

Development of Conversion Tools from V2 to V3 based on caAdapter

엄기성* · 김화선** · 홍해숙[†] · 조훈***

(Ki Sung Um · Hwa Sun Kim · Hae Sook Hong · Hune Cho)

Abstract - Although the goals of HL7 Version 2(V2) and Version 3(V3) are identical, the concepts of the implementation and technological basis are different; this makes their versions inconvertible. This problem interrupts technological innovation advanced from V2 to V3 and has been raised as a new type of barrier in the field of medical information system. This study intends to develop software to convert V2 to V3 which can be utilized in the actual medical environment. Since it is practically difficult to develop the whole tools that automatically change total V2 messages into V3 messages, this article has designed, implemented, and tested the software that allows mapping between V2 and V3 which function as tools. In order to test this program, it has used ADT^A03 of five V2 messages to generate V3 messages. It is expected that the result of this research will be one of the new methods allowing conversion between V2 and V3.

Key Words : HL7 Version 2, HL7 Version 3, Medical Information System, Conversion

1. 서 론

의료정보시스템간의 직접적인 정보 교환 수단인 부재는 사용자로 하여금 별도의 시간과 노력을 투입하여 기존에 작성된 기록을 인쇄하고 재입력하는 등 추가적으로 자원의 소모와 낭비를 감수하게 한다. 이와 같은 문제점들을 해결하기 위한 방안으로 HL7은 한 의료기관 내 또는 서로 다른 독립적인 정보시스템 간의 구조적 변경과 재개발의 표준화, 데이터의 표준화, 데이터 전송의 각 의료 기관간의 정보 공유 및 활용을 위한 표준화에 대하여 연구하고 있다[1][2].

HL7의 본래 역할은 임상진료와 헬스케어 서비스의 관리, 중재, 평가를 지지할 수 있도록 임상자료의 교환, 관리 및 통합을 위한 표준을 제공하는 것이다[1][3]. HL7은 1987년에 개발 당시 기술로 개발됨으로 인해 필드 사이의 구분을 위해 ‘|’, ‘^’ 등을 사용하는 체계를 가지고 있다. 그러므로 반드시 필요한 부분에서 ‘|’를 누락하거나 추가하게 되므로 의미가 있는 메시지를 무의미한 스트림으로 변경될 수 있다. 정보모델 중에서 가장 성공적인 모델 중 하나인 HL7 버전 2(이하 V2)는 호환성(plug-and-play connectivity)과 시맨틱 상호운용성(Semantic Interoperability)을 완벽히 제공하지 않는 단점이 있다[4][5].

이를 해결하기 위해, HL7 버전 3(이하 V3)가 개발되었다. 이 표준은 확장의 관점에서 V2와 비교해 볼 때 새로운 개념에서 출발하였다. 이 새로운 버전은 독립적 시스템 사이의 시맨틱 상호운용성에 초점을 맞추고 시맨틱 웹(Semantic Web)의 기술 및 목적과 일치하며, 의료정보시스템뿐만 아니라 분산형 시스템의 플랫폼 표준으로 사용될 수 있다.

HL7 V2와 V3의 목적은 동일하지만 구현 개념과 기술적 기반이 다르므로 버전 간의 호환이 불가능하다. 이러한 문제는 V2에서 V3로의 기술적 진보를 단절시켜 의료정보시스템분야에 새로운 장벽으로 부각되고 있다.

시맨틱 상호운용을 위한 사용자들의 요구에 의해 V2에서 V3로 기술 전이가 있을 것으로 예상하고 있으나, 아직까지 수많은 기존 데이터와 시스템은 V2 메시지를 사용하고 있다. 실제로 현재 의료정보 분야에서 몇몇 새롭게 개발된 중진 시스템 혹은 실험적 운영을 위한 경우를 제외하고는 V3를 실제 적용하지 못하고 있다. 그럼에도 불구하고 미국의 caBIG(cancer Biomedical Informatics Grid) 프로젝트와 NHIN (Nationwide Health Information Network)와 같이 새롭게 개발되고 있는 시스템 또는 네트워크를 중심으로 V3의 요구는 점진적으로 높아지고 있다[6]. V3에 대한 사용자의 요구는 기술적 흐름으로 매우 자연스러운 현상이나 V2에 기반 한 기존 시스템의 데이터를 어떻게 V3메시지로 표현할 것인가는 매우 중요하고 어려운 문제이다[6].

