

An Integrated Medical Information System using XML

Cheolbum Ahn¹, Yunmook Nah¹, Seunghun Park², and Jungsun Kim³

¹ Department of Computer Engineering, Dankook University, Seoul, Korea
ahn555@dankook.ac.kr, ymah@dankook.ac.kr,

² School of Electronics and Information, Kyung Hee University, Kyunggi-do, Korea
parksh@khu.ac.kr,

³ School of Electrical Engineering and Computer Engineering, Hanyang University,
Kyunggi-do, Korea
jskim@cse.hanyang.ac.kr

Abstract. The advanced medical information systems usually consist of loosely-coupled interaction of independent systems, such as HIS, RIS, and PACS. To support easier information exchange between these systems and between hospitals, and to support new types of medical service such as teleradiology, it becomes essential to integrate separated medical information and allow them to be exchanged and retrieved through internet. This paper presents an integrated medical information system using XML. We analyzed HL7 and DICOM standard formats, and designed XML DTDs for HIS, RIS, PACS, and their integration. We extracted information from HL7 messages and DICOM files, and generated XML document instances and XSL stylesheets based on the proposed XML DTDs. We implemented the web interface for the integrated medical information system, which supports data sharing, information exchange and retrieval between two different standard formats. The proposed XML-based integrated medical information system will contribute to solve the problems of current medical information systems, by enabling integration of separated medical informations and by allowing data exchange and sharing through internet.

1 Introduction

Currently, more and more medical information is being digitalized. For the electronic transfer of X-ray films within hospitals, DICOM(Digital Imaging and Communication in Medicine) format was proposed and commonly utilized in Picture Archiving and Communication Systems(PACS). Patient and prescription information is being represented in electronic records and handled by Hospital Information Systems(HIS). But, current medical information systems are not fully integrated, thus usually consisted of separated subsystems, such as HIS, Radiology Information Systems(RIS), and PACS[1,2]. Moreover, most of medical information systems are incompatible between hospitals. For the exchange of patient information between hospitals, HL7(Health Level 7) standard was proposed. HIS and RIS are usually based on HL7 standard.

Recent advances of web(WWW; World-Wide Web) enable users to access every information through internet. Since its initial announcement in 1998, XML (eXtensible Markup Language) is rapidly becoming a popular data interchange format. There are efforts to represent HL7 and DICOM information in SGML or XML formats, but most of them are in early stages[4,5,6,7].

This paper is an effort to represent various medical information using XML, thus enabling medical information sharing and exchange through internet. The major contribution of this paper is to show how to utilize XML technology to realize more advanced and open medical information systems. The purpose of this paper is: 1) to provide XML DTDs for HIS, RIS, PACS, and their integration, 2) and to show how effectively medical information sharing and retrieval can be realized. By adopting XML technology into medical domain, we can overcome difficulties in medical information integration and compatibility. Also, by utilizing XML query facility, we can deal with more complex queries on integrated medical information.

The remainder of this paper is organized as follows. Section 2 describes background works. Section 3 analyzes HIS, RIS, and PACS formats and presents XML DTDs for each of them and their integration. In section 4, we describe a system realization scheme of an integrated medical information system. Finally, section 5 summarizes the paper.

2 Background

Health Level Seven(HL7) is one of several ANSI-accredited Standards Developing Organization(SDOs) operating in the healthcare arena. "Level Seven" refers to the highest level of the International Standards Organization's (ISO) communications model for Open Systems Interconnection - the application level. HL7 has been actively working with XML technology since the formation of the SGML/XML SIG in 1996. In 1999, HL7 endorsed a recommendation for using XML as an alternative syntax for HL7 V2.3.1 messages[4]. The initial release of Version 3, slated for publication in December 2001, will use only XML encoding. The PRA(Patient Record Architecture) of HL7 V3.0 is written in XML forms based on the "HL7 Document Patient Record Architecture" draft proposed by HL7 SGML/XML SIG in September 1999[6].

Other researches for data exchange between HL7 and DICOM standards are being carried out by HL IMSIG(Image Management SIG) and DICOM WG(Working Group) 20. They also have interests in XML utilization. DICOM WG6 and WG8 studying on DICOM SR(Structured Reporting) are also considering how to utilize XML for DICOM SR[7,10]. But, as we explained earlier, most of these efforts are not finalized yet.

