows the integration of scanners, printers, workstations, servers and hardware networking of different manufacturers. Therefore all the information contains in a DICOM file to must be translated correctly.

## II. OBJECTIVES

The main objective of this study is to build software that can read or retrieve the information contain in medical images (DICOM). The most important part of the DICOM is the image header as it contained important information such as patient's name and date of birth [5]. The internal structure of DICOM files should be examined more closely so that the software is able to read DICOM images with accurate information.

### III. RESEARCH BACKGROUND

In response to the increasing usage of digital images in the medical field, two organizations, ACR (American College of Radiology) and NEMA formed a committee in 1983 to draft a standard format that can be used for the medical images storage and transfer processes. The Committee has first issued an ACR-NEMA standard in 1985. In 1993, the standard was revised and renamed as DICOM. Recent changes been made to DICOM version 3.0 has enabled the transfer of medical images within multi-supplier environment [1].

DICOM Committee is responsible for developing a library of documents that can explain the translation process of data image between imaging devices of the same or different types. An example of the data transfer is between the Computed Tomography Scans and the Picture Archiving and Communication System (PACS).

Development of Digital Image and Communication in Medicine (DICOM) is done by the National Electrical Manufacturers Association (NEMA) to develop a library in which contains documents that describe the data related to transfer images between two different or similar devices as images taken from MRI or CT-scan. According to NEMA (2006), there are five main things that need to be focused when applying DICOM:

- Transfer and accuracy of the complete object (images, documents, wave).
- Rotation and access on objects.
- Smoothness of an action (such as printing of images on film).
- Management flow (to support the work list and information status).
- Quality and consistency of image display/appearance (for print and display).

### IV. METHOD

To develop software that can read and translate the information from the DICOM image, the file structure must be considered carefully. In the DICOM Standard, there are 16 important parts and each of these sections has specific functions (NEMA, 2006). The 16 parts are as follows:

- Part 1-Introduction and Overview
- Part 2-Conformance
- Part 3-Information Object Definitions
- Part 4-Service Class Specifications
- Part 5- Data structures & Semantics
- Part 6-Data Dictionary
- Part 7-Message Exchange
- Part 8-Network Communication Support for Message Exchange
- Part 9-Retired (formerly Point-to-Point Communication Support for Message Exchange)
- Part 10, Media Storage and File Format for Media Interchange
- Part 11, Media Storage Application Profiles

- Part 12, Media Formats and Physical Media for Media Interchange
- Part 13-Print Management Point-to-Point Communication Support
- Part 14, Grayscale Standard Display Function
- Part 15-Security Profiles
- Part 16-Content Mapping Resource
- Part 17-Explanatory Information
- Part 18- Web Access to DICOM Persistent Objects (WADO)

This standard defines the communication of certain transactions for the imaging and the messages associated with imaging through a network and will be written on the media channels that remain to be displayed. It is important to note that the DICOM standard is not limited to increased space for storage or transfer of the image itself, but the rules of the transfer of information from the text of the imaging files are also included.

#### A. DICOM Image Format

DICOM differs from some, but not some other, data formats in that it groups information into data sets. That means that a file of a chest X-Ray image, for example, actually contains the patient ID within the file, so that the image can never be separated from this information by mistake. This is similar to the way that image formats such as JPEG can also have embedded tags to identify and otherwise describe the image.

A DICOM data object consists of a number of attributes, including items such as name, ID, date, date of birth and also one special attribute containing the image pixel data. A single DICOM object can only contain one attribute containing pixel data. For many modalities, this corresponds to a single image. But note that the attribute may contain multiple "frames", allowing storage of cine loops or other multi-frame data. Another example is NM data, where an NM image by definition is a multi-dimensional multi-frame image. In these cases three- or four-dimensional data can be encapsulated in a single DICOM object. Pixel data can be compressed using a variety of standards, including JPEG, JPEG Lossless and JPEG 2000. Figure 1 below shows the example of DICOM image.



