

XML 기반의 바이오 및 임상정보 통합 저장소를 위한 미들웨어의 설계

정종일, 유의혁, 이태현, 신동규, 신동일
세종대학교 컴퓨터공학과

e-mail : {jijeong, solui, leeth, shindk, dshin}@gce.sejong.ac.kr

Design of Middleware for Integration Repository of XML-based Bio and Medical Information

Jongil Jeong, Weehyuk Yu, Taeheon Lee, Dongkyoo Shin, Dongil Shin
Dept. of Computer Engineering, Sejong University

요 약

최근 환자중심의 진료환경 구축을 통한 진료의 질적인 향상을 위해 바이오 및 임상정보 통합의 필요성이 점차 커지고 있다. 따라서 기존의 폐쇄적인 의료시스템들을 개방적인 시스템으로 전환하고 각 시스템간에 정보를 교환하고 공유하기 위해 HL7 등의 표준들이 급속히 확산되고 있다. 그러나 HL7 은 임상문서 교환을 위한 유연성있는 표준이지만 시스템에 의존적인 형태의 메시지는 이질적인 시스템간의 정보교환에는 부적합하다. 따라서 HL7 표준중 XML 기반의 임상기록 저장 구조인 CDA 를 이용하여 환자의 임상정보가 통일되고 통합될 필요성이 있으며 본 논문은 새로운 CDA 를 지원하기 위해 바이오 및 임상정보를 통합하고 통합된 정보를 통합저장소에 저장하는 기능과 시스템 간 자료전송 그리고 외부의 응용프로그램을 통한 통합저장소의 접근제어 기능을 수행하는 미들웨어를 설계한다.¹

1. 서론

최근 환자중심의 진료환경 구축을 통한 진료의 질적 향상을 위해 바이오 및 임상정보 통합의 필요성이 점차 커지고 있다. 따라서 기존 병원 시스템들을 개방형 시스템으로 전환하도록 유도하고 지역적 네트워크의 한계로 인해 접근이 제한되었던 폐쇄적인 의료정보를 교환, 공유 및 통합 관리할 수 있는 환경을 제공하기 위해 HL7 (Health Level 7), PACS (Picture Archiving & Communication System), DICOM (Digital Imaging and Communication Medicine), EDIFACT X.12 같은 표준들이 제정되어 발전을 지속하고 있다. 특히, HL7 은 임상문서 교환을 위한 매우 유연성이 뛰어난 표준이지만 시스템에 의존적인 형태의 메시지가 생성되는 특징 때문에 이질적인 시스템간에 데이터의 통합과 교환에 있어 그 유연성이 오히려 장애요소가 되고있다 [1]. 이를 극복하기 위해서는 HL7 표준중 XML 기반의 임

상기록 저장 구조인 CDA (Clinical Document Architecture)를 이용하여 환자의 임상정보를 통일하고 통합할 필요가 있다.

본 논문은 XML 기반의 데이터 구조로 바이오 및 임상정보를 통합하고 이를 저장하기위한 통합 저장소로 가장 적합한 데이터베이스 시스템을 비교하고 이질적인 병원 시스템간 통합자료 공유와 외부 응용프로그램들의 통합저장소에 대한 접근요청을 제어하는데 필요한 미들웨어를 설계한다.

2. 관련연구

개별적인 병원 시스템에 저장되어 있는 의무기록들이 재정적, 시간적 낭비 없이 효율적으로 이용되기 위해서는 통합을 위해 새롭게 정의된 CDA 에 따라 데이터가 수집되고 공유되고 교환되어 통합될 필요가 있다. 또한 통합된 정보를 모든 병원들이 이용하기 위

¹ 본 연구는 보건복지부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임. (0412-MI01-0416-0002)

해서는 중앙의 임상 데이터베이스가 필요하다. 중앙의 임상 데이터베이스를 구축시에는 문서의 원본을 유지해야 하는 CDA 문서의 특성을 고려해야 한다 [2]. 따라서 이러한 특성을 유지하기 위해 통합 문서들을 어떤 문서 저장관리 시스템에 저장하는 것이 관리적인 측면에서 가장 효율적인지에 대한 비교와 분석이 필요하다.