However, it is clear that XML technology will take key roles in representing and exchanging major medical informations for the near future. Figure 1 shows how XML can be utilized as exchange formats between different standards and different hospitals.

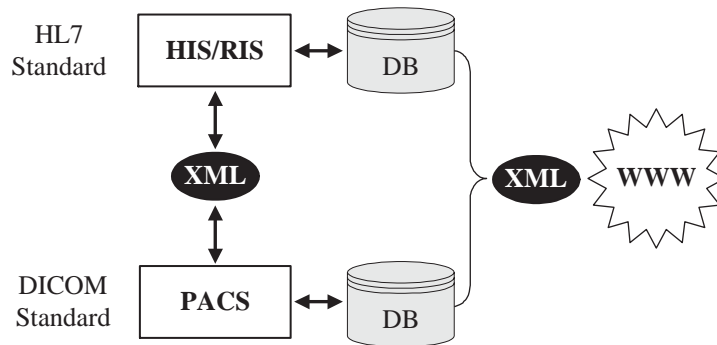


Fig. 1. Conceptual architecture of XML based medical information systems

3 XML DTDs for medical information

3.1 HIS XML DTD

We have used the HL7 Version 2.3 DTD proposed by HL7 SGML/XML SIG[4]. Each HL7 message represents patient-oriented HIS information such as admission, discharge, and transfer. To generate the HIS DTD, we first extract HL7 message table from HL7 standard. All the messages, segments, fields, and data types in this table are then transformed into XML elements according to the appropriate DTD transformation rules. The detail transformation rules are quite tedious, thus omitted from this paper. Figure 2 shows the outline structure of HIS XML DTD and figure 3 shows only some portion of the final HIS XML DTD.

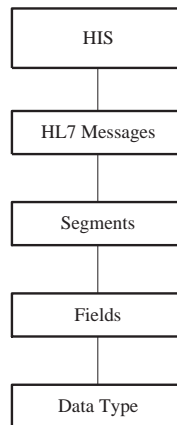


Fig. 2. HIS XML DTD Structure

```

<!ELEMENT HIS
(ADT.A01?,ADT.A04?,ACK?,MPK?,MFN.M01?,ORU.R0
1?,QRY.R02?,ORF.R04?)+>

<!ELEMENT GRP.1.ADT.A01 (NK1)+>

<!ELEMENT GRP.2.ADT.A01 (DB1)+>

<!ELEMENT GRP.3.ADT.A01 (OBX)+>

<!ELEMENT GRP.4.ADT.A01 (AL1)+>

<!ELEMENT GRP.5.ADT.A01 (DGI)+>

<!ELEMENT GRP.6.ADT.A01 (ROL)+>

<!ELEMENT GRP.7.ADT.A01
(PRI,GRP.6.ADT.A01?)+>

<!ELEMENT GRP.8.ADT.A01 (GT1)+>

<!ELEMENT GRP.9.ADT.A01 (IN1,IN2?,IN3?)+>

<!ELEMENT ADT.A01
(MSH,EVN,PID,PD1?,GRP.1.ADT.A01?,PV1,PV2?,GR
P.2.ADT.A01?,GRP.3.ADT.A01?,GRP.4.ADT.A01?,G
RP.5.ADT.A01?,DRG?,GRP.7.ADT.A01?,GRP.8.ADT.
A01?,GRP.9.ADT.A01?,ACC?,UB1?,UB2?)+>

. . . . .

<!ELEMENT VH.1 EMPTY>

<!ATTLIST VH.1 v CDATA #IMPLIED ty CDATA
#FIXED "ID">

. . . . .

```

Fig. 3. HIS.dtd

3.2 RIS XML DTD

Some radiology information systems are subsystems of hospital information systems, but sometimes they are operated as stand-alone systems. RIS contains patient demographic, billing information, procedure descriptions, scheduling, diagnostic reports, patient arrival scheduling, film location, movement, and room scheduling. We have designed RIS DTD based on the Radiology Report of HL7 RIM[5,16]. Figure 4 shows the tree structure of RIS XML DTD. The detail descriptions of each elements are as follows.

