

혈액종합검사에 특화된 온톨로지 구축: 설계와 활용*

제현우, 박유경, 홍원의, 이문용

한국과학기술원 지식서비스공학과
대전광역시 유성구 대학로 291 305-701

jhw4824@kaist.ac.kr park60@kaist.ac.kr laftworld@kaist.ac.kr munyi@kaist.ac.kr

요약: 의료 종합 검진 분야의 전문 의료진들은 환자로 부터 측정된 각종 검사 수치를 기반으로 소견문을 작성한다. 이때 전문가는 자신의 경험에 근거하여 여러 검사 수치 및 환자의 신상 기록, 자동 생성된 단편적인 소견 등을 복합적으로 검토하게 된다. 그러나 현재 검사 항목에 대한 기본적인 소견을 자동으로 생성해주는 Computer-aided Diagnosis 도구는 검사 결과 수치에 따른 단편적인 문구를 생성하는 수준에 그치고 있어 복잡한 사례에는 적용되기 어렵다. 또한 지식 재사용을 위한 지식 베이스(knowledge base)가 체계적으로 개발되어 있지 않아 전문의가 생성해온 상당량의 지식이 버려지고 있는 실정이다. 이 이외에도 전문의들의 문장 구사 습관에 따른 용어 정의가 표준화되지 않아 지식 공유의 문제도 부각되고 있다. 이러한 문제들을 해결하기 위하여 본 논문에서는 경험지식플랫폼(Experiential Knowledge Platform)을 소개하고, 해당 플랫폼에서 지식을 생성하는 메타데이터(meta data)로 활용되는 온톨로지(Ontology) 구축 과정을 제시한다. 특히, 방대한 의료분야에 대한 선행 연구로서 혈액종합검사 분야에 대하여 구축작업을 완료하였다. 본 연구에서 제시하는 혈액종합검사분야의 온톨로지는 현장에 적용되어 그 실효성을 검증받은 후, 타 분야로 확장되어 해당 분야의 전문 문서 검색 및 질의어 확장 모듈, 료 생성 모듈을 생성하기 위한 지식베이스로 활용될 것을 기대한다.

핵심어: 온톨로지, 데이터베이스, 혈액종합검사, 경험지식플랫폼, 의학용어

1. 서론

Computer-aided Diagnosis (CAD)는 의료 검사 결과

의 정확도를 높이는 동시에 검사 절차를 간소화하고 전문 의료인의 작업 부하를 줄이는 방향으로 진화하면서 의료 검사에서 주요한 분야로 자리잡았다. 의료 검사 및 진단 전문가들은 자동 생성된 소견을 참고하여 최종 진단 소견을 작성한다. 이와 같은 상황에서, 정보기술은 의사결정과정에 필요한 절차를 대리하여 전반적인 서비스 품질을 향상시킨다 [1]. 현재 의료분야 데이터베이스는 유전자부터 척수, 질환 등 세부분야로 나뉘어 연구되고 있으며, 이를 통합하는 연구 역시 진행되고 있다[2].

진단의학 과정에서는 검체 샘플을 검사하는 자동화 시스템을 활용해 빠른 시간에 판독을 내리지만, 최종적으로 전문의가 그 결과를 판단하고 의사결정을 하는 과정에서 의사결정지원시스템의 부재로 인해 의료진의 업무 부담이 가중되는 문제점이 있다. 본 연구팀을 포함한 국내외 우수 연구소 및 기업은 컨소시엄을 구성하여 현장전문가의 경험지식 획득 및 활용을 위한 경험지식플랫폼 개발 연구(이하, 경험지식플랫폼)를 추진하였다. 본 연구팀은 경험지식 플랫폼에서 활용되는 온톨로지를 혈액종합검사 분야를 대상으로 개발하였다. 실제 혈액종합검사 분야에 사용되는 용어를 대상으로 검사항목, 검체, 질환과 같은 정보의 관계성을 정의하였으며, KOSTOM(Korea Standard Terminology of Medicine: 한국보건의료표준용어), UMLS(Unified Medical Language System), 그리고 KCD(Korean Standard Classification of Diseases)와 같은 기존 국내외 의료분야 데이터베이스 자료를 다수 수집해 상호 연계하였다. 산재된 용어들을 표준화된 형태로 표현하기 위해 동의어 중복 문제를 해소하였는데, 의미는 같지만 표현이 다른 개념들을 대표하는 대표용어를 유일하게 정의하고, 현장 전문의들의 온톨로지 검토과정을 모사한 구축

* 본 연구는 산업통상자원부 및 한국산업기술평가관리원의 산업핵심기술개발사업(지식서비스)의 일환으로 수행하였음. [10052955, 현장 전문의의 경험지식 획득 및 활용을 위한 경험지식플랫폼 개발 연구]

절차를 통해 결과의 신뢰성을 갖출 수 있었다.

현장전문가의 의사결정을 돕기 위한 경험지식플랫폼 개발연구에서는 혈액종합검사 영역의 논리구조 및 관계와 의학용어를 수록하는 온톨로지 구축이 필수적이다. 해당 온톨로지는 서로 다르게 해석될 수 있는 각종 의료 용어들을 표준화된 형태로 정리하는 역할뿐 아니라 혈액종합검사의 관계성을 내포하고 있기 때문에 그림 1 과 같이 경험지식플랫폼 내부의 질의어 확장 모듈과 를 생성을 보조하는 메타지식으로 활용된다. 질의어 확장 모듈에서는 최종사용자가 사용자 인터페이스에서 만들어내는 질의어를 받아서 질의어와 관련된 내용을 의학 교과서 및 저널에서 자동으로 수집하는 것이 목적이다. 이러한 작업을 위해서는 적절한 질의어를 선정하는 것이 중요하다. 그렇기에 사용자가 부분적인 질의어를 입력하더라도, 해당 질의어가 어떤 의미를 갖고 있는지에 대한 정보와 관련 동의어를 수록한 온톨로지의 지원을 받아서, 더욱 정확한 관련 문서를 찾을 수 있다. 그리고 경험지식플랫폼 시스템 상에서 사례를 판단하는 것과 같이, 노하우 형태로 습득되는 절차적 지식이라는 일종의 룰을 생성하기 위해서 참조가 필요한 메타지식인 온톨로지가 필요하다.

본 연구팀은 혈액종합검사 온톨로지 구축 연구를 위해 국내외에서 연구된 사례가 있는 의료분야 데이터베이스를 수집하였다. 국외 사례로는 UMLS¹, MeSH² (Medical Subject Headings), SNOMED³ (Systematized Nomenclature of Medicine)가 대표적이며, 국내에서는 KOSTOM, KCD가 대표적인 의학 용어 데이터베이스이다. 하지만 이러한 국내외

데이터베이스 자료를 조합하여 경험지식플랫폼을 위한 메타지식으로 즉시 활용하기에는 몇 가지 문제점이 있다.

하나, 혈액종합검사라는 영역에 특화된 온톨로지가 부재하다. 다른 분야의 온톨로지와 상이하게 혈액종합검사 분야의 경우, 해당 분야의 특성으로 인해 검사항목, 검체, 질환, 검사항목 측정치 수준과 같은 개념 사이의 관계 정의가 필요하다.