2.1 XML 문서 저장관리 시스템 기술

XML 문서는 파일 시스템, 관계형 데이터베이스 시스템, 객체지향 데이터베이스 시스템 그리고 XML 전용 데이터베이스에 저장될 수 있다. 파일 시스템은 XML 데이터의 저장이나, 문서 전체에 대한 검색은 용이하지만 문서의 갱신 등과 같은 특정한 질의의 처리에 대한 지원이 부족하다. 객체지향데이터베이스 시스템은 XML 문서가 객체 기반에 구조를 두었기 때문에 XML 문서를 저장하기에 가장 적합한 저장방법이라고 할 수 있지만 현실적으로 대용량의 데이터베이스에 대한 복잡한 질의를 처리하기에는 어려움이 많다. 관계형 데이터베이스 시스템은 일반적인 데이터와 반구조적 데이터가 공존할 수 있고, 대용량 데이터베이스에 대한 복잡한 질의 처리에 효율적이기 때문에 현실적으로 XML 문서를 저장하고 관리하기에는 효율적인 저장 방법 중에 하나라고 할 수 있다. XML 전용 데이터베이스는 XML 에 가장 적합한 계층적 구조이고 XML 문서를 가장 적합하게 표현할 수 있기 때문에 XML 기반의 통합 문서를 저장하기에 가장 적합한 방법이라 할 수 있다.

2.2 관계형 데이터베이스와 XML 전용데이터베이스의 비교

XML 기반의 CDA 문서의 효과적인 저장과 검색을 위해 어떤 종류의 데이터베이스 시스템이 가장 적합한가를 파악하기 위해서는 XML 문서의 저장과 검색의 효율성을 알아볼 수 있는 항목들을 구분하고 비교할 필요가 있다. XML 문서를 저장할 수 있는 대표적인 데이터베이스 시스템인 관계형 데이터베이스와 XML 전용 데이터베이스의 비교를 위한 항목들은 다음과 같이 구분될 수 있다.

- 계층적인 Naming 구조: 관계형 데이터베이스는 지원하지 않기 때문에 데이터를 정규화하거나 단순화하여 데이터의 효율성을 제한해야하고 모든 계층의 Naming 구조는 관계형 데이터베이스 위에 만들어져야 한다. 반면에 XML 전용 데이터베이스는 XML 에 가장 적합한 계층적 구조를 제공하여 XML 문서를 원본형태로 저장할 수 있다.

- 계층적 트리구조 검색: 관계형 데이터베이스의 트리 내 이동은 트리 구조를 따라 올라가고 그곳에서 다른쪽으로 내려가는 방법만 가능하며 intensive join 의 계산이 필요하므로 주요 성능저하의 요인이 되고 있다. 반면에 XML 전용 데이터베이스는 XML 계층 구조에 적합한 XPath, XQuery 를 사용하므로 트리 구조에 따른 자유로운 검색으로 빠르고 효율적인 검색이

가능하다.

- 스키마 변경: 관계형 데이터베이스는 스키마의 변경에 따라 정규화를 다시 해야 하는 반면 XML 전용 데이터베이스는 쉬운 스키마 변경이 가능하다.

- 가변 길이와 형식을 갖는 데이터의 지원: 관계형 데이터베이스는 데이터를 고정된 필드에 저장하므로 고정된 형식과 길이를 갖는 데이터가 필요한 반면 XML 전용 데이터베이스는 XML 객체나 요소들을 원래의 형태로 저장할 수 있다.

