

인체자원은행의 효율적인 운영을 위한 OWL 기반의 데이터베이스 구축에 관한 연구

이상민¹⁾, 김화선²⁾, 조훈³⁾

Study on OWL-based database built for the efficient operation of human resources bank

Sang Min Lee¹⁾, Hwa Sun Kim²⁾, Hune Cho³⁾

요 약

본 연구에서는 인체자원은행의 효율적인 운영과 인체자원은행간의 상호호환을 위해 온톨로지 기반의 데이터베이스를 구축하였다. 본 연구팀은 선행연구에서 7,197,252개의 인체자원정보를 수집하여 고품질 인체자원정보 데이터베이스를 구축하였다. 이 과정에서 1,796개의 임상항목을 분류하고 명확한 용어표현을 위해 국제표준용어체계인 SNOMED-CT와 매핑하였다. 그 결과 동일한 국제표준코드에 대해 서로 다른 용어 및 값 유형, 자료 형을 가지는 826개의 유의어를 발견하였다. 본 연구에서는 유의어로 인한 중복을 최소화하기 위해 76개의 선호어를 선정하였다. 또한 인체자원은행간의 상호호환을 위해 유의어와 선호어 간의 관계를 OWL을 이용하여 정의하고 XML 및 데이터베이스에 저장하였다. 본 연구에서 구축한 OWL 기반의 데이터베이스는 공통된 용어체계를 가지며, 인체자원은행 간의 복잡한 매핑 과정 없이 상호호환 할 수 있다.

핵심어 : 인체자원은행, 유의어, 선호어, 온톨로지, 국제표준용어체계

Abstract

The purpose of this study was to build Ontology based database for efficient operation and interoperability of the human resource bank. The team 7,197,252 of human resources information was collected through previous studies and classifies the 1,796 clinical item through high quality work. For a clear expression of clinical terminology mapping and item of international standard terminology

Received (April 15, 2015), Review Request(April 16, 2015), Review Result(May 01, 2015)

Accepted(May 21, 2015), Published(June 30, 2015)

¹700-842 Dept. of Medical Informatics, Kyungpook National Univ., Gukchaebosang-ro, Jung-gu, Deagu, Korea

email: yth2320@naver.com

²712-715 Dept. of Faculty of Medical Industry Convergence, Daegu Haany Univ., Yugok-dong, Gyeongsan-si, Gyeongsangbuk-do, Korea

email: daspula@daum.net

³(Corresponding Author) 700-842 Dept. of Medical Informatics, Kyungpook National Univ., Gukchaebosang-ro, Jung-gu, Deagu, Korea

email: medinfo@knu.ac.kr

SNOMED-CT it was found to have 826 different synonyms of the term and value types, the data type for the international standard code. In this study, we selected a preferred term of 76 to minimize the redundancy due to the synonyms. Also define the relationship between synonyms and preferred terms using the OWL for interoperability between human resources banks and were stored in XML and databases. OWL-based database built in this study has a common terminology, it can be interchangeably without complex mapping process between the human resource banks.

Keywords : human resource bank, synonym, preferred term, ontology, international standard terminology

1. 서론

세계 각국에서는 국민 건강 증진을 위해 국가적인 차원에서 인체자원을 활용한 연구 사업을 시행하고 있다[1-3]. 인체자원은행은 예방 및 진단, 질병을 치료하는 새로운 방법을 찾기 위해 다양한 인체자원은행으로부터 인체자원 샘플을 체계적으로 수집한다[4]. 이러한 인체자원은 일반적으로 질병의 원인이 되는 유전자 및 생활패턴, 환경적 위험 요인을 식별하는 가장 효율적인 수단이며, 통계적인 분석을 위해 두 개 이상의 인체자원은행으로부터 샘플 데이터를 사용한다[5][6].

인체자원은행은 서로 다른 집단에서 인체자원을 수집하기 때문에, 여러 집단에서 제공하는 데이터는 이질적이며, 용어체계에 대한 서로 다른 이해와 규제로 인해 데이터 통합이 어렵다[5]. 즉 서로 다른 인체자원은행에서 진단 코드를 다르게 해석함으로써 인체자원은행 간의 정보 공유를 어렵게 한다[4][7]. 이러한 문제에 대해 Ontology Web Language (OWL)는 인체자원정보 간의 관계를 의미적으로 비교하고 정의 할 수 있는 도구이며, 데이터 모델의 설계 및 구현을 용이하게 한다[8].