- RIS element: The RIS element is the root element of RIS XML DTD. It consists of one or more RADIOLOGY.REPORT elements.
- RADIOLOGY.REPORT element: The RADIOLOGY.REPORT element represents diagnostic evaluation reports on DICOM images, which are produced by PACS systems. This element contains attributes such as study.id and study.date. These attributes are utilized as keys to retrieve the related DICOM images. This element consists of HEADER and BODY elements.
- HEADER element: The HEADER element represents the head information of reports. It consists of PATIENT.INFO element and PROCEDURE.INFO elements.
- PATIENT.INFO element: The PATIENT.INFO element represents information on patients. This element contains attributes such as pname(patient name), pid(patient unique id), birthdate, and gender. These attribute values are determined by PID(Patient ID) segments of HIS.

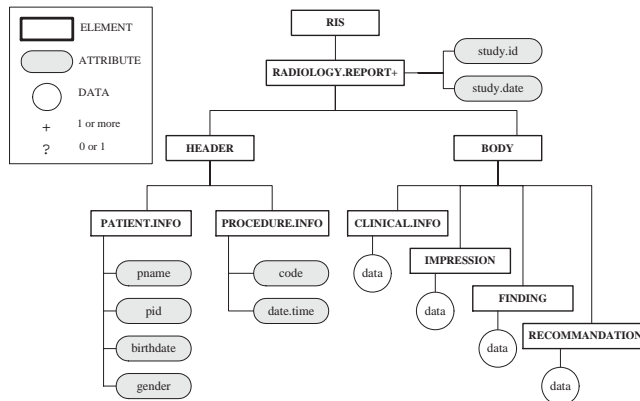


Fig. 4. RIS XML DTD structure

- **PROCEDURE.INFO** element: The **PROCEDURE.INFO** element represents information on procedures. This element contains **code**(procedure code) and **date.time** attributes. These attribute values are determined by PR1 (Procedure) segments of HIS.
- **BODY** element: The **BODY** element represents information on actual diagnostic evaluation results. It consists of **CLINICAL.INFO**, **IMPRESSION**, **FINDING**, **RECOMMENDATION** elements. The values of these sub-elements are text data with **PCDATA** data type.

Figure 5 shows the proposed RIS XML DTD.

```

<!ELEMENT RIS (RADIOLOGY.REPORT)+>
<!ELEMENT RADIOLOGY.REPORT (HEADER, BODY)>
<!ATTLIST RADIOLOGY.REPORT
  study.id CDATA #REQUIRED
  study.date CDATA #REQUIRED>

<!ELEMENT HEADER (PATIENT.INFO, PROCEDURE.INFO)>
<!ELEMENT PATIENT.INFO EMPTY>
<!ATTLIST PATIENT.INFO
  pname CDATA #REQUIRED
  pid CDATA #REQUIRED
  birthdate CDATA #REQUIRED
  gender CDATA #REQUIRED>

<!ELEMENT PROCEDURE.INFO EMPTY>
<!ATTLIST PROCEDURE.INFO
  code CDATA #REQUIRED
  date.time CDATA #REQUIRED>

<!ELEMENT BODY (CLINICAL.INFO, FINDING,
  IMPRESSION, RECOMMENDATION)>
<!ELEMENT CLINICAL.INFO (#PCDATA)>
<!ELEMENT FINDING (#PCDATA)>
<!ELEMENT IMPRESSION (#PCDATA)>
<!ELEMENT RECOMMENDATION (#PCDATA)>

```

Fig. 5. RIS.dtd

3.3 PACS XML DTD

We have designed PACS XML DTD based on DICOM Part3(Information Object Definition) and Part6(Data Dictionary)[7]. Figure 6 shows the tree structure of PACS XML DTD. The detail descriptions of each element are as follows.

