

HL7메시지의 효율적인 저장과 검색을 위한 XML데이터베이스의 설계

이민경⁰*, 전종훈*, 정재현** 최진욱***

*명지대학교 컴퓨터공학과, **서울대학교 컴퓨터공학부, ***서울대학교 의과대학 의공학교실
(mkl, jchun}@mju.ac.kr, cjh@cygnus.snu.ac.kr, jincho@snu.ac.kr

Designing XML Database

for Effective Storage and Retrieval of HL7 messages

Minkyung Lee⁰*, Jonghoon Jun*, Jaeheon Cheong**, Jinwook Choi***

* Dept. of Computer Engineering, Myongji University

**School of Computer Science and Engineering, Seoul National University

*** Dept. of Biomedical Engineering, Seoul National University

요 약

HL7은 텍스트 기반의 의료정보교환 표준이다. HL7메시지는 한 환자가 내원하여 퇴원할 때까지 여러 종류의 메시지가 계속적으로 다량 발생하게 되는데, 그 메시지들의 일부 특정 세그먼트들은 반복적으로 생성되어 사용된다는 특징이 있다. 그러므로 HL7메시지를 데이터베이스에 저장할 때는 불가피하게 특정 세그먼트들에 많은 중복이 발생하게 된다. 또한 HL7메시지는 프로세스 흐름중심으로 발생되므로, 중복을 제거한 환자중심의 데이터베이스 설계가 필요하다. 본 논문에서는 XML데이터베이스를 사용하여 HL7메시지를 저장하는 방식을 제안한다. 이는 XML DTD가 병원정보를 다루는 여러 의료정보표준들 상호간의 정보교환 및 공유를 위한 스키마로서의 역할을 할 뿐만 아니라, 인터넷을 통해 여러 가지 서비스 제공할 수 있도록 효과적인 저장 및 검색을 지원하기 때문이다. 이질적인 병원정보시스템을 가지고 있는 병원들 간에 의료정보를 교환하고, 환자들의 의료정보를 통합관리 하기 위한 XML데이터베이스를 설계하고 구현하였다.

1. 서 론

최근 인터넷을 기반으로 하는 웹의 급속한 발전으로 이제 정보는 인터넷을 통해서 교환되고 공유될 수 있는 개방된 환경에서 정보자체의 가치보다 더 큰 가치를 지니게 되었다.

이러한 경향은 의료분야에도 예외없이 적용되어 지역 네트워크 기반으로 운영되어 오던 기존의 의료정보 시스템도 병원 및 타 기관들과의 데이터공유 필요성, 인터넷을 통한 원격진료 서비스 등으로 인해 개방형 시스템으로의 전환이 요구되고 있다. 따라서, 의료정보의 교환과 공유를 위해 각 의료관련 시스템들에 대한 표준들이 제정되었으며, 거기에 지역적으로 분산되어 있는 HIS (Hospital Information System)의 정보를 교환하기 위한 텍스트기반의 HL7(Health Level 7)을 비롯해, PACS(Picture Archiving & Communication System)를 위한 DICOM(Digital Imaging and Communication in Medicine), EDIFACT, X.12 등이 있다.

그 중에서도 HL7은 이질적인 HIS를 가진 병원들간의 발생되는 많은 텍스트 기반 의료정보들을 교환하기 위한 표준으로 한 환자가 내원하여 퇴원할 때까지 프로세스 흐름을 중심으로 HL7메시지가 발생하게 되는데, 각 병원들은 이질적인 데이터베이스를 가지고 있으므로, 발생되는 HL7메시지의 형식 또한 일정치 못하여 직접적인 교환이 어렵다. 따라서 이러한 이질적인 HIS에서 발생하는 다른 형식의 HL7메시지들에 포함된 의료정보들을 교환하기 위해서는 HL7메시지를 위한 통합된 데이

터베이스를 설계하는 것이 필요하다. 또한 XML DTD를 사용하여 데이터베이스를 설계 할 경우, 병원정보를 다루는 여러 의료정보표준들(DICOM, EDIFACT, X.12 등) 상호간의 정보교환 및 공유를 위한 스키마로서의 역할을 할 뿐만 아니라, 인터넷을 통한 여러 가지 서비스 제공을 위해 효과적인 저장 및 검색이 가능하다.