대한민국은 2007년부터 한국인체자원은행을 설립하여 연구에 필요한 인체자원을 수집, 관리 분량을 하고 있다[9]. 현재 17개 인체자원은행은 인체자원정보를 개별적으로 수집하고 있으며, 통합되지 않은 정보로 인해 정확한 정보 제공이 어렵다. 이에 본 연구팀은 인체자원은행의 임상 및 검체 정보를 수집하고 쉽게 활용하기 위해 선행 연구를 수행하였다. 선행 연구는 15개 인체자원은행의 임상 및 검체 정보를 수집하고 이에 대한 분류체계를 정의하였다. 또한 국제표준용어체계인 Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT)을 이용하여 임상항목에 대한 명확한 표현을 정의하였지만, 이 과정에서 유의어 문제가 발생되었다.

유의어(synonym)는 동일한 국제표준코드에 대해 서로 항목 명과 값 유형(value type), 자료 형(data type)을 가지는 2개 이상 임상항목을 의미한다. 예를 들어, '성별' 항목에 대한 유의어는 '성별(Q00114.01)', 'Donor정보 : 성별'과 같이 서로 다른 항목 명으로 표현된다. 또한 '성별' 항목은 '(F, M)', '(여, 남)', '(1, 2)'와 같은 값 유형에 따라서도 유의어로 분류된다. 이러한 유의어는 인체자원은행의 테이블 정규화 과정에서 1,796개 임상항목 중 826개가 발견되었다. 이처럼 유의어 문제가 지속되게 된다면, 향후 인체자원은행 간의 정보 통합 시 데이터의 상호호환을 어려우며, 데이터를 전송하기 위해서는 데이터베이스 스키마 내의 소스 데이터를 매번 수정해야 한다[7].

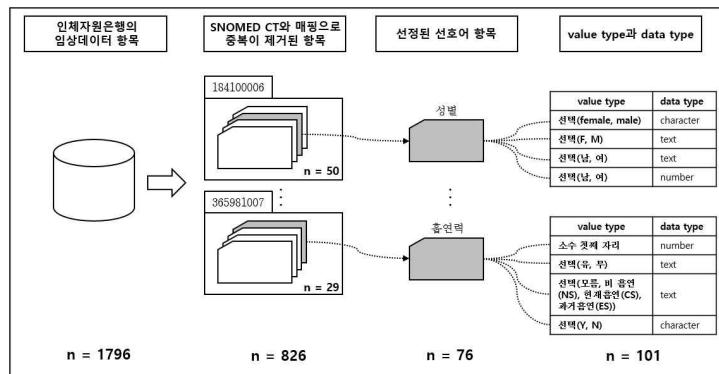
이에 본 연구는 선호어를 선정하여 유의어로 인해 중복된 컬럼(column)을 최소화 하고 인체자원은행 간의 인체자원정보를 상호호환 할 수 있도록 OWL 기반의 데이터베이스를 제안한다. OWL

기반의 데이터베이스는 선호어와 유의어 간의 관계를 정의하며, 인체자원정보를 필요로 하는 인체 자원은행이 분양받을 때 데이터베이스의 컬럼과 자료 형 변경 없이 호환 가능한 기능을 제공한다.