- 전문 검색/질의: 관계형 데이터베이스는 데이터가 잘 정의된 형식의 필드 길이를 갖기를 요구하기 때문에 새로운 형태의 풍부한 다양성을 다룰 수 없고 데이터를 BLOB 로 저장할 수는 있지만 데이터의 검색이나 질의, 인덱싱이 어려운 반면 XML 전용 데이터베이스는 저장 매체로부터 데이터를 얻어내기 위한 전문검색 및 질의(Full-text search/query)가능하다.

- 라운드 트리핑: 라운드 트리핑이란 XML 문서 구조를 그대로 복원할 수 있는 기능을 말한다. 관계형 데이터베이스는 요소의 순서, 속성 등 XML 문서의 구조를 보존하기 어려운 반면 XML 전용 데이터베이스는 완벽한 라운드 트리핑을 지원한다.

관계형데이터베이스와 XML 전용 데이터베이스를 비교한 결과 중앙의 임상 데이터베이스에 저장된 XML 기반의 CDA 문서의 원본을 유지하기 위해서는 관계형 데이터베이스보다는 XML 전용데이터베이스가 보다 효과적인 저장 및 검색 시스템이 될 수 있음을 알 수 있다.

2.3 NXD (Native XML Database) 기술 동향

XML 전용 데이터베이스는 XML 데이터 모델을 지원하는 데이터베이스관리시스템 (DBMS: Database Management System)를 의미한다. NXD 는 XML 문서를 저장의 기본 단위로 하고, XML 문서 구조를 이용해서 질의어를 사용할 수 있으며 다음과 같은 특징을 갖는다 [3].

- XML 문서의 논리적인 모델을 정의하고 이 모델에 따라 문서를 저장하고 검색한다.

- XML 문서를 논리적인 저장단위로 갖는다.

- 어떠한 형태의 물리적인 저장 모델도 요구하지 않는다.

- 질의어 결과는 XML 문서이다.

HL7 메시지에 담긴 정형화 되지 않은 바이오 및 임상정보들을 통합하기 위해서는 통합에 필요한 정형화된 데이터 구조가 미리 정의되어야 한다. 바이오 및 임상정보를 XML 기반의 데이터구조로 통합하여 XML 전용 데이터베이스에 저장하는 경우 다음과 같은 장점을 가질 수 있다 [4].

- 작성된 XML 문서가 미리 정의된 데이터 구조의 규칙 즉, XML 스키마를 잘 따르고 있는지를 체크할 수 있다.

- 미리 정의된 데이터 구조를 기반으로 질의어를 작성할 수 있기 때문에 XML 문서의 구조를 파악하는 과정을 생략할 수 있다.

● XML 전용 데이터베이스는 XML 문서 전체가 한 곳에 저장되기 때문에 여러 테이블에 저장되는 관계형 데이터베이스에 비해서 검색속도가 훨씬 빨라지게 된다.

위와 같은 장점 때문에 XML 전용 데이터베이스를 이용한 XML 기반의 통합데이터 관리는 미리 정의된 통합 데이터 구조와 정확히 일치하는 XML 문서의 생성을 보장하며 미리 정의된 데이터 구조를 통해 쉽게 질의어를 작성할 수 있다. 뿐만 아니라 방대한 바이오 및 임상정보를 검색하는 시간을 단축할 수 있다.

2.4 XML 기반 기술

XML 문서의 구조를 기술하기 위해 XML 스키마를 이용하고 문서의 보안을 위해서는 XML 서명 기술을 이용한다. 두 기술 모두 인터넷상의 문서 표준을 제정하는 W3C (World Wide Web Consortium)에 의해 제정된 기술이다.

1) XML 스키마

DTD (Document Type Definitions)는 기존의 마크업 언어인 SGML 을 간단한 형태로 만든것 이다. 그러나 DTD 의 문법은 데이터 타입의 제한, 재사용성과 확장성의 부족, 데이터 전송과 같은 부분에서 XML 사용자들의 요구사항을 만족시키지 못한다. XML 스키마는 DTD 가 갖는 단점을 개선하기 위해 W3C 가 제안한 표준기술로 DTD 에 비해 다음과 같은 장점을 갖는다 [5].