그래서 본 연구는 실제 의료 환경에서 활용할 수 있는 V2에서 V3로 변환하기 위한 소프트웨어를 개발하는 것을 목표로 하였다. V2 메시지 전체를 V3 메시지로 자동 변환하는 도구 전체를 개발하는 것은 현실적으로 어려우므로 선행 도구로서의 기능을 하는 V2와 V3 간에 개별적으로 맵핑시킬 수 있는 소프트웨어를 설계, 구현 및 테스트 하였다.

[†] 교신저자, 정회원 : 경북대학교 간호대학 간호학과 교수

E-mail : hshong@knu.ac.kr

* 비 회원 : 미국 NIH 객원연구원

** 정 회원 : 경북대학교 의료정보학과 연구교수

*** 비 회원 : 경북대학교 의료정보학과 교수

접수일자 : 2010년 1월 28일

최종완료 : 2010년 3월 5일

2. 시스템 설계 및 기술

2.1 개발환경과 도구

‘원천기술공개’ 원칙을 지키는 caBIG의 모든 산출물은 자바(Java)기술에 기반하고 있으므로 본 개발 역시 호환성을 위해 Java 도구들로 개발하였다. 개발 언어는 Java SDK 1.5, 빌딩 도구(Building Tool)로는 IntelliJ IDEA 7.0을 선택하였다. 하드웨어 환경은 개발과 테스트를 위해서 MS 윈도우 XP 기반의 Pentium 2.7 GHz를 사용하였다[7][8].

2.1.1 caAdapter

본 연구는 caAdapter의 부분 프로젝트로서 진행되고 있으며, caAdapter 프로젝트는 caBIG(cancer Biomedical Informatics Grid) 프로젝트의 하부 프로젝트로서, 이미 널리 알려진 바와 같이 caBIG은 미국정부주도 개발 프로젝트로서 기초과학연구와 임상정보의 교환을 개선하기 위해서 2004년 미국국립암센터(National Cancer Institute)와 미국바이오정보센터(National Center for Bioinformatics)의 정보교환을 위한 정보 인프라 구축을 위해 시작하였다[9].

caBIG 프로젝트는 확장된 다양한 콘텐츠를 개발하기 위해서 그리드 컴퓨팅, caAdapter와 같은 다양하고 새로운 기술들을 적용하고 있다. caBIG의 하부 프로젝트인 caAdapter는 caBIG 프로젝트 시작일과 동시에 진행되었으며, V3 메시지를 지원하기 위한 개발을 하고 있다[10].

caAdapter의 본래의 목적은 단지 CSV(Comma Separated Values) 메타 파일과 V3 MIF(Message Interchange Format)에서 추출된 H3S라고 불리는 V3 구조 정의 간에 맵핑을 통해서 V3 메시지를 생성하는 것이다. 사용자의 대부분은 V2 메시지를 주로 사용하고 있기 때문에 기술적으로 향상된 V3로의 업그레이드의 요구가 높아지고 있다. 그러므로 V2에서 V3로의 자동적인 변환에 대한 관심과 요구가 높아졌다. 또한 caBIG에서도 V3를 사용할 것을 권장하고 있다. 그래서 caBIG 프로젝트의 원칙은 “공개 소스”로서 전체 소스 코드와 관련 개발 문서들을 웹 사이트에 공개되어 관심이 있는 사용자들은 공유하고 무료로 다운로드 할 수 있도록 하고 있다. 단 다운로드 된 소스는 상업용 소프트웨어나 유료 도구로 사용하는 것을 정부차원에서 금지하고 있다 [11].

2.1.2 KNU V2 Parser

caAdapter는 초기 V2에 대해 지원을 하지 않는다. 그래서 V2 메시지 파싱을 위해서 2007년, 2008년 전기학회에 발표하였던 KNU V2 도구를 적용시켰다[2][12]. KNU HL7 Parser는, HL7 메시지를 입력 받아 다수의 세그먼트로 파싱하는 메시지 파싱부, 상기 다수의 세그먼트 각각을 다수의 필드로 파싱하고, 상기 다수의 필드를 컴포넌트로 파싱하는 세그먼트 파싱부, 다수의 컴포넌트 각각을 미가공 데이터로 복원하는 필드 파싱부를 포함하는 것을 특징으로 한다. 이는 정보를 처리하는 가장 작은 객체인 원시데이터유형(primitive data type)과 복합데이터유형(composite data type)을 유효하게 하기 위해서 유형 인터페이스를 참조한다. 아울러 그룹(group), 세그먼트(segments) 및 자료 유형(data

