31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

An XML Based Middleware for ECG Format Conversion

Xuchen Li, Vuk Vojisavljevic and Qiang Fang, *Member, IEEE*

Abstract—**With the rapid development of information and communication technologies, various e-health solutions have been proposed. The digitized medical images as well as the mono-dimension medical signals are two major forms of medical information that are stored and manipulated within an electronic medical environment. Though a variety of industrial and international standards such as DICOM and HL7 have been proposed, many proprietary formats are still pervasively used by many Hospital Information System (HIS) and Picture Archiving and Communication System (PACS) vendors. Those proprietary formats are the big hurdle to form a nationwide or even worldwide e-health network. Thus there is an imperative need to solve the medical data integration problem. Moreover, many small clinics, many hospitals in developing countries and some regional hospitals in developed countries, which have limited budget, have been shunned from embracing the latest medical information technologies due to their high costs. In this paper, we propose an XML based middleware which acts as a translation engine to seamlessly integrate clinical ECG data from a variety of proprietary data formats. Furthermore, this ECG translation engine is designed in a way that it can be integrated into an existing PACS to provide a low cost medical information integration and storage solution.**

I. INTRODUCTION

NOWADAYS, e-health technologies have attracted more and more interests from medical practitioners, government regulation and standard bodies and information and communication technology professionals. The digitized medical images as well as the mono-dimension medical signals are two major forms of medical information that are stored and manipulated within an electronic medical environment such as an electronics medical record (EMR) system [1] or a hospital information system (HIS). Though a variety of industrial and international standards such as DICOM and HL7 have been proposed by international consortium bodies, many proprietary formats are still pervasively used by many hospital information system and PACS [2] vendors. Those proprietary formats virtually segregate the medical information into "isolated islands" and are a big hurdle for the efforts to form a nationwide or even worldwide e-health network. Thus there is an imperative need to integrate those segregated heterogeneous medical data [3]. Moreover, many small clinics, many hospitals in developing countries and some regional hospitals in developed countries, which have limited budget, have evaded adopting the latest medical information technologies due to their high costs.

Health Level Seven (HL7) and Digital Imaging and Communications in Medicine (DICOM) are two mainstream healthcare standards recommended by many standard bodies and government agencies. For example, FDA Annotated ECG [4] complies with HL7 version 3.0 which is based on XML format standard and DICOM ECG is an extension of DICOM standard to cope with one dimensional ECG waveform signals. However proprietary formats such as GE ECG format from General Electric uses its own way to store ECG signals. GE, Philips as well as HL7 v3 ECG formats are based on XML standard so it will be relatively easy to convert both GE format ECG file and HL7 v3 based ECG format into an XML based middleware.

DICOM standard was developed on the basis of the experiences with the American College of Radiology - National Electrical Manufacturers Association (ACR-NEMA) standard. It was initially published in 1993 [5] and was continuously being extended ever since. The aim of this standard was to create an open platform to assist the distribution and viewing of medical images, such as CT scans, MRIs, Ultrasound and ECG signal waveforms. As a standard which restricts image management in PACS, DICOM, besides HL7 also becomes a focus in healthcare field. Not only medical images but ECG signal waveforms are up to DICOM standard. For example, DICOM ECG [6] is one of DICOM applications which display ECG signals complying with DICOM standard. This enables DICOM ECG format to communicate to PACS without any effort, however, it also challenges the communication between HL7 and PACS. Therefore both of HL7 and DICOM standards communicating with PACS will be a macroscopical aim in medical information area.

In conformity with the goal of low cost, researchers at University of Trieste designed a module which called O3-DPACS to integrate with PACS. Open Three Consortium (O3) is an open-source project. O3-DPACS was an evolution of PACS and it was a basic step to manage medical data and images together. In 2007, O3-DPACS was introduced as an open-source application based on Java to improve PACS by designing an interface module complying with both HL7 and DICOM standards. Although it provided a service of ECG storage function, the format of ECG [7] was only up to DICOM standard. In our research, we mainly develop a mono-dimension ECG signal convertor based on GE, Philips ECG and Annotated ECG standards in order to improve the DPACS. A new goal based on this research is to develop an open, scalable, low cost and universal new-fashioned DPACS for clinics and hospitals in metropolitans of developing countries and regional area of developed countries and to meet a new level.