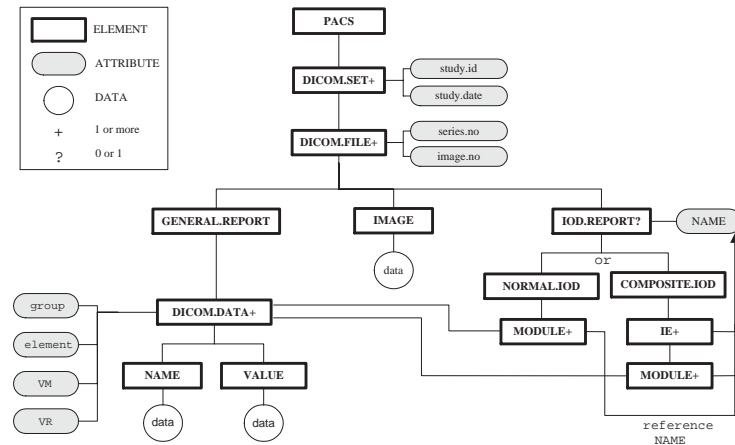


Fig. 6. PACS XML DTD structure

- PACS element: The PACS element is the root element of PACS XML DTD. It consists of one or more DICOM.SET elements.
- DICOM.SET element: The DICOM.SET element represents a set of DICOM files. This element contains attributes such as study.id and study.date. The study.id attribute represents the unique shooting number and it is used as a key to identify the appropriate DICOM.SET element. The study.date attribute represents the shoot date. This element consists of DICOM.FILE elements.
- DICOM.FILE element: The DICOM.FILE element represents a physical DICOM file. It contains series.no and image.no attributes. The series.no attribute represents the unique number given by the shooting direction. For a given direction, ten or twenty images are taken usually. The image.no attribute represents the unique number for one of images for the given direction. This element consists of GENERAL.REPORT, IOD.REPORT, and IMAGE elements. The GENERAL.REPORT is mandatory, while IOD.REPORT is optional.
- GENERAL.REPORT element: The GENERAL.REPORT element is used to represent individual DICOM Data Elements extracted from the given DICOM file. Therefore, it consists of one or more DICOM.DATA elements.

- DICOM.DATA element: The DICOM.DATA element represents one of DICOM Data Element of the given DICOM file. It contains attributes such as group, element, VR, and VM. Each attribute represents one of Data Elements of Data Dictionary of DICOM Part6. This element consists of NAME and VALUE elements.
- NAME and VALUE element: The NAME element represents the name of DICOM Data Element. The VALUE element represents the actual value of the Data Element.
- IOD.REPORT element: The IOD.REPORT element are used to represent DICOM Data Element more hierarchically. This element is based on the DICOM Part3 Data Model. It contains the name attribute. This elements consists of NORMAL.IOD or COMPOSITE.IOD element.
- NORMAL.IOD and COMPOSITE.IOD element: An IOD of DICOM Data Model can be a single object or a composite object, resulting in Normalized IOD or Compositied IOD. The NORMAL.IOD element consists of one or more N.MODULE elements. The COMPOSITE.IOD element consists of one or more IE(Information Entity) elements.
- IE element: The IE element contains the name attribute. This element consists of one or more C.MODULE elements.
- N.MODULE and C.MODULE element: These elements contain the name attributes. Both of them consist of one or more DICOM.DATA elements.
- IMAGE element: The IMAGE element represents the image information extracted from DICOM files.

Figure 7 shows the proposed PACS XML DTD.

3.4 XML DTD for integration

This section defines an integrated DTD for combining HIS DTD, RIS DTD, and PACS DTD. The generation procedure of this integrated DTD is shown in Figure 8.

The integrated DTD was designed from the viewpoint of patients by combining their related HIS, RIS, and PACS information. The INTEGRATED.REPORT is the root element of this DTD followed by the PATIENT elements. The PATIENT element contains attributes such as pname and pid. These attributes are used as key values for the integrated medical information retrieval. Figure 9 shows the tree structure of the integrated DTD.