- XML 스키마는 XML 문법으로 작성되므로 XML 문서 처리에 필요한 SAX [6], DOM[7], XSLT[8] 같은 관련기술들을 사용할 수 있다.
- 프로그래밍 언어에서 쓰이는 데이터형을 지원하므로 각자의 고유한 데이터형을 만들 수 있다.
- 내용 모델과 같이 마크업 구조물들을 확장하거나 재사용할 수 있는 명시적 방법을 허용하는 강력한 클래스 및 형 정의 시스템을 제공한다.
- 복수의 XML 네임스페이스를 지원하므로 다른 네임스페이스에 속하는 이미 정의된 스키마의 구성요소를 재사용할 수 있다.

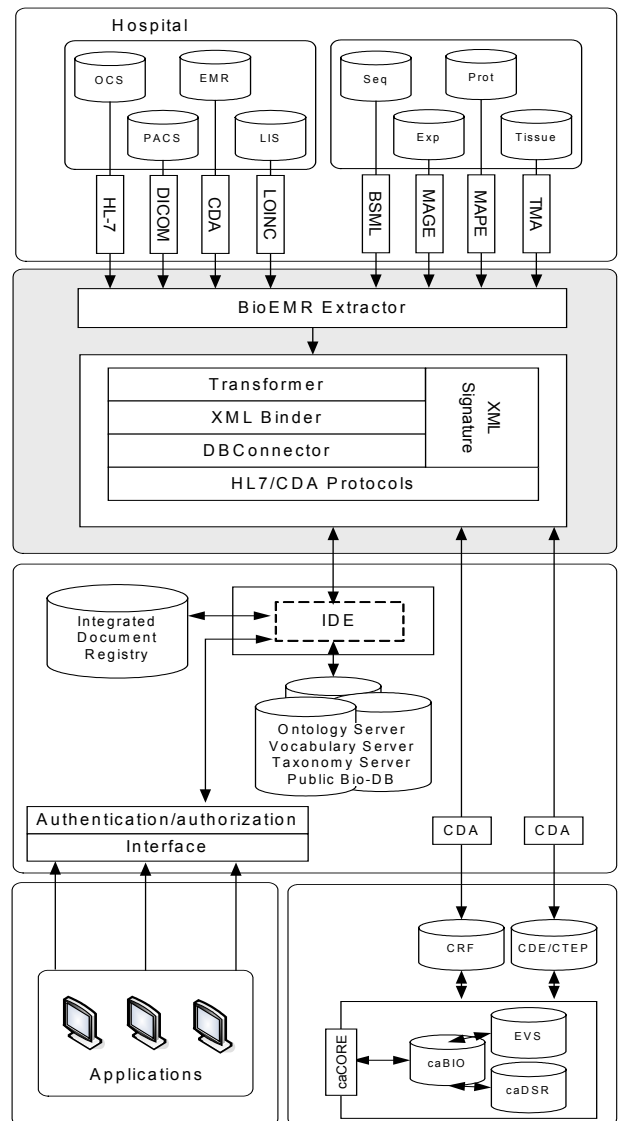
스키마는 문서의 구조만을 정의하고 실제 문서 내용에 대해서는 정의하지 않는대신 XML 문서가 XML 스키마의 규칙을 따르는지를 확인하기 위해서 XML 파서를 사용한다. 따라서 XML 스키마를 이용하여 바이오 및 임상정보들을 통합하기위한 CDA 를 설계 하고 이질적인 시스템들로부터 추출된 바이오 및 임상 데이터들을 통합한 XML 문서가 CDA 의 구조와 일치하는지의 여부를 판단하여 정확한 통합 데이터를 구축 할 수 있다.