2. 관련 연구

2.1 유의어와 선호어

유의어는 동일한 의미를 가지는 2개 이상의 항목을 의미하며, 선호어는 중복되는 항목을 대표하는 고유한 항목을 의미한다. 이처럼 유의어로 인해 중복되는 항목은 선호어를 선정함으로써 항목의 개수를 줄일 수 있다[10]. 본 연구팀은 선행 연구에서 15개 인체자원은행의 인체자원정보를 수집하여 상세화, 분류화, 표준화를 통해 고품질 인체자원정보 데이터베이스를 구축하였다. 15개 인체자원은행은 각 인체자원은행마다 개별적으로 인체자원정보를 수집하고 있어 용어를 정의하는 용어체계가 달랐다. 이로 인해 같은 의미를 가지는 항목이 서로 다른 용어 표현으로 유의어 문제가 발생되었으며, 데이터베이스의 테이블 정규화 과정이 어려웠다. 이러한 문제를 해결하기 위해 2명의 자문위원을 선정하고 1,796개의 임상항목 데이터를 분석하여 국제표준용어체계인 SNOMED-CT의 concept id와 매핑하였다. 매핑 결과, 용어 표현은 다르지만 같은 concept id를 가지는 항목들이 발견되었고, 이들 항목 중 같은 concept id를 가지는 항목을 유의어 그룹으로 정의 하였다. 하나의 concept id에 많게는 50개의 중복된 자료가 제거되었으며, 1,796개 임상항목 중 826개 유의어 항목을 발견 되었다. 유의어를 분석한 결과, 유의어 항목 간에 값 유형과 자료 형이 다양하게 정의 되어 있었으며, 값 유형과 자료 형이 다른 경우 항목이 표현하는 값이 달라지므로 서로 다른 항목으로 정의하였다. 하나의 유의어는 최대 8개의 값 유형이 있었으며, 비슷한 값 유형 간에 들쭉 뭉고 값 유형과 자료 형이 다른 경우에 구분하여 최대 4개의 그룹으로 표현이 하였다. 이러한 유의어 항목 중 값 유형을 가장 많이 표현 할 수 있는 항목을 선호어로 선정하여 '성별', '흡연력'과 같이 76개의 항목이 분류되었고, 선호되는 값 유형과 자료 형에 따라 101개로 구분된다(Fig. 1).



[그림 1] 임상 항목의 선호어 선정 과정

[Fig. 1] The preferred term selection process of clinical item

2.2 Ontology Web Language

OWL은 온톨로지를 표현하는 다른 언어에 비해 표현력이 가장 뛰어난 시멘틱 웹 온톨로지 언어이며, 서로 다른 기종의 응용 프로그램 사이에서 웹 기반의 지식을 처리한다[11][12]. OWL은 OWL Lite, OWL DL 및 OWL Full과 같은 세 개의 하위 언어가 있으며, 후자로 갈수록 표현력이 높지만 효율적인 추론이 어렵다. OWL DL과 OWL Full은 불 연산자를 통해 복잡한 객체의 형태를 정의하고 교집합, 합집합, 여집합을 정의할 수 있어 다양한 표현이 가능하다. OWL Full은 표현력이 가장 높은 반면 추론기능을 완벽히 지원하는 시스템이 없으므로 현재 OWL DL을 가장 많이 사용한다[13]. 따라서 본 논문에서는 온톨로지 표현 언어로 OWL DL 선택하여 연구를 진행하였다.

각 인체자원은행마다 용어를 분류하는 용어체계가 다르기 때문에, 공유 대상이 되는 모든 인체자원은행의 용어체계를 습득하여야 하며, 인체자원은행에 맞게끔 용어체계를 변경해야 하는 문제점이 있다. 따라서 본 연구는 OWL DL을 이용하여 인체자원은행마다 개별적으로 정의한 용어체계 간의 관계를 정의하고 하나의 통합된 용어체계를 구축한다. 이러한 OWL 기반의 통합된 용어체계는 데이터베이스와 연계하여 정보 교환 시 용어체계를 변경해야 하는 번거로움을 줄이며, 효과적인 인체자원은행 관리 및 상호호환이 가능한 통합 관리 시스템을 구축할 수 있게 한다.

2.3 관련 연구 동향

인체자원은행이 수집하는 인체자원정보는 개별적으로 수집되고 인체자원은행마다 임의의 데이터 인코딩을 사용하고 있어 데이터 통합이 어렵다[7]. 특히 의학 용어는 넓고 복잡한 개념을 다루고 있기 때문에 논리적 추론 기반의 데이터베이스 설계가 필요하다[14]. 이에 따라 다수의 인체자원은행에서는 인체자원정보를 효율적으로 검색하기 위해 용어 간의 관계를 표현하고자 OWL을 이용한 개발이 진행되었다. 대표적으로 National Center for Biomedical Ontology (NCBO) Resource Index는 수집된 인체자원을 통합된 데이터베이스로 구축하기 위해 다양한 영역의 연구자에 의해 공개된 200여개의 온톨로지 집합을 사용하였다[15]. Nottingham 대학 병원 NHS Trust에서는 단위 은행으로부터 수집된 인체자원을 코호트기반의 연구를 위해 인체자원은행의 임상 정보 모델을 OWL/XML 기반으로 개발하였다[16].

