

Biological Data의 저장 및 관리를 위한 BioEMR Extractor의 설계 및 구현

이태현^o 정종일 유의혁 김강 신동규 신동일
세종대학교 컴퓨터공학과

{leeth, jijeong, solui, kimkang, shindk, dshin}@gce.sejong.ac.kr

Design and Implementation of the BioEMR Extractor for Storing and Managing Biological Data

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

요 약

최근 인터넷을 기반으로 하는 웹의 급속한 발전으로 정보는 인터넷을 통해서 교환되고 공유될 수 있는 개방된 환경에서 더 큰 가치를 지니게 되었다. 이러한 경향은 의료분야에서도 예외 없이 적용되어 지역 네트워크기반의 기존 의료정보 시스템도 타 기관과의 데이터 공유의 필요성으로 인해 개방형 시스템으로의 전환이 요구되고 있다. 따라서, 의료정보의 교환과 공유를 위해 각 의료관련 시스템들에 대한 표준들이 제정되어 적용되고 있다. 그 중 HL7은 임상문서 교환을 위한 유연한 표준이지만 시스템 의존적이거나 이질적인 시스템간의 정보교환에는 부적합하다. 따라서 본 논문은 이질적인 시스템간의 정보교환 및 공유를 위해 바이오 및 임상정보를 통합하고 통합된 정보를 저장할 수 있는 기능을 주로 하는 BioEMR Extractor를 구현했다. 또한, 이질적인 시스템 통합의 첫 단계로서 RDB기반의 기존 BioEMR 정보를 XML기반의 Data로 변환하는 모듈을 설계 및 구현했다.¹⁾

1. 서 론

최근 인터넷을 기반으로 하는 웹의 급속한 발전으로 이제 정보는 인터넷을 통해서 교환되고 공유될 수 있는 개방된 환경에서 더 큰 가치를 지니게 되었다. 이러한 경향은 의료분야에서도 예외없이 적용되어 지역 네트워크를 기반으로 운영되어오던 기존의 의료정보 시스템도 병원 및 타 기관들과의 데이터 공유의 필요성으로 인해 개방형 시스템으로의 전환이 요구되고 있다. 따라서, 의료정보의 교환과 공유를 위해 각 의료관련 시스템들에 대한 표준들이 제정되어 발전하고 있다. HIS(Hospital Information System)내의 정보를 교환하기 위한 텍스트 기반의 HL7(Health Level 7)을 비롯해, PACS(Picture Archiving & Communication System)를 위한 DICOM(Digital Imaging and Communication in Medicine), EDIFACT X.12 등의 표준이 그 예이다.

최근 바이오 및 임상정보의 통합 필요성이 점차 커지고 있는 상황에서 디지털병원 구축 확산은 지역 네트워크 기반의 폐쇄적인 시스템에서 관리되어오던 바이오 및 임상정보 통합에 필요한 환경을 제공함으로써 이질적인 시스템들로부터 얻은 데이터를 통합 관리할 수 있는 통합 저장소의 구축을 보다 용이하게 할 수 있다. 특히 EMR 시스템에서의 바이오 및 임상정보 통합은 환자들의 체질적이거나 유전적인 질병을 진료기록과 통합할 수 있어 일선 병원에서의 진료의 질 향상에 크게 기여하게 된다.

기존의 HL7메시지로 표현되는 바이오 및 임상정보를 교환하고 통합하기 위해 기존 시스템의 데이터 구조에 독립적인 형태로 통합에 최적화된 CDA의 설계 및 개발이 반드시 필요하다.

정형적이며 구조적인 데이터를 교환하는데 XML문서가 적합하긴 하지만 또 다른 문제는 HIS의 데이터 종류(DB 스키마)에 따른 의료정보의 다양성이다. 현재 각 병원들은 이질적인 HIS를 가지고 있기 때문에 같은 이름의 메시지라 하더라도 병원 자체 DB 스키마에 따라 이질적인 형태의 HL7 메시지가 생성되기 때문에, 이들 간에 직접적인 메시지 교환이 어렵다. 뿐만 아니라 현재 각 병원마다 저장되어 있는 개인 정보와 의무 기록들이 재정적, 시간적 낭비 없이 효율적으로 이용되기 위해서는 모든 병원들이 이용할 수 있는 중앙의 임상 데이터베이스가 필요하다.

이러한 의료정보통합 연구의 필요성을 다음과 같이 정리할 수 있다.