The PATIENT element consists of elements, such as HIS, RIS, and PACS elements, which are root elements of HIS, RIS, and PACS DTD defined in the previous sections. Figure 10 shows the top level integrated XML DTD.

```

<!ELEMENT PACS (DICOM.SET)+>
<!ELEMENT DICOM.SET (DICOM.FILE)+>
<!ATTLIST DICOM.SET
  study.id CDATA#REQUIRED
  study.dateCDATA#REQUIRED>
<!ELEMENT DICOM.FILE (GENERAL.REPORT,IMAGE,IOD.REPORT?)>
<!ATTLIST DICOM.FILE
  series.no CDATA#REQUIRED
  image.no CDATA#REQUIRED>

<!ELEMENT GENERAL.REPORT (DICOM.DATA)+>

<!ELEMENT DICOM.DATA (NAME,VALUE)+>
<!ATTLIST DICOM.DATA
  groupCDATA#REQUIRED
  element CDATA#REQUIRED
  VR CDATA#REQUIRED
  VM CDATA#REQUIRED>

<!ELEMENT NAME (#PCDATA)>
<!ELEMENT VALUE (#PCDATA)>

<!ENTITY % name "
  name CDATA#REQUIRED">

<!ELEMENT IOD.REPORT (NORMAL.IOD|COMPOSITE.IOD)>
<!ATTLIST IOD.REPORT %name;>

<!ELEMENT NORMAL.IOD (N.MODULE)+>
<!ELEMENT N.MODULE (DICOM.DATA)+>
<!ATTLIST N.MODULE %name;>

<!ELEMENT COMPOSITE.IOD (IE)+>
<!ELEMENT IE (C.MODULE)+>
<!ATTLIST IE %name;>
<!ELEMENT C.MODULE (DICOM.DATA)+>
<!ATTLIST C.MODULE %name;>

<!ELEMENT IMAGE (#PCDATA)>

```

Fig. 7. PACS.dtd

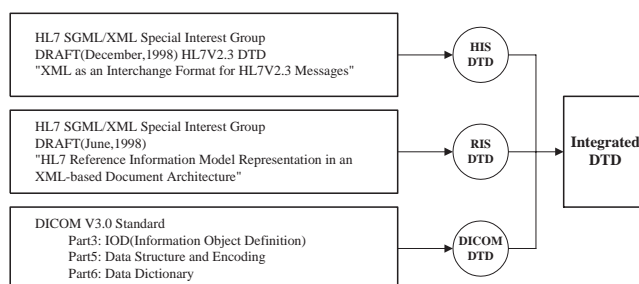


Fig. 8. Integrated DTD generation

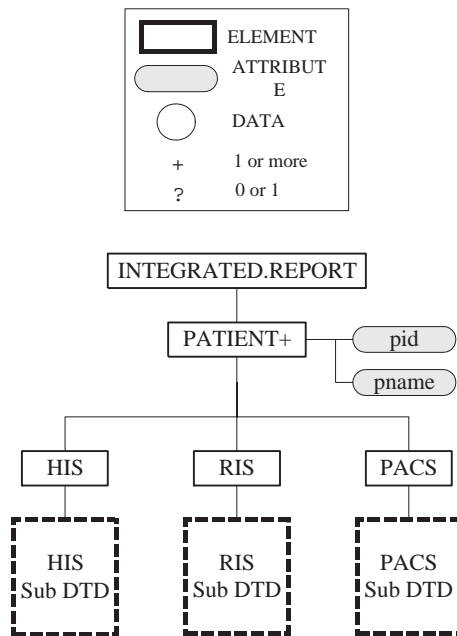


Fig. 9. INTEGRATED DTD structure

```

<!ENTITY % HIS-messages PUBLIC
"-//HIS//DTD HIS message definitions//
EN"
"HIS.dtd" >
%HIS-messages;

<!ENTITY % RIS-reports PUBLIC
"-//RIS//DTD RIS report definitions//
EN"
"RIS.dtd" >
%RIS-reports;

<!ENTITY % PACS-images PUBLIC
"-//PACS//DTD PACS image definitions//
EN"
"PACS.dtd" >
%PACS-images;

<!ELEMENT INTEGRATED.REPORT (PATIENT)+>
<!ELEMENT PATIENT (HIS, RIS, PACS)>
<!ATTLIST PATIENT
  pname CDATA #REQUIRED
  pid CDATA #REQUIRED>

```

Fig. 10. INTEGRATED.dtd

4 System realization

All the target medical information is extracted from the existing HIS, RIS, and PACS systems. The integrated XML document instances are generated according to the integrated DTD described in the section 3. The proposed XML DTD becomes the XML database schema. The XML instances are stored into XML databases with XSL stylesheets containing the presentation information. We have developed the prototype system using ObjectStore eXcelon and implemented the web interface using ASP(Active Server Page) on Window 2000 Server and IIS 5.0 Web server. The system architecture is shown in Figure 11.