국내외의 온톨로지 관련 연구들은 온톨로지의 추론 기능을 이용하여 질병 연구와 관련된 의미 있는 정보를 검색하는 지능형 정보 인프라 구축에 대한 연구가 주로 이루어졌다. 인체자원은행의 관리적인 차원에서는 온톨로지 기반의 데이터베이스 구축이나 온톨로지 간의 매핑을 통해 문제를 해결하였다. 이에 반해, 본 연구는 국제표준용어체계인 SNOMED-CT와 매핑으로 발견된 유의어 항목에서 선호어를 선정하여 공통된 용어 표현을 정의하고, 각 용어가 가지는 속성 간의 관계를 OWL을 이용하여 정의하는 점에서 기존 연구들과 차별된다.

3. 설계 및 구현

인체자원은행 간의 상호호환이 보장되기 위해서는 수집 대상이 되는 인체자원은행의 데이터베이스에 구조와 속성이 반영이 되어야 하며, 의미상 동일한 항목 간에는 통합을 위해 관계 정의가 이루어져야 한다. 이를 위해 본 연구는 인체자원은행 간의 정보가 상호호환 할 수 있도록 OWL 기반의 데이터베이스를 설계하였다. 설계한 OWL 기반의 데이터베이스는 선행 연구에서 발생되었던 데이터베이스 정규화 문제에 대해 선호어를 선정하여 해결하였으며, 상호호환을 위해 유의어와 각 속성들 간의 관계를 OWL을 이용하여 정의하였다. 선호어 선정을 위해 유의어 항목을 분석하여 여러 개로 표현되는 유의어 항목과 항목이 가지는 속성들을 정리하였다(Table 1).

[표 1] '성별' 항목의 속성 분류

[Table 1] Property classification of 'gender' items

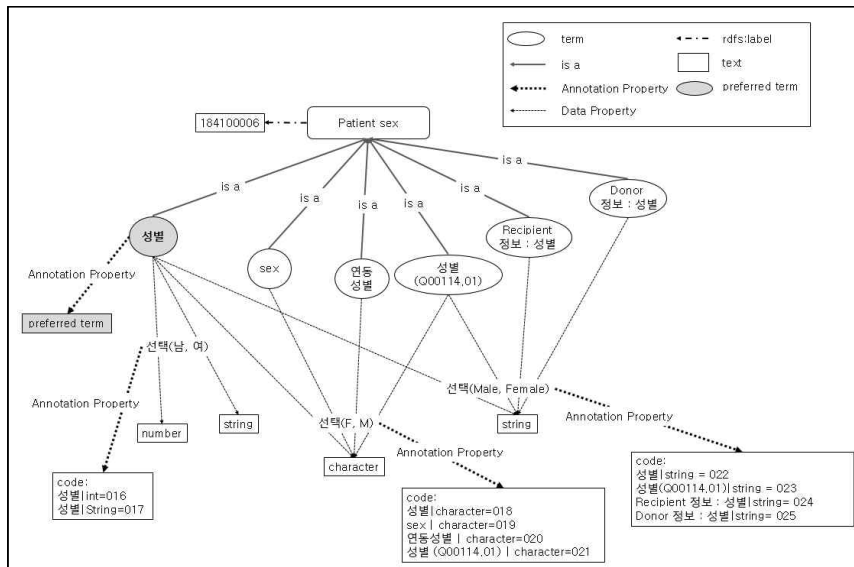
SNOMED-CT Concept	preferred term	synonym	value type	data type
patient sex	성별	성별	선택(F, M)	character
			선택(female, male)	text
			선택(남, 여)	number
			선택(남, 여)	text
		sex	선택(F, M)	character
		Donor정보 : 성별	선택(Male, Female)	text
		연동성별	선택(M, F)	character
	Recipient정보 : 성별	선택(Male, Female)	text	