2. 관련 연구

HL7메시지는 데이터 전송의 최소 전송단위로 순차적인 세그먼트(segment)들의 집합으로 구성되어 있고, 그 세그먼트들은 데이터필드(data field)들의 집합으로 구성되어 있는데, 각각의 세그먼트는 그 기능에 따라 MSH, EVN, PID, PV1 등의 이름이 주어진다 또 각각의 데이터필드들은 컴포넌트(component)들로 구성되고, 그 컴포넌트들은 서브 컴포넌트들로 구성되어 있다. HL7메시지의 구성을 보면 XML과 비슷한 트리모양을 가지고 있음을 알 수 있다.

HL7은 모든 정보를 다루고 있기 때문에 꼭 필요한 필수데이터 뿐만 아니라, 옵션으로 넣을 수 있는 정보도 많고, 경우에 따라서는 중복을 허용하는 세그먼트나 데이터 필드들도 있다. 또한 데이터 필드들은 컴포넌트나 서브 컴포넌트들로 정보를 세분화시켜서 저장되어 있다.

이것은 HL7이 유연성을 가진 표준이기 때문에 필수 항목을 제외한 나머지는 HL7을 사용하는 병원마다 다른 구조의 HL7메시지를 가질 수 있음을 의미한다. 따라서 데이터를 저장하는 데

있어서 규격화하기가 곤란하다. 이러한 HL7메시지는 병원 프로세스에 따른 트리거이벤트에 의해 생성되어 다른 시스템들에 보내어지므로, 각 메시지마다 환자들의 기본정보(PID)가 계속 중복되어, 환자중심의 검색이 어렵다는 단점이 있다.[1]

XML은 HL7과 같이 비정형화된 정보를 표현하고 저장하는데 적합할 뿐만 아니라, HL7메시지를 XML로 변환하여 저장하기 위해 XML문서 저장이 가능한 통합임상데이터베이스를 설계할 경우, 다른 형식의 HL7메시지를 사용하거나, 여러 다른 표준들 상호간의 정보 교환을 위한 의료정보 스키마를 제공하게 되어 XML이 메타언어의 역할을 하고, 환자중심의 웹검색도 용이해진다.[2]

XML문서를 저장하기 위한 방법에는 [3~7]에서 제안한 것처럼 다양한 방식이 있다. 본 논문에서는 비정형화된 여러종류의 메시지를 발생시키고, 메시지의 구성상 트리의 깊이가 깊은 HL7메시지의 특성을 고려하여, 환자중심의 효율적인 검색과 저장을 위해 OODB기반의 XML데이터베이스를 이용해 데이터를 의미적으로 나눠 저장하는 방식을 취하고자 한다.

3. HL7메시지들을 이용한 의료정보의 효율적인 저장

3.1. HL7메시지의 생성범위 및 환자 프로세스 정의

프로세스 중심의 HL7메시지는 병원업무중에 발생하는 이벤트 형식으로 생성된다. 따라서 환자와 직접적으로 연관된 메시지뿐만 아니라, 병원 내에서 발생하는 환자관리를 위한 모든 것을 다루게 되므로, 본 논문에서 구현대상으로 삼은 HL7 메시지는 PatientAdmission, OrderEntry, ObservationReporting, Scheduling의 4가지 메시지타입으로 한정하였다.

3.2. HL7메시지를 XML문서로 변환하기 위한 DTD (Document Type Definition) 정의

HL7메시지를 XML문서로 변환은 데이터베이스에 저장되기 전 단계로, 저장의 효율성을 고려하여야 한다. 따라서 HL7메시지 내의 중복되는 데이터들을 제거할 수 있는 DTD의 정의가 필요하다.