Manuscript received April 23, 2009. This work was supported in part by Australian Postgraduate Awards.

Xuchen Li is with the School of Electrical and Computer Engineering, RMIT University, VIC 3000, Australia. E-mail: xuchenlee@hotmail.com.

Qiang Fang is with the School of Electrical and Computer Engineering, RMIT University, VIC 3000, Australia. E-mail: john.fang@rmit.edu.au.

Vuk Vosjisavljevic, is with the School of Electrical and Computer Engineering, RMIT University, VIC 3000, Australia. E-mail: vuk.vosjisavljevic@rmit.edu.au.

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

II. METHODS

This design is implemented in PHP (Hypertext Preprocessor) scripting language of version of 5.2.9-1. A GE, a Philips and an Annotated ECG files are selected as a sample ECG formats and HL7 v2.3.1 is selected as the interface of the middleware and PACS in this design. Besides DICOM and OPENECG, Philips and GE are also famous ECG device suppliers and their ECG devices are widely used all over the world. However GE and Philips, unlike other medical device suppliers, store ECG files in their own formats. In this design, GE, Philips and Annotated ECG formats are all parsed into HL7 message format and then they can communicate with DPACS via HL7 interface. Below in Figure 1 is the illustration of the design application among the network of XML based middleware, DPACS system and workstation. There are two parsing modules in this design: XML parsing module and HL7 message parsing module.

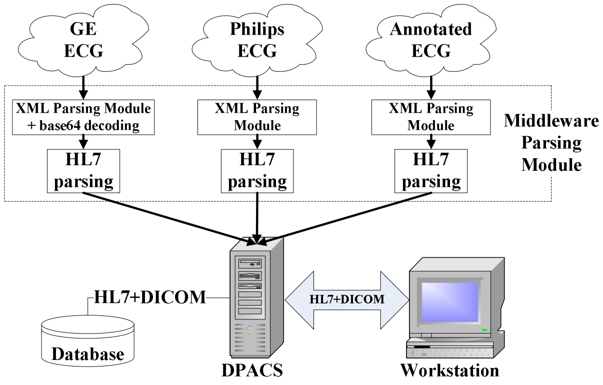


Figure 1. Illustration of the design application among the network of XML based middleware, DPACS system and workstation

A. System Design

The middleware is designed to receive and convert GE, Philips and Annotated ECG format files. There are two parser modules included: a main parser module which is created based on XML format to parse these three ECG format files and an accessorial parser module which is to convert the XML parsing modules into HL7 messages. The middleware will receive ECG data and send them using a common TCP/IP network transport protocol. The basic principal chart is shown in Figure 2.

B. ECG Conversion

All but sample GE ECG format has its own parsing module due to that the digital part of ECG results in GE ECG file is encoded in base64 code which is required to be transferred into a human readable string. Therefore there are two parts in GE ECG XML parsing module.

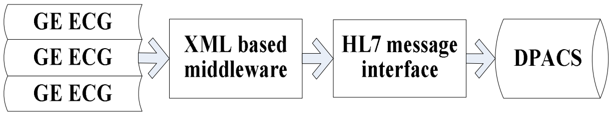


Figure 2. Principal of system design of data exchange

1) Decoder Design: A PHP base64 decoder is designed for the sample GE ECG format file during the conversion. Base64 encoding is one of the most popular encrypting methods which are based on binary data format. It is also used as an encoding scheme that encodes binary data. Besides base64 decoding, the characters in waveforms in the sample GE ECG file are stored in two's complement format which means the digits in waveforms are encrypted twice. If only base64 decoding module is proposed, the result will show that the units of the voltage in waveforms will be over 65,535 and they do not make any sense.

2) XML Parsing Module Design: XML, as a standard applied in most data exchanges, is much suitable to be dealt with and it will be helpful if there is an XML to HL7 conversion library which is available. There are a few Java libraries but nothing free and easy to use. PHP and XML are powerful scripting languages with much confusion but they are not as difficult as they look like because a lot of powerful tools are in XML and PHP scripting such as XPath and XSLT. PHP with its powerful MySQL database provides PACS a convenient working platform. During the parsing process, the structure of the middleware is required to be defined. There are two main types of parsers: tree-based parsers and event-based parsers [8].