type)내로 HL7 메시지들을 수신한다. 메시지, 그룹, 세그먼트, 데이터 유형들 모두 HL7 정의에 의한다. HL7 정의는 메시지를 생성하는 외에도 내부의 메시지 유효성 검사 함수를 이용하여 HL7 메시지의 오류여부를 검사한 뒤 다음 컴포넌트로 이동하는 메커니즘을 가지고 있다[2]. V2 메시지 입력 크기는 제한을 두지 않았다. 적용된 KNU V2 도구의 장점은 V2의 중요한 세그먼트 중 하나인 ‘OBX’ 세그먼트와 같은 다형적 성격을 가진 V2 메시지를 프로세싱 할 수 있다.

2.2 구현 설계

2.2.1 caAdapter의 V3 생성 메커니즘

그림 1은 caAdapter가 V3 메시지를 생성을 위해 4개의 파일이 필요하다는 것을 알 수 있다. 4가지 파일은 메시지 정보 유형 파일인 MIF(Message information format)파일로부터 추출되는 버전 H3S(HL7 V3 Specification), MAP(Mapping File) 파일과 SCS(Source Specification), CSV(Comma Separated Value)파일이다. SCS는 CSV 파일 유형의 데이터의 구조를 나타내고 있으며, caAdapter는 SCS 파일을 생성하기 위한 GUI 환경을 제공하고 있다.

H3S 파일 또한 메타 파일이며 V3 메시지 구조를 나타내고, caAdapter의 H3S 패널기능에서 MIF 파일의 메시지 분기(branch)를 가지치기(pruning) 하거나 이식(grafting) 과정을 통해 생성할 수 있다.

HL7 정의된 메시지 유형(Defined Message Type)의 경우, MIF 파일은 HL7에서 제공되는 V3 CD(Concept Descriptor) 유형 또는 ‘V3 generator’와 같은 도구를 사용하여 획득 할 수 있다. MAP 파일은 SCS와 MIF 간 맵핑 데이터를 포함하고, caAdapter의 맵핑 패널에서 생성할 수 있다. MAP 파일을 사용하여 caAdapter의 V3 메시지 생성기는 최종적으로 가공되지 않은 원재료(raw data)를 포함한 CSV 파일로부터 V3 메시지를 생성한다. 그러나 caAdapter는 CSV 파일을 생성 기능을 지원하지 않기 때문에 사용자가 V3 메시지를 생성하기 전에 준비해야 한다.

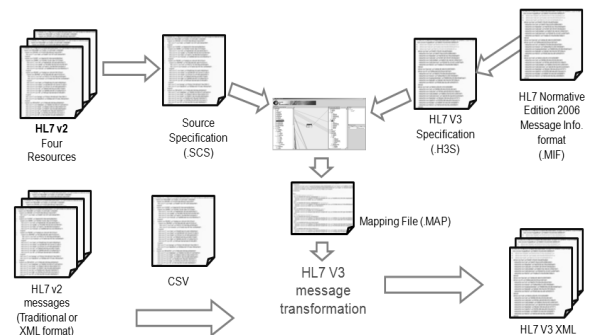


그림 1 caAdapter V3 생성 메커니즘
Fig. 1 V3 Generating Mechanism of caAdapter

2.2.2 V2 변환기(V2 Converter)

본 연구에서 개발된 그림 2의 V2 Converter는 caAdapter의 맵핑 패널과 V3 메시지 생성 도구로서 실행되는 주요 변환 엔진으로서 기능을 하고 있다. HL7 V2 메시지 parser

인 KNU V2 parser를 기반하고 있다. V2 Converter의 핵심 기능은 V2 메시지를 그림 2에서와 같이 CSV 및 SCS 파일로 변환시키는 것과 동시에 caAdapter의 모든 기능들이 원활히 수행될 수 있도록 지원 한다. 이 기능에서 HL7 V3에서 정의한 용어 간 맵핑 기능은 제외된다. 그림 2는 V2 Converter를 포함한 전체 프로세스를 나타내고 있다. 이 중에서 핵심은 'V2 Converter'로서 V2 메시지 파일이 V2 Converter에 의해 SCS 및 CSV 파일이 생성하는 것이다.