둘, 표 1 의 예시와 같이 KOSTOM 온톨로지에는 불필요한 데이터 중복 문제가 존재한다. Abdominal pain 이라는 영문명을 복부통증, 배통증, 배앓이, 복통 등으로 다양한 한글명으로 표현할 수 있는데, 이러한 동의어 관계를 KOSTOM에서는 한/영문 동의어를 한 쌍으로 묶어 ID(용어코드)를 부여하는 방식을 채택했기 때문에 불필요한 데이터 중복문제가 발생하게 된다.

표 1. 혈액종합검사 동의어 예시

용어코드	개념코드	영문명	한글명	UMLS
H00003487	H00003511	Abdominal pain	복부통증	C0000737
H00003492	H00003511	Abdominal pain	배통증	C0000737
H00003506	H00003511	Abdominal pain	배앓이	C0000737
H00003511	H00003511	Abdominal pain	복통	C0000737

셋, 혈액종합검사 온톨로지를 구축 시, 실제 의료 현장에서 해당 구조를 활용 가능하도록 설계하는 것이 중요하다. 그렇기에 실제 의료기관에서 사용중인 시스템과의 연동을 통해 기존의 시스템이 보유하고 있는 경험지식을 포괄할 수 있는 온톨로지 구축의 필요성이 있었다. 이러한 이유로 실제 현장에서 활용되는 시스템에 내포되어있는 현장전문가의 경험지식을 획득할 수 있는 유연한 시스템을 구축하기 위해, 국내 우수 법인인 씨젠의료재단의 전산시스템과의 연동을 추진하였다.

위와 같은 이유로 현장에서 의료진이 사용하는 특수용어를 반영하여 확장할 수 있는 온톨로지 체계가 요구되었고, 본 연구팀은 혈액종합검사 온톨로지를 만들기 위한 기초 자료로 KOSTOM 데이터를 선정하였다. KOSTOM은 2004년부터 보건복지부 산하 보건 의료정보표준화위원회가 장기적으로 구축해온 표준용어체계로 진단, 의료행위, 임상검사, 방사선의학, 치과, 보건, 간호, 기타의 8개 분야 서브셋(subset) 테이블을 포함하는 통합용어테이블이다[3]. KOSTOM을 혈액종합검사 온톨로지 구축의 기초자료로 선택한 것은 정제된 의학용어가 수록됨으로써 데이터의 신뢰성을 보장할 뿐 아니라, 외부 온톨로지 참조를 위한 각종 코드분류체계도 수록되어 있어 활용이 용이했기 때문이다. 또한 KOSTOM에서 검사항목과 관련

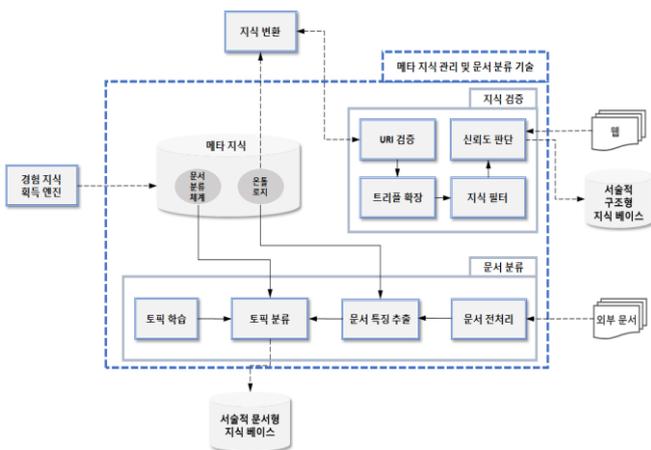


그림 1. 메타 지식 관리 및 문서 분류 기술 세부 동작도

¹ <https://www.nlm.nih.gov/research/umls/>
² <https://www.nlm.nih.gov/mesh/>
³ https://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html

된 용어를 보완하기 위해서 대한진단검사의학회⁴에 있는 의학용어와 각종 유의어를 수집해 반영하였다.

본 연구팀은 이러한 리소스 선정과정을 통해서 국내외의 온톨로지를 상호 연결할 수 있는 구조로, KOSTOM 데이터를 토대로 혈액종합검사를 위한 온톨로지 구축에 대한 연구를 진행하였다. 본 연구팀이 제안하는 혈액종합검사 온톨로지는 현장에서 구축된 데이터베이스를 활용함과 동시에 실제 종합검진 시스템과의 연동성을 고려하여 설계함으로써 실용성을 강화하였다. 실제 의료현장에서 발생하는 소견문에 등장하는 용어 역시 현장전문가의 경험지식으로써 반영하였다. 또한, 한/영 동의어 세트를 독립적으로 구축함으로써 문서검색에 활용되는 질의어 확장 모듈의 지식베이스로 활용할 수 있을 뿐 아니라 불필요한 질의어 생성 가능성을 사전에 방지함으로써 검색 시스템의 효율성을 증대시켰다. 추가로 의료기관의 협조를 통해 온톨로지 관계성 및 용어목록의 신뢰성을 갖출 수 있었다. 마지막으로 대표용어를 선정하는 구체적인 과정을 전문가의 업무 수행 절차를 모사하여 수행하였다.

2. 관련 연구

혈액종합검사 온톨로지를 구축하기 위해서는 먼저 의료분야에서 이뤄진 온톨로지, 데이터베이스, 시소러스(Thesaurus), 표준용어체계의 선행연구 검토와 그 차이점을 파악하는 작업이 선행될 필요가 있다.

온톨로지와 데이터베이스는 설계 목적에서 차이가 있다. 데이터베이스는 데이터들이 통합되고 저장되어있는 데이터의 집합인 것인 반면, 온톨로지는 사물을 개념과 개념간의 관계 정의로 표현해 놓은 것이다. 또한, 데이터베이스의 목표는 효율적인 데이터 관리 및 저장인 반면, 온톨로지는 실세계의 사물을 제약사항 및 개념간의 관계를 상세하고 체계적으로 표현하여 정의한 것이다.

시소러스는 특정한 도메인에서 사용되는 용어와 그들 사이의 관계를 제시한 색인어휘집이다. 시소러스는 주로 색인과 검색 과정에서 디스크립터와 검색어를 선정하기 위한 도구로 사용된다[4]. 그렇기에 정보검색 보조도구로서의 시소러스 개발을 위한 많은 연구와 이를 실현하기 위한 연구들이 진행되고 있다. 온톨로지와 시소러스의 공통점은 개념 간의 의미 관계를 기반으로 지식의 구조를 규정한 방법론이라는 것이다. 온톨로지와 시소러스는 모두 지식을 표현하는 방법이다. 지식 형성은 개념이 구조화된 것으로 생각할 수 있다. 즉, 각기 다른 개념이 정의되고 그

개념간 관계를 정의함으로써 지식은 규정된다. 그렇기에 지식을 형성하기 위한 방법론으로써는 온톨로지뿐만 아니라 시소러스 역시 가능하다.