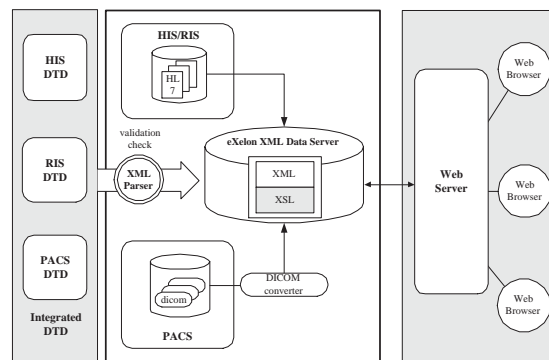


Fig. 11. System architecture

4.1 Generation of integrated XML documents

Currently, the extraction of medical information from HL7 messages and DICOM files is done manually. The automatic generation of XML instances from HIS, RIS, and PACS systems is one of further research topics. The HIS information is extracted from HL7 standard messages. Based on the extracted information and HIS XML DTD, the HIS XML document instances are generated. Figure 12 shows an example of HIS XML document instance generated from a HL7 ACK message.

In our system, the RIS information was extracted from the diagnostic evaluation reports(1996_RSNA_Reports.txt) for DICOM images provided by RSNA (Radiology Society of North America, Inc.). Based on the extracted information and RIS XML DTD, the RIS XML document instances are generated. Figure 13 shows an example of RIS XML document instance.

A DICOM file is a set of Data Elements having Group, Element, VR, Length, and Value in sequence. The Data Element with Group 0x7FE0 and Element 0x0010 contains the pixel data of the actual binary images[3].

In our system, we extracted each Data Element and image information from bi-

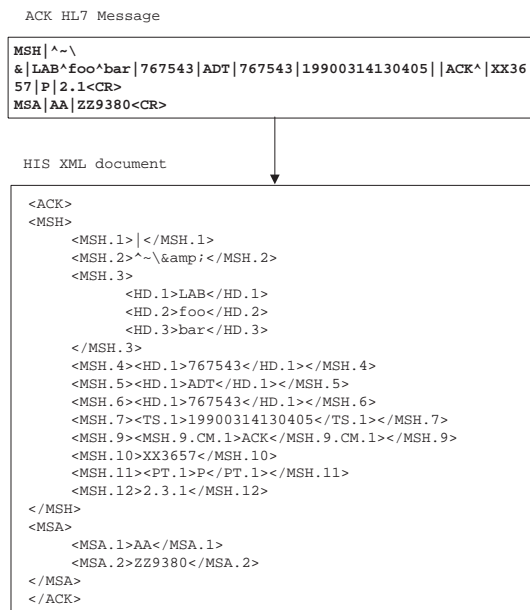


Fig. 12. An example of HIS XML document instance

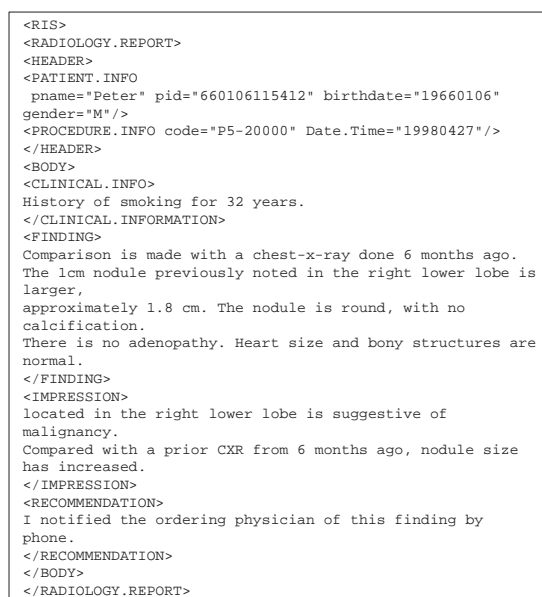


Fig. 13. An example of RIS XML document instance

nary DICOM files by using the DICOM2 converter. Figure 14 and 15 show Data Elements and image information extracted from a DICOM file respectively. In figure 14, each row represents one of DICOM Element information. The structure of figure 14 is the same with Part 6 Data Dictionary, except for the additional column for Data values.