2) XML 전자서명

서명은 데이터의 무결성, 사용자 인증 그리고 부인 봉쇄를 제공한다 [9]. 즉, 디지털 콘텐츠가 불법적으로 위조 및 변조되지 않았음을 보장하고, 메시지의 송수신자가 올바른 사용자임을 확인할 수 있게 하는 인증

방식을 의미한다. XML 전자서명은 XML 의 장점인 구조적 정보 표현 능력 및 확장성을 기반으로 한 전자서명 기술로서 현재 ebXML [10] 메시지의 인증 및 무결성을 위한 표준 기술로 적용되고 있으며 웹 서비스와 함께 XML 기반의 시스템 프레임워크에 다양하게 활용될 수 있다. 특히, XML 서명은 XML 을 포함하는 모든 디지털 콘텐츠에 적용할 수 있다 [11]. 따라서 XML 전자서명은 EMR 연동 시스템에서 교환되는 바이오 및 임상정보를 포함하는 XML 기반의 메시지들에 대한 임의의 변경이나 수정이 가해지는 것을 방지할 수 있는 기술을 제공한다.

3. 바이오 및 임상정보 통합 저장소를 위한 미들웨어의 설계



(그림 1) 바이오 및 임상정보 통합 저장소를 위한 미들웨어

그림 1 은 병원 레거시 시스템에 보관된 의료정보와 DNA 칩이나 단백질 등과 같은 바이오 정보와 환

자의 임상정보를 통합하기 위한 미들웨어이다. 미들웨어의 주요 역할은 이질적인 시스템들로부터 의료 및 바이오 정보를 추출해내고 추출된 정보들을 통합하기 위해 정의된 새로운 CDA (Clinical Document Architecture)를 따르는 통합된 데이터를 생성하여 모든 병원들이 이용할 수 있는 중앙의 임상 데이터베이스 (Integrated Document Registry)에 저장하는 것이다.

미들웨어의 각 모듈에 대한 설명은 다음과 같다.

● **BioEMR Extractor**: OCS, PACS, EMR, LIS 같은 병원 시스템에 의존적인 형태로 존재하는 의료정보 메시지와 패턴 인식을 통해 DNA 칩, 단백질, 세포조직 등으로부터 얻은 임상 유전자 정보로부터 통합에 필요한 정보들을 HL7 등 해당 시스템의 표준기술에 기반하여 추출해내는 역할을 한다.

● **Transformer**: BioEMR Extractor를 통해 얻은 다양한 시스템특징적인 정보들을 통합된 데이터로 변환하는 역할을 한다.

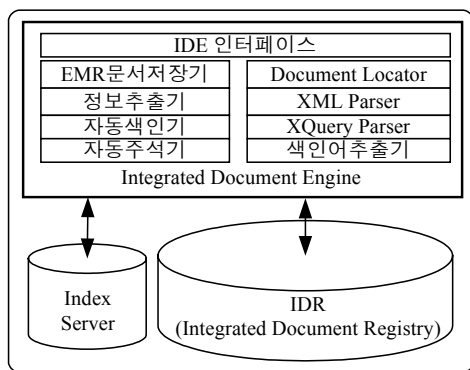
● **XML Binder**: Transformer를 통해 통합된 XML 문서를 응용프로그램이 보다 쉽게 처리할 수 있도록 바이오 및 임상정보 통합을 위해 정의된 XML 스키마를 기반으로 XML 문서를 객체로, 객체를 XML 문서로 변환해주는 역할을 한다.

● **DBConnector**: 통합된 정보를 통합 문서 저장소 (Integrated Document Registry)에 저장하기 위해 데이터베이스와의 통신을 돕는다.

● **XML Signature**: 통합된 정보를 교환하는 중에 발생할 수 있는 데이터의 손실 및 임의적인 변경을 막고 데이터 송수신자를 상호인증하는 역할을 한다.