온톨로지와 시소러스의 차이점으로는 온톨로지는 시소러스와 비교하여 개념 관계를 세분화할 수 있는 표현력을 가진다는 것이다[4]. 다시 말해, 각 개념의 정의와 용어간 관계를 다양한 정의 속성들을 활용함으로써 구조화 시킬 수 있다. 그 결과 시소러스와 비교하여 온톨로지는 인간이 지식을 이해하는 형상과 더욱 비슷한 구조를 나타낸다. 그렇기에 인간은 개념과 개념 사이의 관계성을 기반으로 지식을 파악하는 심적구조를 갖고 있기 때문에, 온톨로지를 활용한다면 개념 간의 관계를 구조화함으로써 부분적인 지식구조 영역의 특성을 더욱 분명하고 세밀하게 표현 가능하다.

온톨로지 언어가 시소러스와 구별되는 또 다른 특성은 일반화 혹은 상호운영성의 규칙을 적용함으로써 구조화된 지식으로부터 새로운 지식을 추론할 수 있다는 점이다[4]. 이는 온톨로지 설계자가 정의한 각기 다른 개념 및 속성 그리고 관계정의를 토대로 추론을 통해 추가로 나타나는 지식을 의미한다. 추가적으로 생성되는 지식 또는 룰은 메타지식을 통한 인공지능과 같은 역할을 시스템 속에서 담당한다.

국민 건강 증진과 의료서비스 분야의 가치 창출을 목적으로 세계 각국에서는 다양한 의료분야 온톨로지를 구축하고 있다. 그 중 대표적인 온톨로지는 UMLS, MeSH, SNOMED 이며, 국내에서는 KOSTOM, KCD 등이 있다. 의료분야 내 국소적인 온톨로지 구축에 대한 노력도 계속 진행되고 있는데, 그 중 대표적인 온톨로지는 Gene Ontology 로써 유전자 정보에 대한 개념과 의미 관계망을 OWL 언어로 표현한 것이다[5]. 그뿐만 아니라 인체의 전반적인 인체자원은행을 온톨로지로 설계하기도 하며, 세부영역인 척추정보만을 도메인으로 한 온톨로지를 연구하기도 하였다[6,7]. 또한 한국 및 중국 문화권에서는 한의학을 대상으로 한 온톨로지 구축에 대한 연구도 지속적으로 이뤄지고 있다[8]. 이러한 국소적인 분야를 다루는 온톨로지는 하나의 모듈로써 단일한 복합 온톨로지 통합하려는 연구가 이뤄지고 있다[2].

국외에서의 대표적인 의료분야 온톨로지는 UMLS, MeSH, SNOMED 등이 있다. UMLS는 미국 국립의학도서관(National Library of Medicine, 이하 NLM)에서 1986년부터 시작한 장기 연구개발 프로젝트로써 다양한 생명과학정보를 통합하고 검색하기 위해 개발된 온톨로지이다. UMLS를 이용함으로써 기존의 동일 개념을 표현하는 데 사용되는 용어의 다양성으로 인한 정보검색의 효율저하 및 통합검색 문제를 해소할 수 있었다[9]. UMLS는 개념 및 용어 데이터베이스인 메타시소러스, 개념들을 분류하기 위한 의미망, 자연어처리 작업을 위한 전문가사전으로 구성된다[9]. 메타시소러스는 생물의학적 개념과 동의어뿐만 아니

⁴ <http://www.labtestsonline.kr/>

라 개념간 관계성에 대한 정보를 포함한다. 의미망은 모든 개념들을 일괄적으로 목록화한 시멘틱 관계(semantic relationship)으로 구분된다. 전문가사전은 자연어처리를 위한 사전정보를 제공하기 위한 도구이다.

MeSH는 미국 국립의학도서관에서 구축한 의학분야의 시소러스로써 의학용어를 유의어, 상위어, 하위어, 관련어 등으로 정의했다. MeSH는 1960년 “Index Medicus(New Series)” 발간과 함께 4,400개의 용어를 수록한 “Medical Subject Headings: Main headings, Subheadings, and Cross references used in the Index Medicus and the National Library of Medicine Catalog”란 서명으로 초판이 발행되었다[10]. 컴퓨터의 발전 및 보편화로 인해, MeSH는 의료현장에서 데이터베이스 색인과 검색에 사용되고 있다. MeSH 표목이 27,000여개로 늘어났으며, 어휘의 상호참조, 기입어를 포함하면 123,000여개에 이른다[10]. 현재 NLM에서는 MeSH를 의료분야의 여러 하위 데이터베이스와 NLM에서 소장하는 목록의 주제색인을 위해 사용하고 있다.

SNOMED는 전자의무기록을 위한 용어시스템으로써 CAP(College of American Pathologists)에 의해서 발전, 유지, 배포되고 있다[11]. 또한 질병, 임상소견, 처치를 표현하는 포괄적 임상용어로 질병을 분류하기 위해 개발한 것이다. SNOMED는 다양한 전문분야와 진료 장소에서 임상자료를 색인, 저장, 검색 및 통합하는 것을 허용한다. 또한 SNOMED의 가장 큰 특징은 특정 질병을 질병이름뿐만 아니라 환자의 상태를 묘사하는 방식으로 코딩이 가능하다는 점이다[11].

국내에서 대표적인 의료분야 용어 및 분류체계 연구는 KOSTOM과 KCD가 있다. KOSTOM은 2004년부터 보건복지부 산하 보건의료정보표준화위원회가 한국보건의료표준용어체계를 구축한 것이다[3]. KOSTOM 구축은 용어표준화를 위한, 보건의료 및 한의학에 사용되는 모든 용어를 개념단위로 군집화하고 각 군집마다 유일한 코드를 부여함으로써 개념과 동의어 처리를 마칠 수 있었다. 이러한 활동은 향후 보건의료기관 HER(Electronic Health Record: HER)시스템의 상호운용성을 보장하기 위한 기초이다[3]. 표준화위원회에서는 보건의료정보 표준화연구를 통하여 KOSTOM에 포함되어있는 KCD 한국어용어를 UMLS에 등재를 함으로써 국외 자료와 연결을 도모했다. 각 분야별 보건의료용어에서는 해당 분야에 맞는 외부의 기본용어체계를 참조함으로써 설계의 타당성을 높였다. 또한 WHO 용어작업에 합류하는 등 국제기구와의 협력을 도모하여 UMLS와 LOINC(Logical Observation Identifiers Names and Codes)에 등재하였다.

국가간 통일된 의학용어체계의 필요성으로

ICD(international classification of diseases)를 사용하고 있다. 우리나라도 ICD-10을 번역하여 국내 실정에 맞도록 질병을 추가하고 상세 분류한 KCD-6(Korean classification of diseases)을 사용하고 있다[12]. 이러한 국제표준 분류체계를 사용함으로써 생기는 장점은 다음과 같다. 하나, 의료기관뿐만 아니라 지역, 국가간 상호 운용성을 확보할 수 있고, 국가간 통계 산출 및 다양한 분야간의 데이터 비교가 가능하다[12]. 이처럼 국가간 통일된 의학용어체계 구축은 의무기록의 전산화 통일뿐만 아니라 국가간 의료서비스의 양과 질적 측면 비교에도 활용될 수 있다.