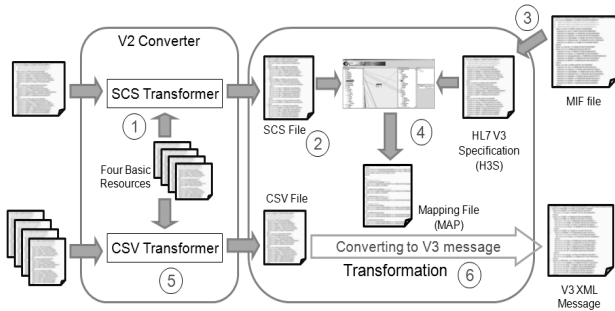


그림 2 V2 변환기와 CaAdapter 간 프로세스
Fig. 2 Process between V2 Converter and caAdapter

구체적으로 설명하면, 사용자의 GUI로부터 파라미터를 입력 받은 직후 KNU V2 Parser로 전송시게 되면, KNU V2 Parser는 파싱된 V2 Message Tree 객체를 반환시켜 SCS 및 CSV 파일을 생성한다.

2.2.1 SCS 및 CSV 파일 생성 메카니즘

KNU V2 Parser에 의해 파싱이 되면 SCS 파일이 생성된다. 그림 3은 이 과정에서 V2 메시지에서 추출된 SCS 구조를 보여주고 있다.

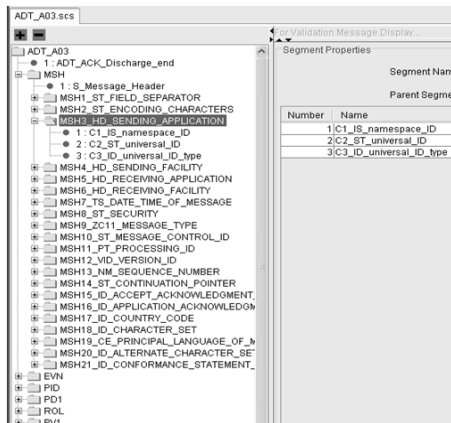


그림 3 caAdapter의 SCS 수정 패널
Fig. 3 SCS editing panel of caAdapter

caAdapter의 SCS 파일은 오직 2개의 레벨; 세그먼트와 필드로 구성되어 있으나 V2는 4개의 레벨; 세그먼트, 필드, 컴포넌트, 서브컴포넌트를 가지고 있다. 그리고 V3는 각 필드와 세그먼트가 반복할 수 있으나, SCS의 경우는 세그먼트

만 반복이 가능하다. 이러한 독립적인 규칙으로 인해 V2 메시지 트리 구조 그대로 SCS 트리로 변환될 수 없다. V2의 모든 세그먼트와 필드 레벨은 SCS 세그먼트로 표현되고 나머지는 SCS 필드로 인식된다.

2.2.2 용어 맵핑 기능의 VOM 파일

V2와 V3는 각 버전의 자체 코드와 독자적인 용어 체계를 가진다. 예를 들어 V2의 경우, 입원 진단 유형은 HL7 테이블 번호 0052에 의해 'A'로 정의되나, V3의 경우, V3 코드 시스템인 2.16.840.1.113883.11.16228 에 의해 'ADMDX'로 정의된다. 이렇게 V2에서 V3로의 메시지 변환 생성을 할 때 의미적으로 동일하게 메시지 변환을 위해 V2 메시지의 'A'에서 'ADMDX'로 정확히 변환이 이루어져야 한다. 그렇기 때문에 V2와 V3의 용어 시스템 간에 정확한 맵핑 기능을 하는 시스템을 필요로 한다. 그러므로 본 연구에서 용어 맵핑 기능을 가진 VOM(Vocabulary Mapping) 파일을 그림 4와 같이 생성하고 접근할 수 있는 기능을 추가하였다. 이는 KNU V2 Parser가 용어 맵핑 기능을 지원하지 않기 때문이다. VOM파일은 XML 유형으로서 "source value"와 "target value"로 1:1 맵핑을 할 수 있도록 생성하였다. 여기에서 "source value"는 V2, "target value"는 V3의 용어를 의미한다.

```
<?xml version="1.0" encoding="UTF-8"?>
<VocabularyMapping name="Test_Example01">
  <comment>
    Source : HL7 0007
    Target : 2.16.840.1.113883.11.16899
  </comment>
  <domain name="maritalStatus">
    <comment>
      This is Example domain about marital status.
    </comment>
    <translation name="single">
      <comment>this means single.</comment>
      <source value="S"/>
      <target value="SGL"/>
    </translation>
    <translation name="married">
      <source value="M"/>
      <target value="MRD"/>
    </translation>
    <.....>
    <elseCase type="errorMessage">
      <inverseElseCase type="taggingPrefix" value="Tag"/>
    </domain>
  <domain name="OtherCategory">
    <.....>
  </domain>
</VocabularyMapping>
```