3.2.1. HL7메시지의 중복 제거

HL7메시지는 프로세스 중심으로 발생되기 때문에 환자를 구분하기 위해 환자주정보(MajorInfo)에 해당하는 PID(환자개인신상정보)세그먼트가 메시지마다 필수적으로 중복되어 나오며, 부정보(MinorInfo)에 해당하는 NK1세그먼트(환자가족정보), AL1(보험 정보) 등은 메시지 종류에 따라 선택적으로 중복되어 나오게 된다. 따라서 중복 제거를 위해 환자의 주정보(PID) 세그먼트와 부정보(NK1, AL1 등) 세그먼트들은 하나의 환자신상정보(patient information)로, 메시지의 정보를 포함하고 있는 MSH세그먼트와 주정보, 부정보를 제외한 나머지 세그먼트들은 환자의료정보(medical information)로 구분하여 중복없는 환자의료정보만을 저장하고자 한다. 따라서 HL7메시지는 그림 1과 같이 두 종류의 정보로 재구성된 후 XML문서로 변환된다.

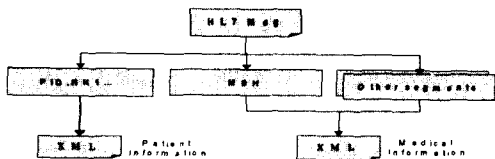


그림 1 HL7메시지를 XML문서로 재구성

3.2.2. HL7메시지에서 환자신상정보

환자신상정보는 메시지들을 저장하기 위한 대상을 확인할 뿐만 아니라, 검색을 위해 사용되어지는 중요한 정보이다. 초기에 환자가 등록될 때는 환자신상정보가 XML문서로 변환되어 데이터베이스에 저장되지만, 한 번 저장된 후에는 갱신이 없는

한 더 이상의 저장없이 이미 저장된 환자신상정보인 PID의 Patient ID(혹은 주민등록번호)를 기본키로 이용하여 환자의료정보를 저장하거나 환자를 검색한다. HL7메시지는 XML로 변환 시 데이터필드들이 그림 2와 같이 텍스트노드로 변환되는데, Patient ID를 이용해서 검색할 경우 Patient의 하위 노드들까지 검색해야만 필요한 정보를 검색할 수 있다. 따라서 Patient ID 데이터필드를 어트리뷰트(attribute)로 변환함으로써 빠르고 효율적인 검색을 지원한다. 또한 의료정보를 저장하기 위해 환자의료정보 데이터베이스의 위치와 이름이 MedicalInfo라는 엘리먼트의 DATA형태로 저장되어 있어, 해당 환자의 의료정보를 효율적으로 검색하고, 많은 의료정보가 발생할 경우 여러 데이터베이스에 나누어 저장하고, 유지보수 할 수 있다.

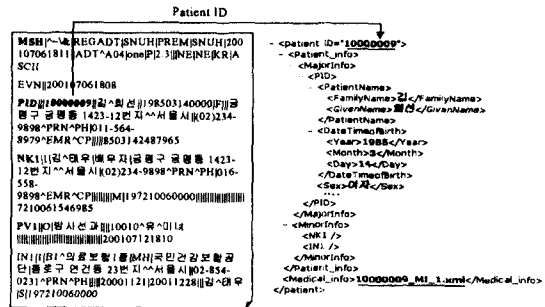


그림 2 HL7메시지에서 환자신상정보를 XML로 변환한 예

그림 3은 환자신상정보 저장을 위한 DTD구조를 보여준다.