- 의료정보화의 진행으로 인한 EMR 및 임상연구 정보 시스템 구축 확산
 - 환자의 진료, 치료시 개인의 유전적인 특성을 고려한 개인별 맞춤의학 시대 도래
 - 인간 유전체 사업의 완성과 첨단 생명과학의 발전으로 High Throughput 바이오 데이터의 획득 및 바이오 데이터 처리의 국제 표준화의 급속한 진행
 - 고속 대용량 정보처리 기술 발달
- 본 논문에서는 의료정보통합을 위한 전체적인 목표시스

1)본 연구는 보건복지부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임.(0412-M101-0416-0002)

템을 제안하고 제안한 시스템 설계 중 Biological Data의 저장 및 관리를 위한 BioEMR Extractor을 설계, 구현한다.

2. 연구의 배경 및 필요성

2.1 EMR(전자의무기록)

의무기록전산화가 추구하는 방향은 CPR과 같은 실현에 상당히 많은 시간을 요하는것과는 반대로 먼저 병원내의 의무기록전산화를 중심으로 하여 개발하여 사용하는 것을 말한다. CPR과의 차이점은 CPR은 전국 어디서나 모든 병원의 환자에 대한 진료 정보를 이용하는 것에 주안점을 두는 것인데 비해 EMR은 병원내의 의무기록 전산화에 주안점을 둔 보다 협의의 개념이라 할 수 있다.

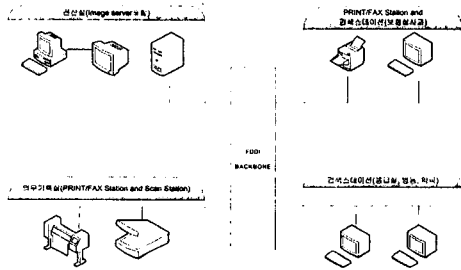


그림 1 EMR 시스템 구성도

EMR은 병원에서 사용되는 종이 문서를 없애고 모든 데이터를 전산매체에 저장하는 방식이다. 현재까지 널리 사용되고 있는 종이 의무기록은 오랜 동안 모아지면 언젠가는 관리 및 보관상에 한계가 오기 때문에 의무기록을 영구보존하는 대부분의 병원에서는 진료 후 일정기간이 지난 비활용 의무기록을 과거에는 마이크로 필름에, 최근에는 광디스크, 컴팩트디스크 등에 수록하여 관리하고 있다. 이들을 전산과 연계 시키고 네트워크를 연결하면 이미지로만 구성된 전자의무기록이 보관 및 검색 목적으로 제한적으로 활용될 수 있는 초기 단계의 전자의무기록이 구성된다.

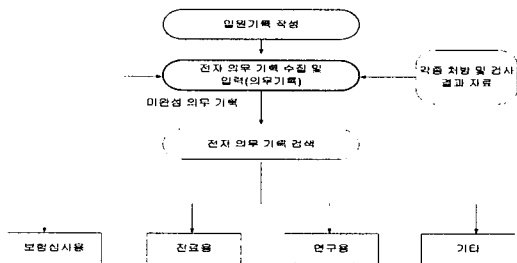


그림 2 EMR 시스템의 업무 흐름

국내에서는 90년대 들어 신설된 병원들을 주축으로, 전 병원을 네트워크환경으로 구축하고 전산시스템도 원무 및 보험청구 중심에서 한단계 발전하여 처방전달시스템 (Order Communication System : OCS)을 도입하는 형

태로 병원전산환경이 발전하였다. 그 결과 OCS 환경에서는 환자의 인적사항, 처방내역, 검사결과 등이 이미 텍스트형태로 입력되어 진료 중에 활용하고 있기 때문에 의료진이 기재하는 진료기록만을 전산에 입력하여 보다 진보된 전자의무 기록이 구축될 수 있게 되었다. 따라서 기록내용이 많은 부분만을 편법으로 수기한 후 스캐닝하거나 전산요원에 의해 입력하여 한 차원 높은 수준의 전자의무기록인, Electronic Medical Record (EMR)을 구현하였다[1].