이처럼 UMLS, MeSH, SNOMED, KOSTOM, 그리고 KCD와 같이, 국내외에서 의학분야 온톨로지 구축에 대한 연구가 지속되고 있다. 하지만 이러한 선행연구 결과를 혈액종합검사 온톨로지를 구축하는데 기초 리소스로 바로 적용하기에는 여러 문제점이 존재했다. 먼저, UMLS는 의학분야에 대한 용어를 정리한 데이터베이스인 메타시소러스뿐만 아니라 개념간의 관계를 정의 및 분류하여 구조적으로 정의를 내린 의미망을 갖고 있지만, 혈액종합검사를 위한 개념간의 의미관계 설정이 되어있지 않았다. UMLS는 단순 개념들간의 상하 구조적 관계를 포함하고 비슷한 개념을 갖고 있는 용어들을 묶는 작업을 중점으로 하였다. 또한, MeSH 역시 의학용어를 유의어, 상위어, 하위어, 관련어로 정의하는 작업을 하였을 뿐, 혈액종합검사를 위한 개념간의 관계설정에 대한 연구는 없었다. 이와 마찬가지로 SNOMED와 KOSTOM, 그리고 KCD는 의학용어를 표준화하여 의학 용어를 개념화시켜 정리한 작업이다. 그렇기에 이러한 선행연구들은 의학용어 표준화 작업 및 개념화 작업에서 큰 기여를 하였지만, 혈액종합검사 온톨로지 구축을 위한 개념간 관계 설정이 되어있지 않은 한계가 있었다. 본 연구팀은 이러한 선행연구의 리소스를 활용해서 혈액종합검사라는 특수한 영역의 온톨로지 구축을 시도했다.

본 연구팀은 혈액종합검사 온톨로지를 구축하기 위해서 단순히 의학용어를 개념화시키는 작업뿐만 아니라, 혈액종합검사를 위해 필요한 특수 개념들의 유기적인 관계정의 작업을 시도했다. 그렇기에 본 연구팀은 혈액종합검사를 위해서 검사항목과 검체, 그리고 질환과의 유기적인 의미 관계망을 정의했다. 또한, 특정 검사항목의 수치 수준에 따라 상이한 질환을 의심할 수 있기 때문에 검사수치 수준에 대한 정보의 연결을 시도했다. 즉, 혈액종합검사 온톨로지 구축을 위해서는 검사항목, 검체, 질환, 검사항목 측정치 수준과 같은 4가지 상위 개념간의 의미관계 설정이 필요했다. 이러한 이유로 인해, 본 연구팀은 앞서 언급한 문제점을 해결할 수 있는 혈액종합검사에 특화된 온톨로지 구축에 대한 연구를 진행하였다.

3. 혈액종합검사 온톨로지 설계

본 연구팀이 제안하는 혈액종합검사 온톨로지는 관계형데이터베이스 기반으로 설계하였다. 의학용어 표준화 작업을 위해 만들어진 KOSTOM은 데이터가 테이블 형식으로 구조화되어 있었다. 그렇기에 본 연구팀은 관계형데이터베이스를 활용하여 검사항목, 검체, 질환, 측정치 수준과 같은 개념간의 관계를 정의함으로써 KOSTOM 데이터를 확장 및 재구조화 시키는 것이 혈액종합검사 온톨로지를 표현하기에 가장 적합하다고 판단했다. 왜냐하면 기존 KOSTOM에 수록된 의학용어와 구조를 반영할 수 있을 뿐만 아니라, KOSTOM에서 발생하는 불필요한 데이터 중복을 최소화하는 스키마를 설계함으로써 정규화 문제를 해소할 수 있다는 장점이 있기 때문이다.

본 연구팀은 혈액종합검사 온톨로지 구축에 필요한 의학용어를 수록하기 위해 각종 리소스 목록을 조사하여 표 2와 같은 최종 리소스를 선정하였다. 이를 토대로 혈액종합검사 영역에서 발생하는 개념간의 관계를 정의하는 혈액종합검사 스키마를 설계했다. 추가로 데이터 정제 및 대표용어 선정 프로세스를 고안하였으며, KOSTOM 데이터가 반영하지 못하는 검사항목 용어와 같은 추가적인 정보를 외부 리소스를 참조하여 온톨로지에 수록하였다.

표 2. 혈액종합검사 온톨로지 리소스

자료명	형태	참조 데이터	목적
2015 쉐젠의 료재단 검사 안내	책	검사항목(Component)	용어, 의미관계 추출
쉐젠 Handbook of Laboratory tests	책	관련질환(Disease)	용어, 의미관계 추출
쉐젠 혈액종합검사 중복 (전산)	엑셀	검사항목(Component) 검체(System)	용어 추출
쉐젠 검사동의어 (전산)	엑셀	검사항목(Component)	동의어 추출
KOSTOM	엑셀	검사항목(Component) 검체(System) 관련질환(Disease)	동의어 추출 KOSTOM 코드 UMLS코드 기타코드(KCD)
Lab Tests Online	웹	검사항목(Component)	동의어 추출
쉐젠 전산 시스템 소견문	엑셀	소견문 내 특이용어	용어 추출

본 연구팀이 혈액종합검사 온톨로지 구축을 위해 선정한 리소스는 표 2와 같다. 경험지식플랫폼의 모듈로 활용될 혈액종합검사 온톨로지를 구축하기에 기초가 되는 리소스로 가장 적합한 것은 KOSTOM이라고 생각하였다. 왜냐하면 의학용어 표준화를 위해 국내외 보건의료에 사용되는 모든 의학용어를 개념단위로 정리하였고, 국제용어체계와의 연결을 위한 작업이 선행된 리소스이기 때문이다.