동일한 SNOMED-CT의 concept에 대해 최대 9개의 의미적으로 동일한 유의어 관계를 발견하였다. 또한 유의어 항목 내에서도 값 유형과 자료 형에 따라 입력되는 데이터의 내용이 달라지므로 동일한 값 유형과 자료 형에 따라 세부 그룹을 구분하였다. 예를 들어 SNOMED-CT의 concept이 'patient sex'인 경우 성별이란 항목이 가지는 값 유형이 '선택(F, M)', '선택(female, male)', '선택(남, 여)', '선택(1, 2)'에 따라 세부 그룹으로 구분된다. 이러한 유의어 중 선호어의 선정 방식은 가장 많은 값 유형을 표현하는 항목을 선호어로 선정하였다. 그 결과 '성별'이란 항목이 가장 많은 값 유형을 가지므로 선호어로 선정되었다.

상호호환을 위한 유의어와 선호어 간의 관계표현은 OWL를 이용하였다. SNOMED-CT의 concept과 매핑되는 유의어는 명확한 용어 표현이 되는 concept의 의미를 상속하고 여러 가지의 용어로 표현됨으로 is a 관계로 정의하였다. 예를 들어, SNOMED-CT의 concept인 patient sex에 대해 is a 관계에 있는 하위 계층은 'Donor 정보 : 성별', '연동성별', '성별(Q00114.01)' 등의 항목들이며, 이들은 유의어 관계이다. 유의어 간에 표현하는 속성들이 각각 다르므로 data type 속성을 이용하여 정의하였다. 유의어 중 선호어 항목에 대해서는 Annotation 속성을 이용하여 선호어를 표시하였다. 항목의 구분하는 고유 식별자는 label를 이용하여 값을 정의하였다.

4. 결과

본 연구는 OWL 기반의 데이터베이스를 구축하여 컬럼 간의 관계를 정의하였다. OWL 기반의 데이터베이스를 구축한 결과, 1,796개 임상항목을 SNOMED-CT의 concept과 매핑하여 826개 유의어 항목 중 76개 선호어를 선정하여 데이터의 컬럼 중복을 제거하였다. 또한 인체자원은행 간의 상호호환을 위해 OWL을 이용하여 정의하였다. OWL을 이용한 관계 표현을 가시적으로 표현하기 위해 개발도구인 Protege 5.0를 이용하였으며, 구현한 결과 자세한 관계 표현을 확인하고자 일부분만 다이어그램 형태로 표현 하였다(Fig. 2). 유의어는 명확한 표현을 위해 SNOMED-CT의 concept을 상속받아 다양한 형태로 표현하였으며, 이들 간에 관계는 상속관계를 표현하는 is a 관계를 이용하여 표현하였다. 유의어 간의 관계는 용어 표현과 값 유형에 따라 구분되며, 하나의 용어는 여러 개의 값 유형을 가진다. 이는 서로 다른 인체자원은행에서 사용하고 있는 값 유형이므로 상호호환을 위해 표현하였다. 차후 인체자원은행의 수집대상이 늘어날 경우에, 구축한 OWL은 추가 항목을 SNOMED-CT의 concept id와 매핑하여 유의어 항목 간에 관계와 속성 간의 관계를 정의한다. 따라서 OWL기반의 데이터베이스는 수집한 인체자원항목을 OWL기반으로 매핑하여 불명확한 용어표현과 항목 중복으로 인한 데이터베이스 정규화 과정의 어려움을 줄 일 수 있다.



