

임상문서표준규격내 검사실 용어의 LOINC 매핑을 위한 LMOF 전처리 도구

도형호⁰¹ 김일곤¹ 이성기¹ 광연식² 송준현³ 이성현¹ 예정훈¹

경북대학교 전자전기컴퓨터학부¹, 경북대학교 의과대학 의료정보학 교실²,

지능형 진료지원 및 정보 공유 시스템 개발 센터³

{hhdo⁰¹, ikkim¹, sklee¹, yskwak²}@knu.ac.kr, {jhsong0135³, xenex¹}@dreamwiz.com, 4ngel¹@naver.com

The LMOF Preprocessing Tool for Mapping Laboratory Vocabulary to LOINC in Clinical Document Architecture

Hyoungho Do⁰¹, Il Kon Kim¹, Sungkee Lee¹, Yun Sik Kwak²
, Joon Hyun Song³, Sung Hyun Lee¹, Jung Hoon Yeah¹

School of Electronical Engineering & Computer Science, Kyungpook National University, Korea¹

Department of Medical Infomatics, School of Medicine, Kyungpook National University, Korea²

Intelligent Health Information Sharing System Development Center, Korea³

1. 서론

최근 컴퓨터의 성능 향상과 컴퓨터 네트워크의 발달에 따라 기관 내 업무의 전산화를 넘어서 많은 정보들을 외부 기관으로 교환 및 공유하고 있다. 이러한 현상은 의료 분야에서도 예외가 아니어서 병원 내부 전산화(전자 의무 기록: EMR)를 넘어서 국가 차원의 보건 의료 전산화(평생 전자 건강 기록 : EHR)가 국가 정책으로 자리 잡으며 각종 보건 의료정보를 공유하는 시대가 도래 하였다. 이로 인하여 서로의 정보를 주고 받기 위해서는 서로간의 전송 포맷 등의 통일을 위한 표준화가 필요하게 되었다[1].

평생전자건강진료정보(EHR, Electronic Health Record)에서는 병원 내 정보의 교환 및 공유를 위하여 HL7(Health Level Seven) 메시지 및 CDA(Clinical Document Architecture) 문서 제작을 하게 되는데, 여기에서 사용되는 용어들의 통일이 필요하며 이를 위해 LOINC(Logical Observation Identifiers Names and Codes)[2], SNOMED(Systematized Nomenclature of Medicine)와 같은 용어 체계를 사용하고 있다.

LOINC는 그 양이 매우 방대하므로 병원의 검사 코드들만을 LOINC로 매핑 시키는 작업만으로도 결코 쉬운 일이 아니다. 지금까지 로컬 검사 코드와 LOINC와의 매핑에 관한 연구는 로컬 검사 코드와 LOINC를 어떻게 자동으로 매핑 할 수 있을지가 주요 관심 대상이었다. Kambiz는 LOINC를 자동으로 매핑하려는 시도를 하였다[3]. 하지만 이러한 자동 매핑을 위한 노력은 해당병원의 임상 병리 전문가의 도움 없이는 병원 코드의 특성 등을 모르기 때문에 매핑의 한계가 있고, 사용자의 의도와는 맞지 않게 잘못 매핑 되는 문제점을 가지고 있다. 다른 관련 연구들도 모두 매핑된 정보가 정확하다는 검증이 되어 있지 않고, 자동 매핑의 한계점을 드러내고 있다.

2. 본론

Regenstrief Institute에서는 LOINC 매핑을 쉽게 할 수 있도록 도와주는 RELMA(Regenstrief LOINC Mapping Assistant)라는 도구를 제공한다. RELMA를 이용하여 기존 로컬 검사 코드를 LOINC로 매핑하는 작업을 수행하기 위해서는 RELMA에서 요구하는 LMOF(Local Master Observation File) 포맷에 맞게 로컬 데이터베이스가 제공되어야 하는 제약이 있다[4]. 하지만 이 포맷은 아주 이상적인 형태이고, 일반 병원의 데이터베이스는 하나의 필드에 모든 정보를 넣어두거나 한글 혼용 등으로 인해 이러한 요구사항을 충족시킬 수 없는 경우가 많다.

서론에서 소개한 바와 같이 병원 검사 코드와 LOINC를 자동으로 매핑 할 수 있는 가능성은 아주 낮다고 볼 수 있다. 본 논문에서는 LOINC 자동 매핑 대신 최선의 대안으로 임상 병리 전문가의 매핑 작업을 최대한 편리하게 할 수 있도록 병원 로컬 검사 코드 정보를 전처리 해주는 도구를 개발하여 매핑작업의 효율성과 작업속도의 향상에 공헌한다. 본 논문에서 제안하는 LOINC 매핑을 위한 전체적인 흐름은 그림 1과 같으며, 제작한 전처리 도구는 Local DB를 Revised LMOF로 바꾸어 준다.



그림 1. LOINC 매핑을 위한 전체적인 흐름

Preprocessing Tool은 하나의 필드에 혼용되어 있는 정보를 해당하는 필드로 옮겨주는 역할과 한글 키워드를 제외시켜주는 역할을 하는데, 개략적인 의사코드는 그림 2와 같이 나타낼 수 있다. 본 논문에서 사용하는 2차급 병원의 전처리 되지 않은 병원 검사 데이터베이스에는 ACR_NAME 이라는 필드에 LMOF의 TestName, MaximumValue, MinimumValue, Units 에 해당하는 정보가 모두 들어 있고, 또한 한글이 혼용되어 사용 되므로 검색의 효율성이 떨어진다. 본 논문에서는 모든 정보가 들어있는 ACR_NAME 필드에서 한글 정보가 포함된 모든 정보를 comments 필드로 옮겨 놓음으로서, 만약에 검사의 정보를 다시 알고 싶을 때 항상 볼 수 있도록 한다.