하지만 KOSTOM에 존재하는 데이터를 혈액종합검사 온톨로지에 포함되는 데이터로 바로 활용할 수는 없었다. 왜냐하면 하나, KOSTOM에는 혈액종합검사뿐만 아니라 의료 전 분야에 대한 용어표준화 작업이 진행된 데이터가 포함되어있기 때문이다. 그렇기에 본 연구팀은 혈액종합검사에 관련된 의학용어를 추출할 필요가 있었다. 둘, KOSTOM 구조를 혈액종합검사 온톨로지에 적용시키는 것은 데이터의 중복문제를 심화시키는 단점이 있었다. KOSTOM 데이터는 표 1에서 언급한 것과 같은 구조로써 발생하는 동의어 중복문제가 있었다. 이러한 구조를 그대로 활용하여 검사항목, 검체, 질환, 측정치 수준을 나타내는 속성 정보를 포함하도록 온톨로지를 설계한다면 데이터 중복문제가 심각해진다는 단점이 있었다. 그렇기에 대표용어를 선정함으로써 KOSTOM에 있는 동의어 중복문제를 해소하는, 데이터베이스 정규화 작업이 필요했다. 셋, KOSTOM에는 혈액종합검사를 위한 특수한 영역에서 발생하는 검사항목에 대한 용어를 모두 다루고 있지 않았다. 그렇기에 본 연구팀은 대한진단검사의학회의 자료를 참조하여 KOSTOM에서 수록되어 있지 않은 혈액종합검사 검사항목에 대한 용어 및 동의어를 온톨로지에 반영했다. 대한진단검사의학회에서는 보건 의료에서 중요한 역할을 하고 있는 진단검사의학 분야의 각종 검사에 대해 이해하기 쉽게 기초적 정보를 환자 및 그 보호자들에게 제공하고 있을 뿐만 아니라 직접 모든 자료를 제작 및 감수를 시행함으로써 데이터의 신뢰성을 높였다. 또한, 미국 AACCC(American Association for Clinical Chemistry)와의 협력에 의해 구축된 범세계적 사이트이기에 온톨로지 데이터 수록에 대한 신뢰성을 갖춘 리소스라고 판단했다. 넷, KOSTOM은 혈액종합검사 온톨로지 구축을 위한 개념간의 관계정의 정보가 부재했다. 그렇기에 본 연구팀은 표 2의 리소스를 참조하여 혈액종합검사를 위한 검사항목, 검체, 질환, 그리고 검사항목의 측정치 수준과 같은 개념들간의 관계를 정의했다. 그 후에, 혈액종합검사에 필요한 개념들간의 관계를 정의한 구조를 나타내는 스키마 설계를 시도했다.

현장전문가의 경험지식을 획득하는 것에 중점을 두고 있는 경험지식플랫폼 연구 특성을 고려하여 실제 사용중인 전산시스템 소견문에서 의료진이 사용한 용어를 추가로 온톨로지에 수록하였다. 이러한

작업을 통해서, 온톨로지는 국내외 표준에 맞는 의학용어체계뿐만 아니라 의료현장에서 실제로 사용되고 있는 용어를 상호 연결함으로써, 시스템 자체의 유연성을 높였다.

현장에서 요구하는 혈액종합검사에 대한 세부적 검사항목 종류와 그에 따른 검사수치는 병, 의원마다 차이가 있을 수 있다. 이러한 이유로, 본 연구팀은 씨젠의료재단에서 사용하고 있는 검사항목 및 검사수치에 따른 질환의 의심정도를 판단하는 기준에 맞추어 온톨로지를 구축하였고, 추후 사용자의 요구에 따라 변경(customization) 및 확장 가능하도록 하였다.

본 연구팀이 제안하는 혈액종합검사 온톨로지의 스키마는 그림 2 와 같다. 혈액종합검사 온톨로지 에서 중심이 되는 테이블은 KEYWORD 테이블이다. 해당 테이블은 대표용어를 수록하기 위해 고안되었다. 해당 테이블을 구성하는 속성은 3 가지이다. EKP_CODE 는 대표용어와 매핑되는 주 키(primary key)를 뜻한다. 해당 코드를 활용함으로써 대표용어와 관련된 동의어 집합을 탐색할 수 있다. 또한 TYPE 속성으로 해당 대표용어가 어떤 카테고리에 묶이는지에 대한

정보를 포함한다. 현재 혈액종합검사에서 대표용어는 검사항목, 검체, 질환, 카테고리, 기타와 같은 5 가지 타입(type)으로 구분된다. 대표용어를 지정하는 테이블과 동의어 테이블을 구분함으로써, 기존 KOSTOM 데이터베이스에서 발생하는 데이터 중복을 최소화하는 스키마 설계를 고안하여 정규화 문제를 해결할 수 있었다.

동의어 사전 역할을 할 수 있는 테이블은 KR_SYN, EN_SYN 과 같이 두 테이블로 분할하였다. 왜냐하면, 온톨로지를 활용해서 문서를 검색할 경우, 한/영 동의어를 구분 없이 모아두면 검색에 불필요한 질의어를 만들어 낼 가능성이 존재하기 때문이다. 이러한 이유로 한/영 동의어 테이블을 따로 설계함으로써 질의어를 확장하는 작업에서의 검색 정확도 및 검색 효율을 온톨로지 구조 자체에서 강화시켰다.

KEYWORD_CODE_MATCHING 테이블에는 외부 온톨로지와의 연계를 위해 각종 코드정보를 수록하였다. 그렇기에 특정 대표용어와 관련된 UMLS, KOSTOM, KCD 코드를 참조하여 외부 온톨로지와 연동 및 탐색이 가능할 수 있도록 하였다. 추가로 UMLS, 및 KOSTOM 정보가 갱신될 때, 해당 코드를 참조하여 혈액종합