[그림 2] 온톨로지를 이용한 개념 간의 관계 표현

[Fig. 2] Representation of the relationship between the concept of using ontologies

Protege 5.0으로 개발한 OWL은 XML문서로 나타낼 수 있다(Fig. 3). OWL을 XML 형태로 나타낼 경우 데이터 표현 및 관리하기 용이하고 데이터가 가지는 속성이 명시 되어 데이터 간의 상호

운용이 가능하다. OWL Class의 rdf:ID는 SNOMED-CT의 concept을 표현하고 하위 계층으로 표현되는 subClass에는 유의어 항목이 입력된다. OWL에는 두 가지의 Property가 있으며 용어 간의 관계 표현을 위해서는 Object Properties를 이용하여 표현하고 용어의 값 유형을 표현하기 위해서는 Data Type Properties를 이용하여 표현한다. 유의어 항목의 Data type Property는 owl:Restriction의 내에서 표현된다. Data type Property는 XML schema의 attributes의 요소와 매핑 한다. Data Type Properties 통해 76개의 선호어가 가지는 101개의 값 유형과 자료 형을 정의한다. domain은 표현될 데이터 용어를 나타내고 range는 용어의 속성인 자료 형을 나타낸다. Data type Property의 ID는 값 유형을 의미하며 중복된 표현이 있으므로 1에서 101까지 코드 값을 정의하였다.

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns="http://www.semanticweb.org/ism/ontologies/2015/1/untitled-ontology-27.owl" xmlns="
3 http://www.semanticweb.org/ism/ontologies/2015/1/untitled-ontology-27.owl#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdfs="
4 http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rd="http://www.w3.org/2000/01/rdf-schema#"
5 <owl:Ontology rdf:about="BioBank"/>
6 <owl:Class rdf:ID="age_at_starting_smoking">
7   <rdfs:SubClassOf rdfs:resource="#증연시작"/>
8   <owl:Restriction
9     <owl:DatatypeProperty rdf:ID="free text"/>
10    <rdfs:domain rdfs:resource="#증연시작"/>
11    <rdfs:range rdfs:resource="#text7"/>
12  />
13 />
14 <owl:Class rdf:ID="blood_Pressure">
15   <rdfs:SubClassOf rdfs:resource="#임상_항목_혈압">
16   <owl:Restriction
17     <owl:DatatypeProperty rdf:ID="순환기/이원기">
18     <rdfs:domain rdfs:resource="#임상_항목_혈압"/>
19     <rdfs:range rdfs:resource="#text7"/>
20   />
21 />
22 <rdfs:SubClassOf>
23 <rdfs:SubClassOf rdfs:resource="#혈압">
24 <owl:Restriction
25   <owl:DatatypeProperty rdf:ID="free text"/>
26   <rdfs:domain rdfs:resource="#혈압"/>
27   <rdfs:range rdfs:resource="#text7"/>
28 />
29 />
30 />
31 <owl:Class rdf:ID="blood_product_Unit_Identifier">
32   <rdfs:SubClassOf rdfs:resource="#혈액변호">
33   <owl:Restriction
34     <owl:DatatypeProperty rdf:ID="free text"/>
35     <rdfs:domain rdfs:resource="#혈액변호"/>
36     <rdfs:range rdfs:resource="#text7"/>
37   />
38 />
39 />
40 <rdfs:SubClassOf>
41 <rdfs:SubClassOf rdfs:resource="#혈액샘플번호">
42 <owl:Restriction
43   <owl:DatatypeProperty rdf:ID="free text"/>
44   <rdfs:domain rdfs:resource="#혈액샘플번호"/>
45   <rdfs:range rdfs:resource="#text7"/>
46 />
47 />