그림 4 XML 유형의 VOM 파일 예제
Fig. 4 Example of VOM file with an XML format

3. 구현 결과

총 11개의 클래스와 8,000라인의 자바 코드를 프로그래밍 하였고, V2 Converter GUI에서 사용자는 유효한 V2 메시지에서 SCS 및 CSV를 생성할 수 있었다. 구현 테스트를 위해서 우리는 V2 메시지 중 ADT^A03를 총 5개 사용하였다. 이 메시지는 경북대학교병원에서 제공된 실제 사례이기 때문에, 환자의 이름, 주소 등의 개인적인 정보 메시지는 임의로 한글이 아닌 영어로 변경하였다. V2의 ADT^A03와 일치되는 V3 메시지는 PRRA_MT402003로서, HL7에서 제공하는 PRRA_MT402003.MIF 파일에 기반 한 PRPA_MT

402002.H3S 파일은 caAdapter H3S 패널을 활용하였다. 본 연구에서는 V2 Converter를 사용하여 V2 메시지서 ADT_A03.SCS 및 ADT_A03_01.CSV 파일을 생성하였다. 나머지 4개의 CSV 파일은 V3 메시지 생성과정에서 API 파일을 이용하여 자동적으로 생성되었다. 용어 매핑을 위해서 VOM 파일 생성도구가 아직 개발되어 있지 않기 때문에 이 파일은 XMLspy 소프트웨어[14]를 활용해서 XML 유형으로 수기로 입력하여 ADT_A03_To_02003.VOM 파일을 준비하였다. 다음은 가장 중요한 단계로서 ADT_A03_To_402002.MAP 파일을 생성하는 것이다. 이 파일은 ADT_A03.SCS 파일과 PRPA_MT402003.H3S를 이용하여 caAdapter 매핑 패널에서 생성하였다. 마지막 단계는 ADT_A03_01.CSV 파일을 이용하여 caAdapter V3 변환 패널에서 V3 메시지를 생성하므로 총 5개의 메시지를 생성할 수 있었다.

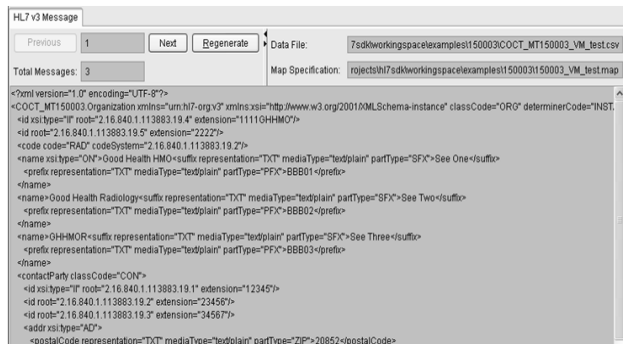


그림 5 생성된 V3 메시지
Fig. 5 Generated V3 message

4. 결론 및 토의

본 연구는 서로 다른 큰 대륙 간의 교류를 위해서 교량을 구축하는 것과 동일하다고 볼 수 있다. 아직까지 연구는 진행 중으로 불완전하고 많은 과제가 남아있다. V2와 V3 간의 개념적 차이는 이 두 버전 사이에 커다란 장벽을 만들었다. 이는 HL7 뿐만 아니라, 세계의 의료정보시스템의 급격한 기술적 진보에 따른 문제점이 낳은 결과라고 할 수 있다.

V3라는 진보된 기술이 발표되어 있음에도 불구하고 V2.4, 2.5 그리고 2.6과 같은 낮은 버전이 지속적으로 개발되고 유지되고 있는 점은 실제와 이상적 기술의 격차라고 할 수 있다.