Fig.1 Dicom image

## B. System Design

To develop a good system various steps and processes are involved and one of them is the process of analysis process and system design. To understand the objectives and solving problems, analysis should be done while the design is needed to achieve the objectives or purposes of this study. To determine the objectives of this project, generally it involves several parts:

- Opening of DICOM file
- Displaying DICOM image
- Reading information from the DICOM image

## C. System Development

To develop this software, pixelmed library was chosen as the dictionary for DICOM files. Pixelmed was developed by Dr David A Clunie (2006) using java programming. It is a tool that can read information from DICOM images. This library is chosen because it has a complete DICOM dictionary and easy to use. The java coding below shows how to use the pixelmed library in developing software.

```
com.pixelmed.display.SourceImage import;
com.pixelmed.dicom.AttributeList import; import
com.pixelmed.dicom.Attribute; // This part read image
attributes
com.pixelmed.display.SingleImagePanel import;
com.pixelmed.display.WindowCenterAndWidth import;
```

1) *Reading Information and Attributes from DICOM* This section provides attributes to identify and describe the information of an image. The information may consist of the position and orientation of the image, image pixel and image type and that describe the image data [6]. Examples of modules in this section are General Image Module, Patient Module, Image Pixel Module, and so on. Table 1 shows the modules of the pixel image, while table 2 and table 3 show the general image of the module and patient module.

Table 1 Module for Pixel Image (NEMA,2006)

Attribute Name	Tag	Attribute Description
Samples per Pixel	(0028.0002)	Number of samples (planes) in this image
Photometric Interpretation	(0028,0004)	Specifies the intended interpretation of the pixel
Rows	(0028,0010)	Number of rows in the image.
Columns	(0028,0011)	Number of columns in the image
High Bit	(0028,0102)	Most significant bit for pixel sample data. Each sample shall have the same high bit.
Pixel Representation	(0028,0103)	Data representation of the pixel samples. Each sample shall have the same pixel representation.

Table 2 General image module (NEMA, 2006)

Attribute Name	Tag	Attribute Description
Instance Number	(0020,0013)	A number that identifies this image.
Patient Orientation	(0020,0020)	Patient direction of the rows and columns of the image.
Content Date	(0008,0023)	The date the mage pixel data creation started.
Content Time	(0008,0033)	The time the image pixel data creation started.
Image Type	(0008,0008)	Image identification characteristics.
Acquisition Number	(0020,0012)	A number identifying the single continuous gathering of data over a period of time which resulted in this image.
Acquisition Time	(0008,0032)	The time the acquisition of data that resulted in this image started
Acquisition Datetime	(0008,002A)	The date and time that the acquisition of data that resulted in this image started.

Table 3 Patient module (NEMA,2006)

Attribute Name	Tag	Attribute Description
Patient's Name	(0010,0010)	Name of patient
Patient ID	(0010,0020)	Patient identification
Patient's Birth Date	(0010,0030)	Patient DOB
Patient's Sex	(0010,0040)	Gender
Patient's Weight	(0010,1030)	Weight

## 2) Java Coding

To read the information or attributes from DICOM images, Java coding used are as follows:

```
// Opening DICOM file
AttributeList al = new AttributeList ();
System.out.println
("Name:" + al.get(com.pixelmed.dicom.TagFromName.PatientName));
System.out.println
("Name:" + al.get (com.pixelmed.dicom.TagFromName.PatientSex));
System.out.println
("Pixel:" + al.get (com.pixelmed.dicom.TagFromName.PatientID));
System.out.println
("Pixel:" + al.get (com.pixelmed.dicom.TagFromName.BirthDate));
System.out.println(""+al.get(com.pixelmed.dicom.TagFromName.MediaStorageSOPClassUID));
ij.plugin.DICOM varDicom = new ij.plugin.DICOM ();
varDicom.open (filename);
```

With this software users can choose to read the desired data. Based on the coding section, four patient's image information, *PatientName* (0010,0010), *PatientSex* (0010,0040), *PatientID* (0010,0020) and *PatientBirthDate* (0010, 0030) will be read by software. Figure 2 shows how the user can select the information required when writing a program using Netbeans.

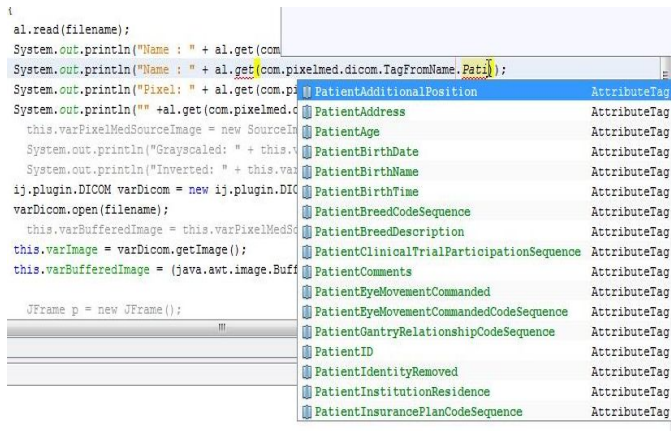


Fig 2 List of attributes in the DICOM image



Fig. 4 Patient DICOM image (PPUKM, 2010)

Figure 3 shows how the software can open and read files in DICOM format (\*. DCM)