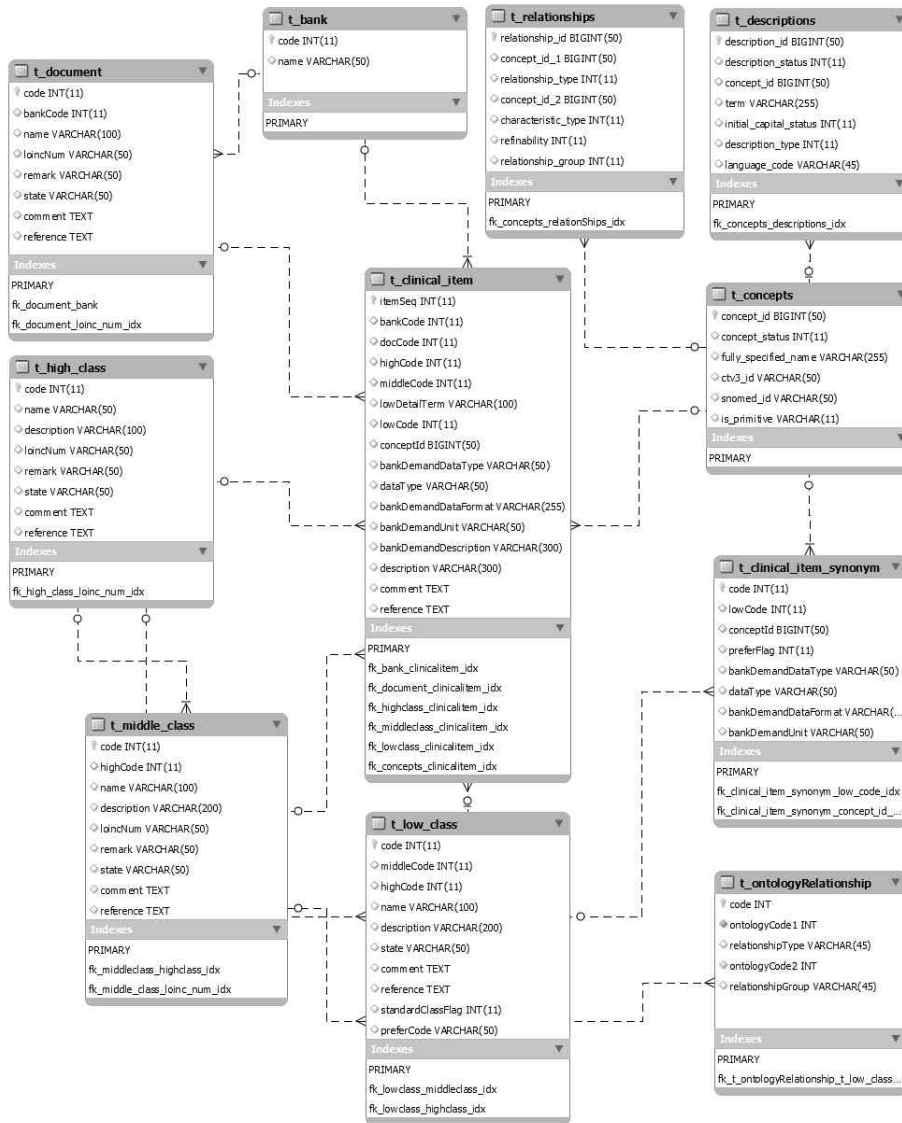
```

[그림 3] 선호어와 유의어를 관계를 표현한 OWL/XML 문서

[Fig. 3] OWL/XML document showing the relationship between preferred term and synonyms

OWL를 이용하여 정의된 관계는 효율적인 활용과 관리를 위해 데이터베이스에 저장하였다(Fig. 4). 임상항목은 t_clinical_item 테이블을 중심으로 총 11개의 테이블로 구성된다. t_clinical_item 테이블은 각 테이블의 고유 key 값들로 관계를 맺고 있으며, 임상항목을 SNOMED-CT와 매핑하여 SNOMED-CT의 세 가지 핵심테이블을 t_concepts, t_descriptions, t_relationships 테이블에 저장하였다. 인체자원정보는 선행연구의 분류화 과정을 통해 대분류, 중분류, 소분류로 나뉘어 졌으며 t_high_class, t_middle_class, t_low_class 테이블에 각각 저장하였다. 이중 상세화된 임상항목은 t_low_class 테이블에 저장되어 있으며, t_low_class 테이블에서 유의어 문제가 있는 항목들을

t_clinical_Item_synonym 테이블에 저장하였다. 선호어와 유의어 간의 관계가 저장된 OWL기반의 정보를 t_ontologyRelationships 테이블에 저장하였으며, t_ontologyRelationships 테이블의 relationshipType 컬럼에는 is a와 Data type Property등등 OWL에서 정의된 관계 표현이 저장되었다. 유의어 항목 간의 관계표현을 위해 ontologyCode1 컬럼은 domain을 ontologyCode2 컬럼에는 range를 나타내었다.



[그림 4] OWL 기반의 데이터베이스 설계
[Fig. 4] Design of OWL-based database

5. 결론

온톨로지 언어인 OWL을 이용하면 인체자원정보 간의 관계를 의미적으로 비교하여 정의 할 수 있으며, 인체자원정보 공유 시 인체자원은행 간의 공통된 용어체계를 가지게 한다. 이는 인체자원은행 간의 상호호환을 가능케 하고 향후 추가될 인체자원은행과 기본 인체자원은행 간의 복잡한 과정 없이 공통된 용어체계를 통해 매핑하는 장점이 있다.

이를 위해 본 연구에서는 한국인체자원은행에서 사용되고 있는 인체자원정보를 분석하여 유의어 항목에서 선호어 항목을 정의하였고, 데이터베이스 구축 시 항목중복으로 인해 공유가 어려웠던 문제를 OWL 기반의 통합된 데이터베이스 구축을 통해 해결하고자 제안하였다.

제안한 OWL 기반의 데이터베이스는 인체자원정보의 수집된 항목을 통합하기 위해 각 인체자원은행의 항목 데이터가 선호어 항목으로 통합하고 인체자원정보 분양 시 다시 원본 항목으로 수정 없이 전송 가능하도록 관계 정보를 이용하여 매핑된 인체자원정보를 보내게 된다.

OWL 기반의 통합된 데이터베이스를 통해 인체자원은행의 데이터는 효율적인 데이터를 수집과 분양이 가능하고 다른 인체자원은행에서 사용하는 OWL 기반의 데이터베이스와 매핑을 통해 국내 뿐만 아니라 국외의 인체자원을 모두 공유가 가능하다.

본 연구의 한계점은 임상항목에 데이터만 이용하여 항목의 수가 적으며 이로 인해 정의 할 수 있는 관계 또한 적게 나타났다. 차후 연구에서는 수집대상을 늘리고 임상 및 검체항목 전체에 대한 OWL 기반의 데이터베이스를 구축하고자 하며 차후연구 목표로서 항목 간의 매핑을 인체자원은행 담당자가 직접 할 수 있도록 온톨로지 매핑도구를 개발 할 계획이다.

References