2.2 마이크로어레이 데이터베이스의 필요성과 기능

마이크로어레이 기술이 발전함에 따라 생물학자들은 한번의 실험에 수많은 데이터를 관리해야하는 번거로움이 생겨났다. 이러한 데이터 관리를 용이하게 하기 위한 데이터베이스시스템이 생겨났으며, 그것을 이용하여 중앙관리식으로 데이터를 저장, 관리, 분석할 수 있게 되었다. 이러한 데이터베이스 시스템은 주로 웹서버에 데이터를 관리하며 사용자는 웹을 통해 데이터를 등록, 검색할 수 있게 된다[2]. 본 논문에서는 TMA(Tissue Microarray)의 데이터베이스 시스템[3]을 조사하고 그 구조를 통합하는 스키마를 추출 한다.

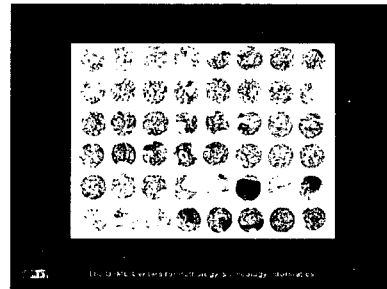


그림 3 TMA Slide

```
<?xml version="1.0" ?>
- <histo>
- <tma>
  <header />
  - <block>
    <slide />
    <core />
  </block>
</tma>
</histo>
```

그림 4 XML Format으로 표현한 TMA 구조

2.3 BSML(Bioinformatic Sequence Markup Language)

BSML은 의미있는 생물학적 관계를 얻기 위한 방법으로 바이오인포매틱스 연구정보에 개방된 XML언어이다. DTD에 의해 표현되는 XML의 응용으로써 사용자가 조정할 수 있는 표나 그림들로서 서열정보를 보여준다. 특히 서열의 표현뿐만 아니라 생체 물리화학적 특성까지 표현 가능하다.

이러한 장점을 보유한 BSML은 지놈(genome)정보의 교환 및 표현, 응용 분야에 사용된다[4].

```
<!DOCTYPE Bsm1 SYSTEM "BSML2_2.DTD">
<Bsm1>
<Attribute name="project" content="sample document"/>
<Attribute name="author" content="Simon St.Laurent"/>
...
```

그림 4 Metadata를 제공하는 BSML Document의 예제

3. 제안하는 시스템 설계 및 구현

본 논문에서는 Bio정보 통합을 위한 전체 시스템을 설계하고 그 시스템의 첫 단계로써 TMA-DTD 요구사항을 분석하여 이중 스키마간의 converting을 위한 모듈을 설계, 구현하였다.

3.1 Biological Data 통합을 위한 전체 시스템 구성

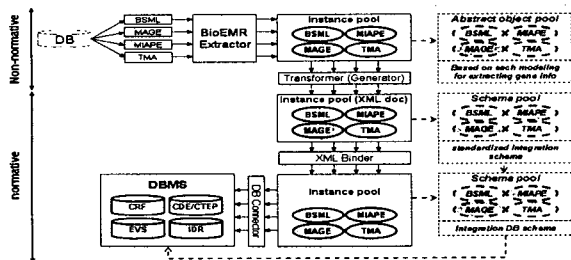


그림 5 Biological Data 통합을 위한 시스템 구성도

그림 5는 전체적인 시스템 구성도로써 크게 Non-normative(비-표준)와 Normative(표준)로 분류하였다. 실질적인 스키마로 구성된 Non-normative 부분에서 바이오정보 통합을 위한 통합된 새로운 CDA(Critical Document Architecture)를 따르는 데이터(Normative부분)를 생성하여 DBMS에 저장하는 시스템을 설계하였다.

전체적인 시스템 흐름은 다음과 같다.

- ① BioEMR Extractor : 실질적인 스키마에서 통합에 필요한 데이터 추출
- ② Transformer : BioEMR Extractor에서 얻은 정보를 통합된 데이터로 변환
- ③ XML Binder : Transformer에서 얻은 정보를 정의된 XML Schema기반으로 변환
- ④ DB Connector : 통합된 정보를 DBMS에 저장

3.2 TMA Generator 설계 및 구현

구현의 첫 단계로 BioEMR Extractor의 한 부분인 TMA에 대한 Generator(그림 6)를 설계하고 구현하였다. TMA Generator설계 및 구현을 위해 ACSR(AIDS and Cancer Specimen Resource)[5]에서 제시한 TMA-DTD의 Histo, TMA, Header, Block, Slide, Core에 대한 80여가지의 요구사항을 분석하여 정의된 element를 생성해내는 모듈을 구현하였고, 요구사항 분석에 대한 구현은 사용자 논리에 따라 각 모듈을 조합하는 방식으로 구현하였다.