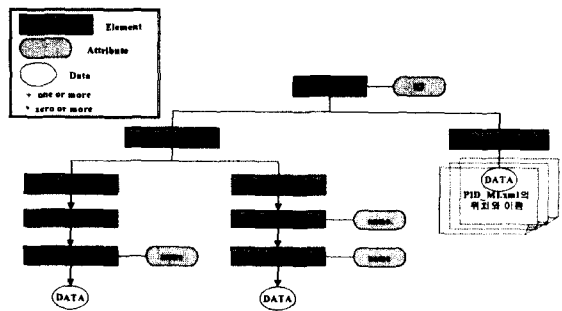


그림 3 환자신상정보 저장을 위한 DTD 구조

3.2.3. HL7메시지에서 환자의료정보

의료정보는 다양한 기관에서 새로운 정보가 계속적으로 발생되기 때문에, 날짜별 검색과 여러 병원이나 시스템에서 발생하는 메시지의 출처를 구분하기 위해서 그림 4와 같이 메시지가 발생하는 시간과 수신지(Receiving Application/Facility), 발신지(Sending Application/Facility)를 어트리뷰트로 정의하였다.

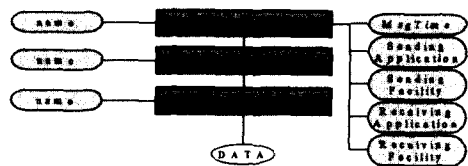


그림 4 환자의료정보 저장을 위한 DTD구조

3.3.XML데이터베이스 설계

CCDB(Central Clinical DataBase)는 개념적으로 모든 환자들의 신상정보를 하나의 XML파일로 저장된 통합환자정보데이터베이스(CPIDB:Central Patient Information DataBase)와 환자 개인의 의료정보가 환자마다 하나의 XML파일로 저장된 개인 임상정보데이터베이스(PCIDB:Personnel Clinical Information DataBase)로 나눌 수 있다. 즉, CPIDB에는 환자신상정보만, PCIDB에는 환자의료정보만 저장되지만, CPIDB의 경우는 모든 환자들의 신상정보가 하나의 XML파일로 묶여있는 형태이고, PCIDB는 각 개인의 의료정보들이 하나의 XML파일로 묶여서 여러개의 XML화일이 존재하게 되는 것이다. 그리고 CPIDB에 있는 환자와 PCIDB의 의료정보를 연결하는 고리가 환자정보의 MedicalInfo가 된다. 따라서 CPIDB에서는 ID를 통해 환자를 빠르고 효과적으로 찾고, 검색된 환자의 MedicalInfo로부터 환자의료정보데이터베이스의 이름과 위치를 찾아 그 환자의 PCIDB에서 의료정보를 찾는 과정을 거치게 된다.

그림5에서 보면 HL7메시지가 발생된 후 환자 정보는 XML로 변환되어 CPIDB에 있는 patients.xml안에 Patients 루트(root) 밑에 하위엘레먼트 형태로 저장되고, 의료정보는 변환 후 PCIDB에 있는 PID_ML_1.xml안에 MedicalInfo루트 밑에 년도와 월별로 나누고, 메시지타입별 나누어 저장된다. 이때 환자의 의료정보가 많을 경우 여러 이름의 XML화일로 나누어 저장될 수 있다. Year와 Month의 어트리뷰트num은 해당 년도와 월을 나타내며, MedicalSection의 name은 4가지(PatientAdmission, OrderEntry, ObservationReporting, Scheduling) 중 하나의 메시지 타입을 낸다. 이러한 설계는 많은 의료정보가 발생될 경우나 유지보수하기에 편할 뿐만 아니라, 환자중심의 빠른 검색이 가능하다는 장점이 있다.