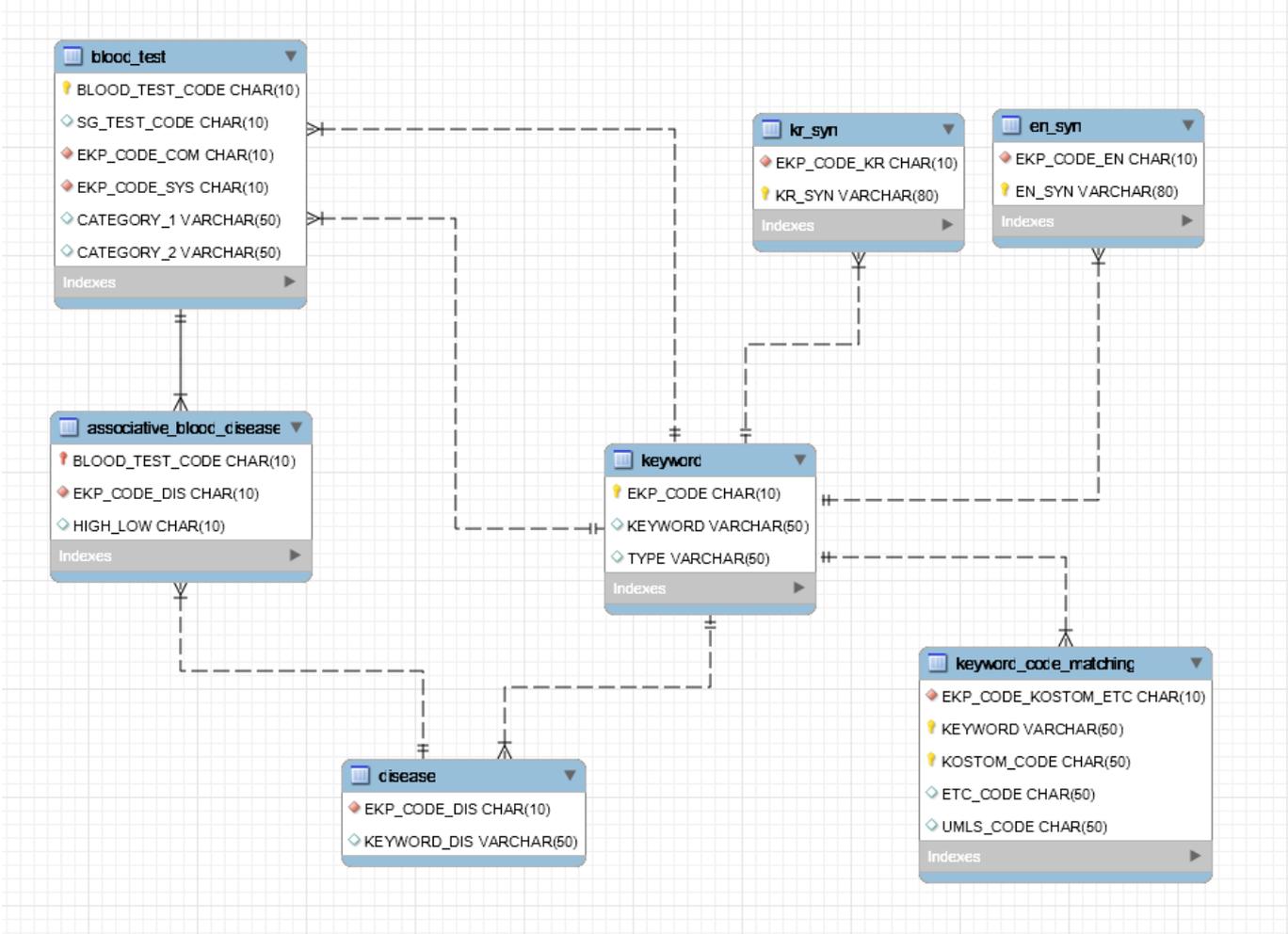


그림 2. 혈액종합검사 온톨로지 스키마

검사 온톨로지의 수정사항 반영을 용이하도록 설계하였다.

혈액종합검사의 전체적인 논리구조 및 의미관계를 포함하는 테이블은 BLOODTEST 이다. 혈액종합검사 구체적 하위 항목은 병, 의원마다 다소간 차이가 있다. 그렇기에 본 혈액종합검사 온톨로지는 씨젠의료재단의 혈액종합검사 기준을 따랐다. 이런 이유로, 혈액종합검사 항목과 관련된 질환 정보와 측정치 수준 정보를 담고 있는 ASSOCIATIVE_BLOOD_DISEASE 테이블과 BLOODTEST 테이블을 분리하여 병, 의원마다 발생하는 검사항목의 차이를 BLOODTEST 테이블에 유연하게 반영시킬 수 있도록 스키마를 설계하였다. 또한, 기존에 운영되고 있는 전산시스템의 의료정보 체계를 반영하기 위해서 SG_TEST_CODE 를 부여하여 기존 시스템과의 연동을 도모했다. 검사항목과 검체의 코드정보를 담고 있는 EKP_CODE_COM 와 EKP_CODE_SYS 의 관계성을 BLOOD_TEST_CODE 라는 주 키를 활용하여 수록했다. 또한, 해당 항목이 실제 진단을 목적으로 하는 병, 의원 현장에서 쓰이는 분류 용어와 의학 교과서에서 쓰이는 분류 용어와의 불일치로 속성값 CATEGORY_1, CATEGORY_2 를 구분하여 해당 정보를 따로 수록하였다. 실제 현장에서 사용되고 있는 분류체계와 학술정보를 탐색할 수 있는 정보를 모두 포함함으로써, 해당 온톨로지를 활용하여 문서검색의 효율성을 높일 수 있었다.

ASSOCIATIVE_BLOOD_DISEASE 테이블에는 검사항목과 검체의 관계 정보를 담고 있는 BLOOD_TEST_CODE 와 질환을 나타내는 코드 정보인 EKP_CODE_DIS 그리고 검사항목 측정치의 수준 정보를 나타내는 HIGH_LOW 속성값이 있다. EKP_CODE_DIS 와 같은 코드정보를 활용한 이유는 동의어 중복문제를 해결하기 위함이다. 다양하게 표현될 수 있는 동의어들을 모두 수록하여 불필요한 데이터 중복을 막기 위해서, 각 동의어 세트를 대표하는 대표용어의 코드를 활용하였다. 또한, 구체적인 측정치는 병, 의원마다 차이가 있는 것을 고려하여 메타지식인 온톨로지에서는 단순히 특정 검사 수치 수준이 높은지 또는 낮은지에 대한 정보만 HIGH_LOW 속성을 통해서 수록하였다.

끝으로 DISEASE 테이블은 질환 사이의 계층정보의 반영을 고려해서, KEYWORD 테이블과 독립성을 유지하였다.

이로써 혈액종합검사 온톨로지는 경험지식플랫폼 내부에서 활용되는 메타지식의 역할을 할 뿐만 아니라, 혈액종합검사라는 의료서비스 세부분야 온톨로지를 설계 및 구조화하는 방법을 제시하였다.

본 연구팀이 활용한 의료용어 데이터는 KOSTOM 을 사용하였다. 그 이유는 KOSTOM 은 보건의료정보표준화위원회가 한국보건의료표준용어체계를 구축한 것으로서, 의학용어를 개념을 토대로 구성하고 UMLS, KCD 와 같은 국내외 대표적인 온톨로지와의 호환성을 보장하는 코드 정보를 수록하고 있기 때문이다.

KOSTOM 데이터를 기반으로 앞서 본 연구팀이 설계한 스키마에 맞도록 1차적으로 용어를 분류하고, 삭제 및 수정 또는 추가적인 검토가 필요한 용어들은 2차 전문의 검토를 위하여 남겨두었다. 2차 전문의 검토를 마친 데이터는 본 연구팀이 재확인 후, 해당 결과를 온톨로지에 반영하였다. 지식공학자들의 충분한 의학 도메인 지식은 부정확한 데이터베이스 설계 및 구축을 야기할 수 있기 때문에 정확한 용어 분류 및 정제가 필요했다. 그렇기에, 본 연구팀은 실제 현장에서 근무하는 전문가들의 협조를 받아 실제 스키마 구조와 데이터 정확성 및 분류체계의 검토를 받은 후, 해당 내용을 본 온톨로지에 반영함으로써 스키마와 데이터의 신뢰성을 갖출 수 있었다. 이러한 접근법을 활용한 이유는 의료용어를 분류 및 정제하는 과정에서 오류가 발생한다면, 혈액종합검사 온톨로지의 오염을 일으킬 뿐만 아니라 경험지식 플랫폼에서 온톨로지를 활용하는 모듈의 성능에도 악영향을 끼칠 수 있기에 2차 검토를 기반으로 온톨로지 데이터를 수록하였다.

본 연구팀은 각 동의어 묶음을 대표하는 대표용어를 정의하기 위해 KOSTOM 개별코드 및 대표코드를 활용했다. KOSTOM 개별코드는 각 항목별로 부여된 주 키의 역할을 하고 있다. KOSTOM 대표코드는 비슷한 용어를 하나의 개념으로 묶는 역할을 한다. 본 연구팀은 해당 코드를 그대로 활용할 수 없었다. 왜냐하면 혈액종합검사란 특수 분야에 관련된 의학용어가 KOSTOM 에 모두 수록되어 있지 않았기에, 추가로 외부 리소스에서 용어를 수록할 필요가 있었기 때문이다. 또한, 실제 현장에서 사용하는 전산시스템에서 발생하는 의학용어를 수록하기 위해서 추가적인 코드 체계가 필요했다.