Element Description	Group/Element	VM	VR	Data(value)
Study Date	(0008,0020)	1	DA	[1995.01.19]
Modality	(0008,0060)	1	CS	[US]
Patient's Name	(0010,0010)	1	PN	[NAPPER^MARGRET]
Patient's ID	(0010,0020)	1	LO	[ACN000001]
Patient's Birth Date	(0010,0030)	1	DA	[1950.04.20]
Performing Physician's Name	(0008,1050)	1-n	PN	[KLOFAS,EDWARD]
.....

Fig. 14. An example of text information from DICOM file



Fig. 15. An example of image information from DICOM file (chest.png)

Based on the extracted information and PACS XML DTD, the PACS XML document instances are generated. Figure 16 shows an example of PACS XML document instance.

4.2 Storing and retrieving mechanism of integrated XML documents

The resulting XML instances and XSL documents are stored in the XMLStore of ObjectStore eXcelon. The XMLStore parses XML data and supports the indexing mechanism on XML tags. The application programs for the integrated medical information system were developed using ASP. The query data on the web browser is transferred to the web server. The ASP program on the web server generates XQL(XML Query Language) command strings and calls eXcelon API using this XQL command as a parameter.

This call makes Server Extensions(a kind of user defined stored procedures

```

<PACS>
<DICOM.SET study.id="2019" study.date="19980209">
<DICOM.FILE series.no="1" image.no="1301">
  <GENERAL.REPORT>
    <DICOM.DATA group="0008" element="0016" VR="UI" VM="1">
      <NAME>SOP Class UID</NAME>
      <VALUE>1.2.840.10008.5.1.4.1.1.1</VALUE>
      .....
    <DICOM.DATA group="0028" element="0011" VR="US" VM="1">
      <NAME>Columns</NAME>
      <VALUE>440</VALUE>
    </DICOM.DATA>
  </GENERAL.REPORT>
  <IMAGE>chest.png</IMAGE>
</DICOM.FILE>
</DICOM.SET>
</PACS>

```

Fig. 16. An example of PACS XML document instance

written in Java) to be executed. Finally, the transferred XQL is executed by XMLStore through DOM(Document Object Model). The executed query returns results in XML forms, which are combined with XSL by MS XML, thus finally producing HTML documents for web browsers. Figure 17 summarizes this storing and retrieving mechanism.

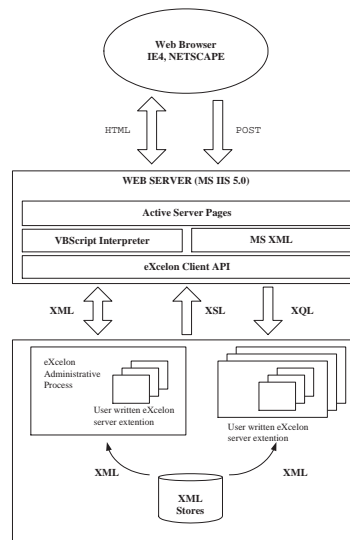


Fig. 17. Storing and retrieving mechanism

4.3 Web interface and retrieval examples

When a user logs in, the system shows the main screen. Users provide patient name and ID as query terms (Figure 18.a), and the system shows HIS, RIS, PACS lists for the given patient (Figure 18.b). Each of the lists are linked to the appropriate information. By clicking links, users can retrieve HIS information, PACS images, and diagnostic reports (Figure 18.c, Figure 18.d, Figure 18.e). If users click 'Patient List' link on the left frame in Figure 18.a, the system shows the list of all patients.

The query interface of our system is quite flexible. We can expand the web query interface (like Figure 18.a) by adding more complicated XQL queries and defining appropriate XSL stylesheets for each additional XQL queries. Of course, if users enter XQL directly on the eXcelon query interface, they can retrieve all kinds of information from XML medical databases with more complex query terms.