본 연구에서 V2에서 V3가 명확히 변환되기 위해 V2 Converter에 의한 SCS, CSV파일의 생성과 caAdapter를 활용한 MAP 파일의 생성하였다. 이를 통해 caAdapter의 V3 메시지 생성기는 최종적으로 가공되지 않은 raw data를 포함한 CSV 파일로부터 V3 메시지를 생성할 수 있다. 두 버전 간 개념적 차이로 인해 완전한 자동화된 툴을 개발하는 것이 현재로서는 불가능할 일로 보이지만, 사용자가 메시지 항목마다 의료기관에서 수작업을 하는 것은 더욱이 기술적인 진보를 늦추는 결과를 초래할 것이다. 아직까지 caAdapter는 완전자동화 전 단계에 사용할 수 도구라고 할 수 있다.

본 연구에서 생성된 파일은 다른 사례에 충분히 재사용이 가능할 것이다. 이 중 VOM 파일은 재사용가능하며, 실제로 V2에서 V3로의 변환뿐 아니라 코드 시스템 간 시멘틱 변환을 위한 일반적 목적으로도 사용할 수 있을 것이다.

감사의 글

본 연구는 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원(No.2009-0066546)과 지식경제부 및 한국산업기술평가관리원의 지식서비스 산업원천기술개발사업(KI10033576)의 일환으로 수행하였음

참고 문헌

- [1] Health Level seven. HL7 Version 3 Development Framework (HDF), Ann Arbor, MI: Health Level Seven, Inc, 2005.
- [2] HS. Kim, CB. Park, HS. Hong, H. Cho, "Adoption of MFER HL7 Standard for Shared Electronic medical Record," Trans. KIEE, vol. 57D, no. 3, pp. 501-506, 2008.
- [3] DJ. Marotta, "HL7 in the 21st century. Integrating medical information exchange," Healthc Inform, vol.17, no. 4, pp. 46-48, 2000.
- [4] HH. Kim, H. Cho, T. Tung, HS. Hong, HS. Kim Development of HL7 Middleware for Patient Monitoring Device, Trans. KIEE, vol. 56, no. 9, pp. 1680-1687, 2007.
- [5] SM. Huff, WD. Bidgood, Jr., JJ. Cimino and WE. Hammond, "A proposal for incorporating health level seven (HL7) vocabulary in the UMLS Metathesaurus," Proc AMIA Symp, pp. 800-804, 1998.
- [6] caBIG, available at:<http://cabig.cancer.gov>
- [7] Java SDK 1.5, available at:<http://java.sun.com>
- [8] IntelliJ IDEA 7.0, available at:<http://www.jetbrains.com>
- [9] caAdapter, cancer Biomedical Informatics Grid 2007 Annual Meeting, Feb. 2007, Washington, DC.
- [10] S. Liu, et. al., HL7 SDK - Health Level Seven Software Development Toolkit, NIH Research Festival, Oct. 2005, Bethesda, Maryland.
- [11] Open source of caBIG project, available at:http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/caAdapter
- [12] T. Tran, HS. Kim, HH. Kim, H. Cho, "An Efficient Algorithm for HL7 Message Parsing," Trans. KIEE, vol. 55D, no. 6, pp. 274-278, 2006.
- [13] XMLspy, available at:<http://www.altova.com/XMLSpy>

저 자 소 개



엄 기 성 (嚴 基 成)

89년 연세대학교 보건학석사학위를 받고 04년 경북대학교 의료정보학박사학위 후 04년부터 미국 ISIS를 거쳐 현재 NIH의 CBIIT에서 caAdapter 프로젝트 연구원으로 활동 중이다.



김 화 선 (金 和 仙)

03년 인제대학교 컴퓨터공학석사학위를 받고 07년 경북대학교 의료정보학박사학위 후 동 대학에서 연구초빙교수로서 XML 기반의 병원정보시스템, 객체지향 기반 한 HL7 CDA, RIM, V3 개발, 표준 용어코드체계 등을 연구하고 있다.



홍 해 숙 (洪 海 淑)

77, 79년 경북대학교 간호학·석사학위 후 경북대학교 간호학과 교수로 현재 재직 중이며, 90년도 생물학박사학위 취득 후 간호정보학 중 HL7 V3 스키마 개발에 관심을 가지고 연구 중에 있다.



조 훈 (趙 勳)

현재 경북대학교 의료정보원천기술연구소 연구소장으로서 86년 미국 남캐롤라이나 주립대학 전산학석사, 92년 유타주립대학 의료정보학박사학위를 취득한 후 병원정보시스템, 영상 및 Tele-health에 관한 연구를 활발히 하고 있다.