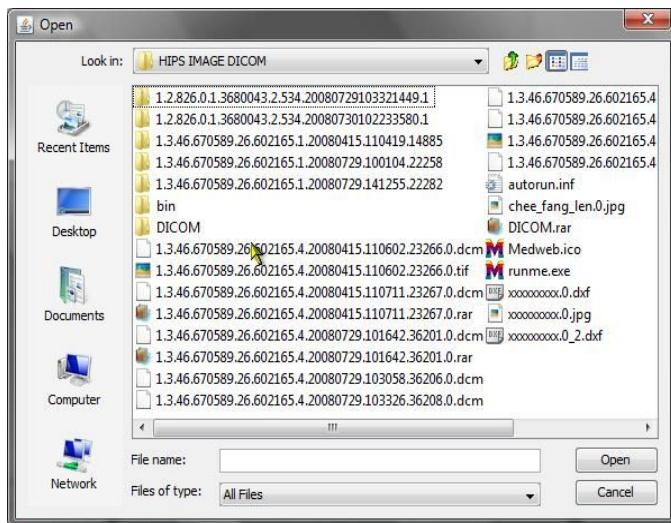


Fig. 3 Open DICOM image files (\*. dcm)

### V. OUTPUT

Figure 4 and figure 5 show the output produced by the DICOM image information retrieval software. Each image contained of hundreds of information. This information is needed because it can help physicians in carrying out their research on the patient's X-Ray. As discussed in the previous section, the software will read four patient's information, *PatientBirthName*, *PatientName*, *PixelSpacing* and *BirthDate*. Some patient data were used to test the accuracy of the software to read DICOM patient information.

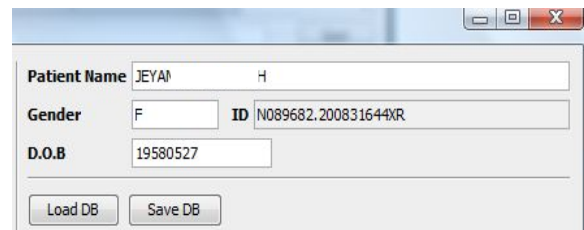


Fig. 5 Patient's information

### VI. CONCLUSION

DICOM (*Digital Imaging and Communication In Medicine*) is a file format that has been used extensively in the medical field. However, its file format is too large and has a variety of ways and definitions in the search for information contained in a DICOM image. DICOM standard allows the integration of scanners, printers, workstations, servers and network hardware from different manufacturers. Therefore, attributes or information in an image should be interpreted or searched and transferred carefully so as not to lose information due to the integration process [1].

A DICOM data object consists of a number of attributes, including items such as patient's name, patient's ID, patient's sex and patient's birth date. By using a number of patient data, the developed software is able to read all the information contained in medical images and is very helpful in the search for patient information quickly and accurately.

#### ACKNOWLEDGMENT

This research project conducted in collaboration with Dr. Abd Yazid Mohd Kassim and Dr. Hamzaini Abd Hamid from the Department of Orthopedics and Traumatology, Medical Centre of Universiti Kebangsaan Malaysia. This department has provided medical image data (DICOM) to be used in this research. This research also funded by University Grant UKM-OUP-ICT-34-171/2009 and UKM-GUP-TMK-07-01-035.

#### REFERENCES

- [1] *Digital Imaging and Communication on Medicine (DICOM)*, National Electrical Manufacturers Association (NEMA) std, 2006.
- [2] R.N.J. Graham, R.W. Perriss, and A.F. Scarsbrook, "DICOM Demystified: A Review of Digital file formats and their use in radiological practice," *Clinical Radiology*, vol.60, pp. 1133-1140, 2005
- [3] M. Klein, "Using Imaging Data in Making Orthopedic Diagnoses," in D. C. Chhieng, & G. P. Siegal, *Advances in Experimental Medicine and Biology*, 2005, paper 2, p. 104-111
- [4] C. Homorodean, D. Olinic, S. Nedevschi, and N. Olinic, "Templates implementation for structured DICOM diagnosis reporting in echocardiography," in *Proceedings of 2005 IEEE International Conference Computers in Cardiology*, September 25-28 2005, Lyon, France, pp. 379 – 382.
- [5] L.W. Kit, "Permodelan Anatomi Manusia Secara Tiga Dimensi Berdasarkan Imej Perubatan DICOM Dalam Persekitaran CAD," Msc thesis, Universiti Kebangsaan Malaysia, Bangi, Malaysia, 2007.
- [6] (2010) The NEMA website. [Online]. Available: <http://www.nema.org/>