Tree-based parsers: these parsers parse XML document into a tree structure, analyze the whole document and provide access to the tree elements, such as XML Document Object Model (DOM).

Event-based parsers: these parsers view XML document as a series of events such as Event-based parser (EXPAT) and Simple API XML (SAX) parser [9] which is an event driven parser. When a specific event occurs, they will call a function to define it. Event-based parsers focus on the content of the XML documents but not the structure. Therefore, event based parsers can access data faster than tree-based parsers.

EXPAT is a non-validating parser which ignores any DTDs (Document Type Definition). As an event-based and non-validating XML parser, EXPAT [10] is fast and small that it matches fast web applications. It must be noticed that XML documents must be well-formed otherwise EXPAT will generate an error.

A regular expression which is a substitute of SAX in this design is supported by nearly all major programming languages such as Java and PHP since it can change the tag values but the structure of an XML file. The method of reading and manipulating XML document is DOM, but this method needs to read the whole document and save it into a tree-based structure. This process is slow, un-efficient and resource-overused; therefore, EXPAT parser replaces DOM. However a regular expression parser, but EXPAT and SAX, is selected as the parser in the XML parsing modules because the easier the structure of XML parsing modules is the easier to parse them into HL7 message.

Creating first parser module starts by generating a new DOM document object and loading the ecg.xml files into the object where a loading function has been called. During the regular expression parsing, three sample ECG files are parsed into new XML files which are similar to each other, thus the parser in HL7 parsing modules can work lightly.

3) HL7 Parsing Module Design: The task of the parser in this module is to parse/convert the result of XML parsing module. In this design HL7 v2.3.1 is selected as the communication interface to PACS. Findings have shown that XML can serve as an implementable message specification for HL7 v2.3.1 messages and the ability to explicitly represent an HL7 requirement in XML conferring the ability to validate the requirement with an XML parser.

The proposal for mapping is when parsing hits Patient Identification (PID) and is used to match a patient. If no match is found, it will display a dummy "unmatched" patient. The information can be capsulated into both ORU^R01 and ORU^W01 segments. In HL7 message segments, OBX is to be parsed because the message has PatientArchives field which is tree-base formed such as "id", "firstname" and "lastname". One useful option is to let the user define a database account in MySQL which is to be used for importing source results.

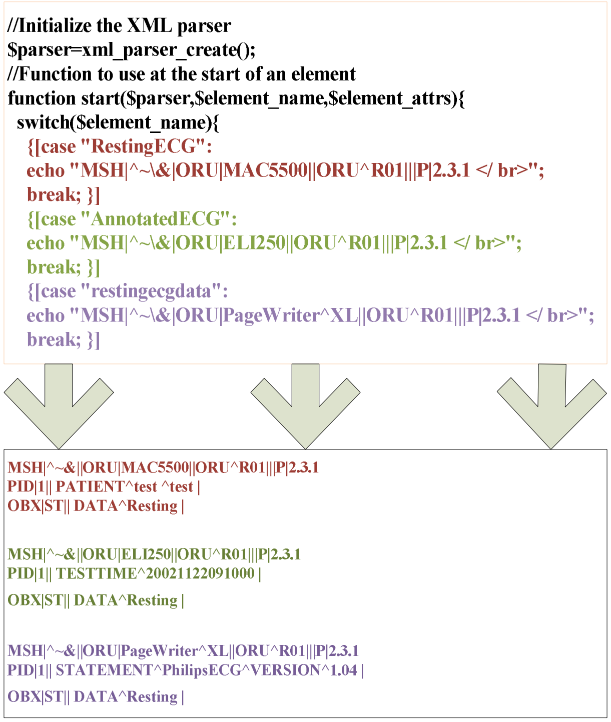
The time fields of acquisition time and acquisition date may need to be matched to existing or extra fields. If a test request is not selected according to the requested date time one link to a dummy, "XML error" affair can be created. This helps to say when someone outside forwards a result for a common patient to you. When we do not find a request for incoming data we could either make one up or just store the test results without linking them to any request. This is possible since we link them via a table in database that does not enforce direct linking inside test results. It should not be in any case because test results are wider in scope than test requests. Once the lab request is found or created, the OBXs are parsed according to their type: numeric (NM), string (ST) and structured numeric (SN). The segment specification of HL7 messages is shown in TABLE I.