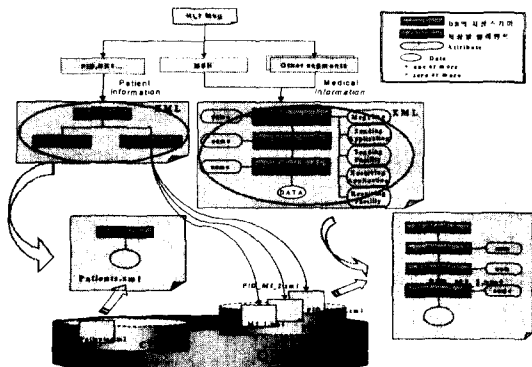


그림 5 XML데이터베이스에서 의료정보의 저장형태

4. 구현결과

그림 6은 전체 시스템 아키텍처이다. HIS내의 HL7메시지 서버는 이질적인 HIS의 데이터베이스로부터 HL7메시지를 만들어 XMLConverter로 보내고, XMLConverter는 XML로 변환시켜 CCDB에 전송하게 되면 웹에서의 검색이 가능하게 된다. 구현은 Windows2000환경 하에서 Visual C++을 사용하였고, 데이터베이스는 eXcelon 2.6을 사용하였다. XML DTD생성과 HL7을 XML로 매핑하기 위한 tool은 Symphonia 3.0을 사용하였고, Web 검색을 위해 JSP와 XSLT를 사용하였다.

그림7은 메시지타입이 PatientAdmission인 ADTA04메시지와 메시지타입이 ObservationReporting인 ORUR01메시지가 XML로 변환을 거쳐 CCDB에 저장된 후, 웹에서 환자 ID로 환자 신상정보와 환자정보를 검색한 예이다.

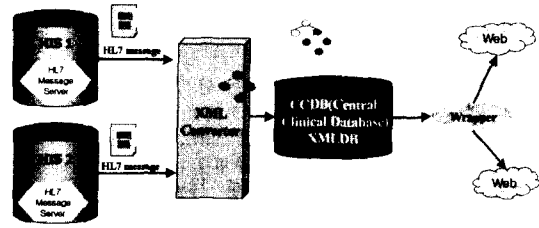


그림 6 시스템 아키텍처

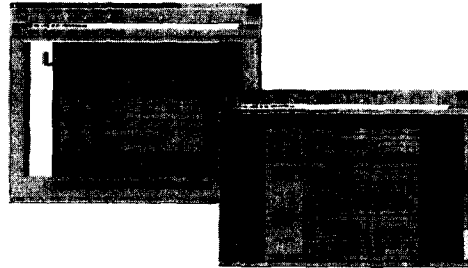


그림 7 환자신상정보와 의료정보의 검색 예

5. 결론 및 향후 계획

본 연구에서는 HL7메시지의 효율적인 저장과 검색을 위한 XML데이터베이스를 설계하였다. 이를 위해 하나의 HL7메시지를 환자신상정보와 환자의료정보로 의미적으로 분리하여 XML로 변환 후 XML데이터베이스에 저장하였다. 따라서 중복을 제거하고, 유연성을 높이며, 환자중심의 검색이 용이하게 되었다.

향후로는 모든 의료정보를 효과적으로 저장하고 검색할 수 있는 데이터베이스의 설계뿐만 아니라, 의료정보와 흡사하게 계속적으로 발생하는 다량의 데이터를 효과적으로 저장, 검색할 수 있는 개인중심의 데이터베이스를 설계하는 것이 필요하다.

참고문헌

- [1]HL7 organization, HL7 standard version 2.3 , health level seven, Inc, 1997
- [2]W3C, Extensible Markup Language(XML)1.0, <http://www.w3.org/TR/2000/REC-xml-20001006.html>, Oct,2000
- [3]Alin Deutsch, Mary Fernandez, and Dan Suciu. Storing Semistructured Data with STORED , SIGMOD, 1999
- [4]Daniela. Florescu and Donald Kossman, Storing and Querying XML Data using an RDBMS , IEEE Data Engineering Bulletin 22(3), p.27~34, 1999
- [5] J. Shanmugasundaram, K. Tufte, C. Zhang, G.. He, D. J. DeWitt, and J. F. Naughton, Relational Databases for Querying XML Documents: Limitations and Opportunities , Proc. Of 25th Intl Conf. Con VLDB, Edinburgh, Scotland, UK. P.302~304, 1999
- [6]Takeyuki Shimura,Masatoshi Yoshikawa, and Shunsuke Uemura, Storage and Retrieval of XML Documnets Using Object-Relational Database, DEXA 1999
- [7]Jason McHugh, Serge Abiteboul, Roy Goldman, Dallan Quass, and Jennifer Widom. Lore: A Database Management System for Semistructured Data. SIGMOD Record, 26(3), 1997