이러한 이유로 본 연구팀이 제안한 혈액종합검사 온톨로지에서는 EKP_CODE 를 활용함으로써 추가적인 데이터 확장을 가능하게 했으며, KOSTOM 의 코드 정보를 독립적으로 수록함으로써 기존 KOSTOM 의 원래 구조와도 연동이 가능하도록 하였다. 본 연구팀은 EKP_CODE 를 활용하여 대표용어와 매핑을 시켰다. 대표용어가 명확히 지정되지 않은 경우에는 데이터 정제 프로세스에서 전문의의 협조를 통해 동의어 목록을 검토한 후, 각 용어를 가장 잘 표현하고, 사용빈도가 높은 단어를 대표용어로 선정하여 EKP_CODE 를 부여하였다.

본 연구팀이 혈액종합검사 온톨로지를 설계하며 고려한 부분 중 하나는 기존에 사용중인 전산시스템과의 연계성 보장과 외부 리소스를 자유롭게 추가할 수 있는 온톨로지의 확장가능성이었다. 경험지식플랫폼 개발연구에서 중점을 두고 있는 사항 중 하나는 현장전문가의 경험지식을 획득하여 활용하는 것이었다. 그렇기에 기존에 현장에서 사용하고 있는 전산시스템에 저장되어 있는 정보를 온톨로지에 수록하려고 시도했다.

현장전문가인 의료진이 혈액종합검사 결과치를 보고 실제로 어떠한 의료용어 및 표현을 소견서에 작성하는지에 대한 정보를 온톨로지에 반영함으로써, 현장 및 기존 전산시스템이 갖고 있는 경험지식을 획득할 수 있다고 판단했다. 이를 위해 현장에서 사용된 소견문을 전처리하여 일반명사를 추출했다. 소견문에서 등장하는 용어 중 온톨로지에 이미 수록된 용어들을 제거한 후, 본 연구팀과 현장 전문가들의 이중 검토를 거쳐 소견문에서 발생하는 현장용어를 추가로 온톨로지에 반영하였다.

KOSTOM은 의학용어를 표준화하면서 수많은 종류의 질환을 개념화하고 국내외 의료 온톨로지에 대한 코드를 수록한 장점이 있었지만, 혈액종합검사를 위한 검사항목에 대한 용어 정리에 대해서는 부족함이 있었다. 그렇기에 본 연구팀은 이러한 한계를 보완하기 위해서 대한진단검사의학회에 나타나는 검사항목 정보를 웹크롤링을 통하여 검사항목 용어 및 동의어를 추출하였다. 그리고 한글명, 영문명과 같이 독립적으로 분류한 후, 해당 데이터를 온톨로지에 반영하였다.

본 연구팀은 혈액종합검사 온톨로지를 만들기 위한 각종 리소스 목록을 탐색하여, 혈액종합검사와 관련된 의학용어를 수록하기에 가장 적합한 기초 리소스로 KOSTOM을 선정하고, 혈액종합검사 스키마를 설계했다. 또한, 데이터 정제 및 대표용어 선정 프로세스를 고안하였으며 KOSTOM 데이터가 반영하지 못하는 정보를 외부 리소스를 참조하여 추가함으로써 온톨로지의 신뢰도 및 활용성을 높였다.

4. 결과

본 연구팀이 혈액종합검사를 위한 온톨로지 구축에 사용한 데이터는 표 3과 같다. KEYWORD 테이블은 검사항목 105개, 질환 221개, 검체 3개, 그리고 검사구분 22개에 대한 대표용어를 수록한다. 또한, EN_SYN, KR_SYN 테이블에서는 대표용어들을 다르게 부를 수 있는 영문 동의어 1195개 한글 동의어 1044개가 수록되어 있다. 또한, 혈액종합검사 온톨로지를 구성하는 개념간 관계를 정의한 테이블은 BLOODTEST와 ASSOCIATIVE_BLOOD_DISEASE으로써 각각 105개 377개의 관계정보를 수록하고 있다. 그리고 각 의료용어에 대응되는 외부 온톨로지의 코드 정보는 약 694쌍이 존재한다.

본 연구팀이 구축한 혈액종합검사 온톨로지는 다양한 측면에서 활용 및 응용이 가능하다. 먼저 본 온톨로지는 경험지식플랫폼의 하위 모듈인 문서 분류 및 처리작업에서 활용될 수 있다. 적절한 문서를 찾기 위해서는 적절한 질의어가 필요하다. 그렇기에 올바르게 구조화되고 정제된 용어사전이 필요하다. 더 나아가, 질의어를 만들기 위한 단순 동의어 사전

표 3. 혈액종합검사 온톨로지 데이터 확보 현황

자료명	개수	참조 테이블
Component (검사항목: 대표용어)	105 개	KEYWORD
Disease (질환: 대표용어)	221 개	KEYWORD
System (검체: 대표용어)	3 개	KEYWORD
Category (검사구분: 대표용어)	22 개	KEYWORD
English synonym(영문 동의어)	1195 개	EN_SYN
Korean synonym(한글 동의어)	1044 개	KR_SYN
검사항목-검체 관계	105 개	BLOODTEST
검사항목-검체-질환-측정치 관계	377 개	ASSOCIATIVE_BLOOD_DISEASE
외부 코드(KOSTOM, UMLS, KCD)	694 쌍	KEYWORD_CODE_MATCHING

역할뿐만 아니라 해당 용어와 관련되어 나타나는 다양한 의미관계 정보(검체, 검사항목, 질환, 수준)가 포함되어 있어, 더욱 정확한 검색 결과를 기대할 수 있도록 한다.

또한, 혈액종합검사 온톨로지는 의학용어 분야를 도메인으로 다룸으로써 발생할 수 있는 잘못된 데이터 분류 및 정제 작업의 문제점에 주목하였다. 이러한 문제점을 예방하기 위해 구체적인 용어 정제 절차를 고안하고, 이를 반영한 결과를 온톨로지에 수록하였다. 또한, 전문의 협조를 통해 온톨로지에 수록된 용어목록과 관계성을 전수 검토함으로써 데이터 신뢰성을 갖출 수 있었다.

혈액종합검사 온톨로지를 구성하는데 토대가 된 KOSTOM의 의학용어를 수록하는데서 그치지 않고, UMLS, KCD와 같은 외부 온톨로지와의 확장적 연계를 고려하여 관련 코드를 수록함으로써 의학용어 국제 표준화와의 연계를 도모했다.

끝으로 본 연구는 경험지식플랫폼 개발연구의 하위 모듈로 활용될 혈액종합검사 온톨로지 구축 연구로써, 경험지식플랫폼 내부에서 사례판단을 위한 룰을 생성하는 메타지식으로 활용 가능하다. 그리고 실제 의료기관의 전산서비스 체계와의 연동을 도모함으로써 유기적인 온톨로지 시스템 확장을 가능하게 한다.