TABLE I
SEGMENT SPECIFICATION OF ORU_R01 MESSAGE

Segment	**Specification**
MSH	Message Header
EVN	Event Type
PID	Patient Identification
[{ OBX }]	Observation / Result
[{ OBX_1 }]	Set ID
[{ OBX_2 }]	Value Type
[{ OBX_3 }]	Observation Identifier
[{ OBX_4 }]	Observation sub ID
[{ OBX_5 }]	Observation value
[{ OBX_6 }]	Units
[{ OBX_11}]	Observation Results Status ID
[{ OBX_13}]	User Defined Access Checks
[{ OBX_14}]	Date/Time of the Observation
[{ OBX_15}]	Producer’s ID
[{ OBX_16}]	Responsible Observer
[{ OBX_17}]	Observation Method
[{ OBX_18}]	Equipment Instance Identifier
[{ OBX_19}]	Date/Time of the Analysis
[{ OBX_20}]	Observation Site

Segments in this table are the key segments used in ORU_R01 message to send ECG observation result. The ‘[ ]’ means optional while ‘{ }’ means repeatable.

In HL7 parsing module there are two functions required to be declared, one to handle the element data and one to handle the character data within the elements. This is where the code which reflects the XML file is required to be changed which is to change the element and attribute names. This consists of two steps: a function to detect the start of real data and a function to detect when an element comes to an end in case to register when more than one contact is specified. For an element handler, each function is called once for each node. A switch statement is used to decide what action to take depending on which node is being processed. The parser will take care of the name and attribute variables.



```
//Initialize the XML parser
$parser=xml_parser_create();
//Function to use at the start of an element
function start($parser,$element_name,$element_attrs){
 switch($element_name){
   {[case "RestingECG":
   echo "MSH|^~\&|ORU|MAC5500||ORU^R01|||P|2.3.1 </ br>";
   break; }]
   {[case "AnnotatedECG":
   echo "MSH|^~\&|ORU|ELI250||ORU^R01|||P|2.3.1 </ br>";
   break; }]
   {[case "restingecgdata":
   echo "MSH|^~\&|ORU|PageWriter^XL||ORU^R01|||P|2.3.1 </ br>";
   break; }]
```

```
MSH|^~&||ORU|MAC5500||ORU^R01|||P|2.3.1
PID|1|| PATIENT^test ^test |
OBX|ST|| DATA^Resting |

MSH|^~&||ORU|ELI250||ORU^R01|||P|2.3.1
PID|1|| TESTTIME^20021122091000 |
OBX|ST|| DATA^Resting |

MSH|^~&||ORU|PageWriter^XL||ORU^R01|||P|2.3.1
PID|1|| STATEMENT^PhilipsECG^VERSION^1.04 |
OBX|ST|| DATA^Resting |
```

Figure 3. HL7 message parser module flow chart

The function "xml_set_element_handler()" has been abbreviated here by removing the other case $name="": { } statements. Furthermore a character handler is also required. Finally, the parser has been told which functions to use, read the data from the opened file and parse the contents. The data from the XML file is now held in “$data” and can be accessed using a standard PHP statement cycle. The flow chart of for HL7 message parser module is shown up in Figure 3. At the end, HL7 parsing module completes the whole XML file by releasing the parser.

This principal of this method can be applied in other versions of HL7 standard [11] because there is little difference among the HL7 message standards. As is always the case with XML when processing with a validating processor using a regular expression match function "preg_match", the extra blank space between elements can be removed in actual message instances.

III. DISCUSSION

A common network transmission protocol TCP/IP is setup and used in a LAN with one server which acted as a service class provider (SCP) and a client laptop which acted as a service class user (SCU). During the converting test with sample GE, Philips and Annotated ECG files are exported and tested from MAC 5500, ELI250 and PageWriter XL respectively. HL7 message codes are written into MySQL database. Two parsing modules work well.