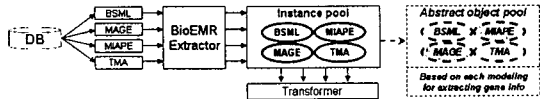


그림 6 BioEMR Extractor

그림 6에서 DB에 RDB기반으로 정의되어진 BioEMR 데이터를 XML문서로 변환하는 모듈을 Java기반으로 구현하였고, DOM API를 이용하여 내용을 추출한 후 Xerces를 이용하여 데이터추출과 동시에 Valid-check를 하도록 구현하였다.

그림 7은 구현한 XML Generator를 이용 RDB에서 XML기반으로 TMA를 추출한 문서의 일부이다. LDE(Local Data Element)의 정의는 사용자 혹은 Bio정보를 수집하는측에 의존적이기에 이 구현에서는 제외하였지만 개별적인 모듈을 확장해서 구현이 가능하도록 하였다.

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml?>
<?header?
<!--TMA blocks Available for Research form ODU ACSR Mid-Region/>
<header?
<!--
<!-- identifier TAO-050-Child identifier
<!-- purpose: f- cardiac disease, f- HIV 1 of 10<Ancl purpose>
<!-- description: heart tissue from 7 Autopsy patients</local_description>
<!-- meta numbers: heart TAOB-050/001_001</meta_numbers>
<!-- related blocks: TAOB-050 TAOB-051 TAOB-052 TAOB-053 TAOB-054 TAOB-055 TAOB-056 TAOB-057 TAOB-058 TAOB-059</related_blocks>
<!-- creation: f- histology_creator
<!-- creation_date: 2001-11-16</creation_date>
<!-- protocol: manual</block_protocol>
<!-- number_of_cores: 2</block_number_of_cores>
<!-- core: jv12</block_core_core>
<!-- core_spacing: 4</block_core_spacing>
<!-- assay: histol inspection</block_assay>
<!-- status: Production</block_status>
<!-- meta_numbers: 01A 01B 01C 01D 01E 01F 01G 02A 02B 02C 02D 02E 02F 02G 03A 03B 03C 03D 03E 03F 03G 04A 04B 04C 04D 04E 04F 04G 04H 04I 04J 04K 04L 04M 04N 04O 04P 04Q 04R 04S 04T 04U 04V 04W 04X 04Y 04Z 05A 05B 05C 05D 05E 05F 05G</block_meta_numbers>
</!--
<!-->
<!--this is the first example of header element</header?
<!--this is the first example of header element</header?
<!--this is the first example of header element</header?
</-->
```

그림 7 Generator를 이용 추출한 TMA문서

4. 결론 및 향후 과제

의료정보분야의 급속한 발전에도 불구하고 각 분야별로 실질적으로 존재하는 개별적인 표준을 구현한 시스템들로부터 데이터를 통합하여 운영하는 것은 상당히 어려운 작업이다. 따라서 각 표준간의 상호운영성을 확보하기 위해 XML 기반의 새로운 표준의 필요성이 대두되었다.

본 논문에서는 임상정보와 바이오정보 통합을 위한 전체적인 시스템을 구성하였고, 그 첫 번째 단계로써 TMA를 포함한 Critical Data를 converting하는 모듈을 설계, 구현하였다.

향후 연구에서는 현재 설계한 시스템을 기반으로 미 구현된 모듈의 계속적인 구현과 함께 통합된 바이오 및 임상문서의 저장과 관리를 위한 대용량 XML 문서 저장관리시스템에 대한 응용연구가 지속 될 것이다.

참고문헌

- [1] 김윤년, 전효찬, 김윤정, 박황백, 김도형 "Web을 이용한 회송환자에 대한 EMR 공유 시스템"
- [2] 이미경 "Survey on MicroArray Database System"
- [3] Jules J Berman "The tissue microarray data exchange specification: implementation by the Cooperative Prostate Cancer Tissue Resource"
- [4] <http://bsml.org>
- [5] <http://www.acsr.mid-region.org/>
- [6] W3C Standard., "XML Schema Part 0: Primer Second Edition", <http://www.w3.org/TR/xmlschema-0/>
- [7] W3C Standard., "Document Object Model (DOM)", <http://www.w3.org/DOM/>
- [8] <http://www.xmldb.org/>