● **HL7/CDA Protocols**: 통합 정보 전송을 위한 프로토콜의 사용을 용이하게 하는 역할을 한다.

● **IDE (Integrated Document Engine)**: 미들웨어를 통해 통합된 정보를 통합 문서 저장소에 저장하고 외부 응용프로그램을 통해 통합 문서 저장소에서 관리되는 문서에 접근을 제어하는 역할을 한다.



(그림 2) 통합 문서 처리 엔진

그림 2는 그림 1에 보여진 IDE (Integrated Document Engine)의 구조이다. IDE는 통합 문서를 통합 문서 저장소 (IDR)에 저장하고 검색하는 기능을 제공한다. XML 기반의 문서를 효과적으로 처리하기 위해 XML 문서의 구문을 분석하고 XML 데이터의 질

의어를 분석하는 기능을 제공한다. 또한 XML 문서를 저장할 때 저장되는 문서들에 대한 색인을 부여하여 색인서버에 보관하는 기능을 제공하고 XML 데이터를 검색할 때는 색인 서버에서 추출한 색인어를 기반으로 검색을 수행함으로써 검색속도를 향상 시키는 기능을 제공한다.

4. 결론 및 향후과제

최근 진료의 질적 향상을 위한 바이오 및 임상정보 통합의 요구에 따라 해당 분야별로 표준기술들이 구현되어 있다. 그러나 각 표준간 상호운용성이 결여되어 각 표준을 구현한 시스템들로부터 데이터를 통합하는 일은 매우 복잡하다. 따라서 각 표준간 상호운용성을 확보하기 위해 XML 기반의 새로운 데이터 구조를 정의하고 이를 저장하기 위한 통합 저장소의 역할이 보다 중요해졌다. 본 논문에서 제시한 바이오 및 임상정보 통합 저장소를 위한 미들웨어는 XML 기반의 의료정보 통합을 위한 기능을 제공하고 다양한 시스템들로부터 통합된 정보를 통합저장소에 저장하는 기능과 시스템 간 자료전송 그리고 외부의 응용프로그램을 통한 통합저장소의 접근제어 기능을 수행하는 기능을 제공하여 다양한 형태의 의료정보 메시지를 교환하고 통합할 수 있는 가능성을 제시하고 있다.

향후 연구에서는 현재 설계된 미들웨어를 기반으로 바이오 및 임상문서의 표준기반 EMR (Electronic Medical Records) 연동엔진 개발, XML 기반 EMR 문서 저장기, 정보추출기, 자동색인기, 자동주석기 개발에 대한 연구가 지속될 것이다.

참고문헌

- [1] 이민경, 정재현, 전종훈, 유수영, 김보영, 최진욱, "The LEX System : HL7을 사용하는 전자의무기록의 효율적인 교환과 공유를 위한 XML 기반 통합의료 환경의 구축", 정보처리학회, 2002
- [2] L. Alschul, R. H. Dolin, Clinical Document Architecture Framework release 1.0, hl7, 2000
- [3] <http://www.xmldb.org/>
- [4] T. W. Leung, "Professional XML Development with apache tools", Wrox, 2004.
- [5] W3C Standard., "XML Schema Part 0: Primer Second Edition", <http://www.w3.org/TR/xmlschema-0/>
- [6] SAX, <http://www.saxproject.org/sax1-roadmap.html>
- [7] W3C Standard., "Document Object Model (DOM)", <http://www.w3.org/DOM/>
- [8] W3C Standard., "eXtensible Stylesheet Language Transformations (XSLT)", <http://www.w3.org/TR/xslt>
- [9] W3C Standard., "XML-Signature Requirements", <http://www.w3.org/TR/xmlsig-requirements>
- [10] ebXML, <http://www.ebxml.org>
- [11] W3C Standard., "XML-Signature Syntax and Processing", <http://www.w3.org/TR/xmlsig-core/>