<pre> WRITE FILE(backup original file) WRITE(SQL): Update LMOF3 SET Comments = '%s' WHERE TestCode='%s',strTestDescription,strLineName READ(SQL): SELECT * FROM LMOF3 SHOW the all field REPEAT IF ASCII of TestDescription[index] is 40 WRITE index REPEAT IF ASCII of TestDescription[index] is 41 WRITE subindex END IF UNTIL index to LengthOfTestDescription END IF </pre>	<pre> UNTIL 0 to LengthOfTestDescription WRITE SubTestDescription Part of TestDescription(index to subindex) REPEAT IF ASCII SubTestDescription is 45 WRITE index UNTIL 0 to LengthOfSubTestDescription Parse the string according to this condition Write the Database corresponded fields. IF FlagHangul is TRUE THEN IF Comment <> ASCII_range THEN Exclude Hangul character from TestDescription END IF END IF SHOW the all field </pre>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

그림 2. 전처리 과정의 개략적 의사 코드

전처리 되지 않은 병원 데이터베이스는 표 1과 같고, 제안한 도구를 이용하여 전처리 시킨 LMOF의 형태는 표 2와 같다.

표 2. 제안하는 도구를 이용하여 전처리 된 결과

표 1. 전처리 전 병원 데이터베이스

TestDescription
소아 Bilirubin. T(0.1-1.2 mg%)
BUN(8-23mg%)
Glucose.FBS(70-110mg%)
Sodium(135-145 meq/1)

Test Description	Comments	Min value	Max Value	units
Bilirubin. T	소아 Bilirubin. T(0.1-1.2 mg%)	0.1	1.2	mg%
BUN	BUN(8-23mg%)	8	23	mg%
Glucose.FBS	Glucose.FBS(70-110mg%)	70	110	mg%
Sodium	Sodium(135-145 meq/1)	135	145	meq/1

본 논문에서는 국내 2차급 병원의 로컬 검사 데이터베이스를 중심으로 다음과 같은 실험을 하였다. 먼저 병원 데이터베이스에서 가져온 ACR_NAME 필드에 있는 데이터베이스를 바로 LMOF의 TestDescription 필드에 삽입하여 전처리 없이 RELMA를 사용을 하였을 때와 본 논문에서 제안한 도구를 이용하여 한 단계 전처리 과정을 거친 후, RELMA를 사용을 하여 검색을 하였을 경우를 서로 비교한다. 비교한 결과를 우선 RELMA 사용의 측면에서 봤을 때는 혼란스러움이 줄어드는 효과를 얻을 수 있다. 기존 전처리 없이 RELMA를 사용하면 필요외의 키워드들이 많아 사용자가 불필요한 키워드의 체크박스를 해제해 주어야 하는 불편함이 따른다. 또한, 전처리를 하지 않고 RELMA를 이용하여 검색하여 찾아지는 레코드 수와 제안한 방법으로 전처리 한 후에 RELMA로 검색했을 경우에 검색되는 레코드 수를 비교하였다. 표 3은 2가지 방법을 사용하였을 경우에 검색 되는 레코드의 개수를 서로 비교한 것이다. 아래와 같이 MinValue, MaxValue, Units 필드가 채워짐으로서 좀 더 세분화 된 검색이 가능하므로 표3과 같은 성능의 향상을 기대할 수 있다.

표 3. 성능 비교

검사명	전처리 하지 않는 방법	제안한 방법	비율
Protein. T	53	51	0.96
Protein. A	73	72	0.98
Protein. G	58	55	0.89
Bilirubin. T	58	55	0.89
Bilirubin. D	58	22	0.38
평균	60	51	0.85

3. 결론

본 논문에서는 RELMA를 이용하여 병원 검사코드들을 LOINC로 매핑하기 위하여 기존 병원 검사 데이터베이스를 전처리 하는 방법을 고찰하고 그 도구를 제작하였다. 또한 도구를 이용하여 국내 2차급 병원의 검사실 데이터베이스를 전 처리한 후 LOINC로 매핑하는 실험을 하여, 사용자의 편의 향상 및 성능의 향상을 보였다. 따라서 병원 검사실 용어의 표준화를 하기 위한 방법으로 자동으로 매핑하는 방법의 문제점을 줄이는 동시에, 로컬 데이터베이스를 아무런 처리 없이 RELMA를 사용하여 매핑하는 방법 보다 효율성을 올리는데 공헌을 하였다. 향후 과제로는 실험하였던 500여개의 검사실 데이터를 가지고 있는 2차급 병원보다 복잡하고 방대한 데이터를 가지고 있는 3차급 병원의 LOINC 매핑을 도울 수 있는 좀 더 안정성 있는 도구를 제작하는 것이 목표이다. 또한, 현재 LOINC에서 사용하고 있지 않는 키워드들을 RELMA에서 인식할 수 있는 언어로 바꿔줄 수 있는 지식기반의 방법을 사용하여 좀 더 향상된 성능을 기대 할 수 있을 것이다.

4. 참고문헌

- [1] 박연식, 평생전자건강진료정보: 규격정의, 종류 및 표준적용, 대한의료정보학회지, 제11권, 제1호, 25페이지, 2005
- [2] LOINC References <http://www.regenstrief.org/medinformatics/loinc/>
- [3] Kambiz et al, Design a Medical Terminology Interpreter as an Agent, Proceedings of the China-Japan-Korea Joint Symposium on Medical Informatics, 113-121, 2006.
- [4] RELMA manual <http://www.regenstrief.org/medinformatics/loinc/relma>