5. 결론

본 연구팀은 경험지식플랫폼 개발연구에서 활용될 하위모듈인 혈액종합검사 온톨로지를 구축했다. 기존에 UMLS, MeSH, SNOMED, KOSTOM, 그리고 KCD 와 같은 의학 분야의 온톨로지가 존재했지만, 이러한 연구결과를 바로 본 연구에 적용할 수 없는 문제가 있었다. 왜냐하면 혈액종합검사를 위한 검사항목, 검체, 질환, 그리고 측정치 수준과 같은 개념 사이의 관계 정의가 없었기 때문이다. 이러한 문제로 인해 본 연구팀은 KOSTOM 리소스를 기초로 활용해서, 부족한 의학용어를 외부 리소스를 활용하여 보강한 후, 혈액종합검사에 필요한 개념간의 관계를 관계형데이터베이스로 정의함으로써, 혈액종합검사 온톨로지 구축을 시도했다.

본 연구팀이 제안한 혈액종합검사 온톨로지는 다음과 같은 장점을 지닌다. 하나, 본 연구는 경험지식플랫폼의 하위 모듈 개발에 대한 연구일 뿐만 아니라, 혈액종합검사를 위한 개념간 관계성 정보를 데이터베이스를 활용하여 구축함으로써 다양한 현장에서 사용될 수 있는 혈액종합검사 온톨로지를 구축하였다. 둘, 대표용어 및 동의어 목록 구축으로 질의어 확장과 같은 문서 검색을 위한 지식베이스로 활용 가능하다. 셋, 체계적인 용어 검토 프로세스를 고안하고 반영함으로써 혈액종합검사 온톨로지에 대한 신뢰성을 갖추었다. 넷, KOSTOM, UMLS, KCD 와 같은 다양한 외부 온톨로지에 대한 코드정보를 수록함으로써 외부정보 참조 및 확장을 용이하도록 만들었다.

본 연구에서의 한계점으로는 의료현장에서 기존에 사용하고 있던 전산시스템 내부의 소견문 데이터에 등장하는 현장용어를 사전에 한번만 추출하여 온톨로지 반영한 부분이다. 이로써 추가로 발생하는 용어를 지속적으로 온톨로지에 반영하지 못한다는 한계가 있다. 무조건적으로 소견문에 나오는 용어를 모두 온톨로지에 수록하는 것은 온톨로지의 오염을 초래하는 문제를 야기할 수도 있으나, 소견문에 등장하는 용어에 대한 지속적 모니터링 기능을 개발한다면, 소견문에서 발생하는 유사단어를 판별하여 자동적으로 속성에 맞게 분류하고 온톨로지를 확장하는 세부 모듈 개발을 연구할 수 있을 것이다.

본 연구팀은 다음과 같은 후속 추가 연구를 제안한다. 하나, 지식경험플랫폼의 범위가 혈액종합검사뿐만 아니라 장기적으로는 조직검사로 확장될 필요가 있다. 그렇기에 해당 범위를 커버하기 위해 온톨로지 역시 확장될 필요가 있다. 조직검사 온톨로지가 구축된다면, 다양한 질환과 검사항목들 간의 관계성에 대해 구조적인 지식체계를 갖추으로써 조직검사를 대상으로 하는 의사결정지원시스템에 적용할 수 있을 것이다. 둘, 현재 혈액종합검사 온톨로지에는 검사항목, 검체, 질환, 검사항목 측정치 수준에 대

한 개념 사이의 관계성을 포함하고 있다. 추후에는 더욱 정교한 관계설정을 위해 사람의 나이, 성별 등 혈액종합검사에 영향을 미칠 수 있는 인적 요인의 관계정보를 추가할 필요가 있다. 끝으로 의료진이 입력하는 소견문에서 새로운 의학용어가 발견되면, 자동적으로 동의어를 추출하여 온톨로지를 확장하는 기능에 대한 개발연구 필요성이 존재한다. 이러한 후속 연구가 진행된다면 의료진단검사 영역에서 유용하게 활용될 수 있는 메타지식을 구축하는데 기여를 할 수 있을 것이다.

본 연구팀이 수행한 혈액종합검사 온톨로지 구축을 위한 연구는 경험지식플랫폼의 모듈로 활용되는 메타지식을 구축하기 위한 연구일 뿐만 아니라, 혈액종합검사 온톨로지 구축이란 세분화된 의료영역의 온톨로지 구축에 대한 방법을 제시한다. 그렇기에, 본 연구팀의 혈액종합검사 온톨로지 구축 연구는 혈액종합검사 스키마 설계에 대한 방향성을 제시하고 있다. 또한, 실제 의료시스템과의 연동이라는 실용적인 측면을 강조함으로써, 본 연구팀이 제안한 온톨로지가 실제 혈액종합검사를 위한 전산시스템에 적용 가능하도록 설계하였다. 끝으로, 본 연구는 혈액종합검사라는 부분적인 진단검사 영역에 대한 온톨로지 구축 연구를 했지만, 본 연구팀이 제안한 온톨로지 구조와 구축 방법론은 추후 조직진단 검사 등 확장된 분야에서의 후속연구 방향을 제시할 수 있을 것이다.

참고문헌

- [1] Doi, Kunio. "Computer-aided diagnosis in medical imaging: historical review, current status and future potential." *Computerized medical imaging and graphics* 31.4 (2007): 198-211
- [2] 최호섭, et al. "온톨로지 구축 방법과 사례." *정보과학회지* 24.4 (2006): 31-44.
- [3] 안선주, et al. "의료정보의 의미적 상호운용성 보장을 위한 임상콘텐츠 모델." *정보과학회논문지: 소프트웨어 및 응용* 37.12 (2010): 871-881.
- [4] 고영만. "시소러스 기반 온톨로지에 관한 연구." *성균관대학교, 정보관리* 5 (2006).
- [5] Gene Ontology Consortium. "The Gene Ontology (GO) database and informatics resource." *Nucleic acids research* 32.suppl 1 (2004): D258-D261.

- [6] 이상민, 김화선, and 조훈. "인체자원은행의 효율적인 운영을 위한 OWL 기반의 데이터베이스 구축에 관한 연구." (2015).
- [7] 이승우, et al. "노령 척추 정보 데이터베이스 및 관리 도구 설계." 한국콘텐츠학회 2010 춘계 종합학술대회 (2010): 539-541.
- [8] 김철, 김상균, and 송미영. "논문분석과 구축사례 조사를 통한 한의학 온톨로지 연구동향 분석." 한국한의학연구원 논문집 14.2 (2008): 121-129.
- [9] Bodenreider, Olivier. "The unified medical language system (UMLS): integrating biomedical terminology." *Nucleic acids research* 32.suppl 1 (2004): D267-D270.
- [10] 정소나, and 이춘실. "NLM Medical Text Indexer를 활용한 우리나라 의학문헌의 MeSH Semi Indexing 방안." 제 17 회 한국정보관리학회 학술대회 논문집 (2010): 21-28.
- [11] Stearns, Michael Q., et al. "SNOMED clinical terms: overview of the development process and project status." *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001.
- [12] 김지영. "보건의료정보 용어표준의 국제동향." *전자공학회지* 40.1 (2013): 30-38.