EXPAT does not validate although it does read and demand external entities. If there is a well-formed parser which is suitable, EXPAT is a definitely fast one. However, since both parsers are really fast, this may only be of interest if one aims for maximum speed.

Additionally if a DOM structure is created in memory which typically needs a lot of space, the node structures will not be negligible, so that the overall speed depends not only on the raw parser speed, but on the quality of the DOM building code.

Another factor depending on the size of the XML file which may be important is that DOM trees typically need a lot of space in memory. This depends of course on how much markup in the document and how many accessories it has in the document. Although the XML DOM trees need notable less memory than other DOM implementation, it is not the best solution.

IV. CONCLUSION

In this paper, an XML based middleware is designed to convert a variety of ECG formats into HL7 message formats and to integrate with an existing PACS systems which can be seen more as storage facilities than as true processing environments. In order to integrate ECG data converting module with PACS it is required to realize the conversion among ECG formats in both data type and image type. Since there are a great many of ECG formats, more format conversions will be implemented in the near future. For instance, SCP ECG, Siemens ECG and DICOM ECG convertors will also be added in the middleware. Nevertheless this task is feasible and the final result will be affirmative.

With the evolution of HIS and PACS the task of hospital system integration tends to be easier and less expensive. Yet merely one or two researches considered the cost and convenience of medical systems and this is the most important aspect in current medial and e-healthcare environment nowadays, so that design of cheap middleware to facilitate medical system integration is the theme of this research.

REFERENCES

[1] W. MacKinnon and M. Wasserman, "Integrated Electronic Medical Record Systems: Critical Success Factors for Implementation," in *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, 2009, pp. 1-10.

[2] L. Boqiang, L. Xiaomei, L. Zhongguo, Y. Qingwei, and Y. Xiaohong, "Design and implementation of information exchange between HIS and PACS based on HL7 standard," in *Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 2008, pp. 552-555.

[3] M. Struck, S. Pramatarov, and C. Weigand, "Method and system for standardized and platform independent medical data information persistence in telemedicine," in *Computers in Cardiology, 2008*, 2008, pp. 257-260.

[4] W. Wesselink, N. Goga, A. J. Mooij, and R. Spronk, "Formal methods impact on ANSI standard HL7/IM - filling gaps in MSC theory," in *Electrical and Computer Engineering, 2005. Canadian Conference on*, 2005, pp. 1656-1659.

[5] B. Blazona and M. Koncar, "HL7 and DICOM based integration of radiology departments with healthcare enterprise information systems," *International Journal of Medical Informatics,* vol. 76, pp. S425-S432, 2007.

[6] DICOM Supplement 30: Waveform Interchange, Nat. Elect. Manufactures Assoc.: ACR-NEMA, Digital Imaging and Communications in Medicine, NEMA, Washington D.C., 2000.

[7] P. Inchingolo, M. Beltrame, P. Bosazzi, D. Cicuta, G. Faustini, S. Mininel, A. Poli, and F. Vatta, "O3-DPACS Open-Source Image-Data Manager/Archiver and HDW2 Image-Data Display: An IHE-compliant project pushing the e-health integration in the world," *Computerized Medical Imaging and Graphics,* vol. 30, pp. 391-406, 2006.

[8] Y. Ishitani, K. Fume, and K. Sumita, "Table structure analysis based on cell classification and cell modification for XML document transformation," in *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, 2005, pp.1247-1252 Vol. 2.

[9] P. Yinfei, Z. Ying, and K. Chiu, "Hybrid Parallelism for XML SAX Parsing," in *Web Services, 2008. ICWS '08. IEEE International Conference on*, 2008, pp. 505-512.

[10] G. Zhenghong, P. Yinfei, Z. Ying, and K. Chili, "A High Performance Schema-Specific XML Parser," in *e-Science and Grid Computing, IEEE International Conference on*, 2007, pp. 245-252.

[11] A. Schloegl, F. Chiarugi, E. Cervesato, E. Apostolopoulos, and C. E. Chronaki, "Two-way converter between the HL7 Annotated ECG and SCP-ECG data formats using BioSig," in *Computers in Cardiology, 2007*, 2007, pp. 253-256.