5 Conclusion

In this paper, we focused on the adoption of XML technology into healthcare domain, thus overcoming difficulties in medical information integration and compatibility.

We have analyzed HL7 and DICOM standard formats, and designed XML DTDs for HIS, RIS, PACS, and their integration. We extracted information from HL7 messages and DICOM files, and generated XML document instances and XSL stylesheets based on the proposed XML DTDs. We implemented the web interface for the integrated medical information system, which supports data sharing, information exchange and retrieval between two different standard formats. The proposed XML-based integrated medical information system will contribute to solve the problems of current medical information systems, by enabling integration of separated medical informations and by allowing data exchange and sharing through internet.

There should be further research on automatic generation of XML instances from traditional medical systems. Also, the proposed DTDs should be extended to include ongoing standard efforts, such as HL7 SGML/XML SIG, IMSIG, and DICOM WG20.

Acknowledgements

This work was supported by the Radiotherapy Medical Instruments Industrialization Technology Development Project of the Ministry of Science and Technology.

References

1. Lee, D.S., *An Analysis of Medical Image Storage and Transfer System Construction Model using DICOM*, Master Thesis, Yonsei Univ., 1998.
2. Yoon, M.J., *Implementation of MINI-PACS using DICOM converter on the Web*, Master Thesis, Chosun Univ., 1998.
3. Kwon, G.B. and Kim, I.K., "Web-based Medical Information System using DICOM Specification," in Proc. Korean Database Conference 2000, May 2000, pp.215-222.
4. ———, "XML as an Interchange Format for HL7 V2.3 Messages," HL7 SGML/XML SIG, <http://www.mcis.duke.edu/standards/HL7/committees/sgml/WhitePapers/hl7v2/hl7xml2.zip>, Dec. 1998.
5. ———, "HL7 Reference Information Model Representation in an XML-based Document Architecture," HL7 SGML/XML SIG, http://www.mcis.duke.edu/standards/HL7/committees/sgml/WhitePapers/hl7v2/xml_rim.rtf, June, 1998.
6. ———, "HL7 Document Patient Record Architecture DRAFT-Frame work Document," HL7 SGML/XML SIG, http://www.hl7.org/Special/Committees/sgml/PRA/HL7_PRA_9_27_1999.rtf, Sept. 1999.
7. ———, "DICOM 1999 - Parts (1-14) - DRAFT VERSION," ACR-NEMA Committee, <http://medical.nema.org/dicom/1999.html>.
8. ———, "Hospital Information System," <http://www.med.usf.edu/CLASS/his.htm>.
9. Gum, W. and Kim, I.K., "Design and Implementation of DICOM DB with ATLCOM," in Proc. Korean Database Conference 2000, May 2000, pp.243-249.
10. Horiil, S.C., Prior, F.W. and Bidgood, W.D., "DICOM: An Introduction to the Standard," http://www.xray.hmc.psu.edu/dicom/dicom/dicom_intro/DICOMintro.html.
11. ———, *ObjectStore eXcelon User Guide*, Object Design, <http://www.odi.com>, June, 1999.
12. ———, "Digital Imaging and Communications in Medicine," ACR-NEMA Committee Working Group VIS 255, 1993.
13. Boumphrey, F., et al., *Professional XML Applications*, Information & Culture Press, 1999.
14. ———, "Extensible Stylesheet Language(XSL) Version 1.0," W3C, <http://www.w3.org/TR/2000/CR-xsl-20001121>.
15. ———, "Extensible Markup Language(XML) 1.0(Second Edition)," W3C, <http://www.w3.org/TR/2000/REC-xml-20001006>.
16. ———, "HIS/RIS/PACS Interconnection," Korea Radiotherapy Association, <http://www.krta.or.kr/guest/pacs/pacs11.htm>.
17. Kim, J.H. and Han, M.C., "Integration of Hospital Information System and PACS," <http://radhome.snu.ac.kr/Department/Section/pacslab/pacscombine/pacscombine.htm>.
18. Chae, J.S. *Producing XML Home Pages*, Hong Pub., 1999.