- [1] P. Elliott and T. C. Peakman, The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *International Journal of Epidemiology*. (2008), Vol.37, No.2, pp.234-244.
- [2] Z. Chen, J. Chen, R. Collins, Y. Guo, R. Peto, F. Wu and L. Li, China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *International journal of epidemiology*. (2011), Vol.40, No.6, pp.1652-1666.
- [3] O. Park, S. Y. Cho, S. Y. Shin, J. S. Park, J. W. Kim and B. G. Han, A strategic plan for the second phase (2013 - 2015) of the Korea biobank project. *Osong Public Health and Research Perspectives*. (2013), Vol.4, No.2, pp.107-116.
- [4] M. KRESTYANINOVA, O. SPJUTH, J. HASTINGS, J. DIETRICH and D. REBHOLZ-SCHUHMANN, Biobank metaportal to enhance collaborative research: sail. *simbioms. org. Proceedings of ICTA*. (2011).
- [5] A. Q. Andrade, M. Kreuzthaler, J. Hastings, M. Krestyaninova and S. Schulz, Requirements for semantic biobanks. *Stud Health Technol Inform*. (2012), Vol.180, No.1, pp.569-573.

- [6] J. Muilu, L. Peltonen and J. E. Litton, The federated database - a basis for biobank-based post-genome studies, integrating phenome and genome data from 600 000 twin pairs in Europe. *European Journal of Human Genetics*. (2007), Vol.15, No.7, pp.718-723.
- [7] R. Wynden, M. G. Weiner, I. Sim, D. Gabriel, M. Casale, S. Carini and R. J. Cucina, Ontology mapping and data discovery for the translational investigator. *AMIA Summits on Translational Science Proceedings*. (2010), Vol.66.
- [8] M. Brochhausen, M. N. Fransson, N. Kanaskar, M. Eriksson, R. Merino-Martinez, R. A. Hall and J. E. Litton, Developing a semantically rich ontology for the biobank-administration domain. *J. Biomedical Semantics*. (2013), Vol.4, No.23.
- [9] S. Y. Cho, E. J. Hong, J. M. Nam, B. Han, C. Chu and O. Park, Opening of the National Biobank of Korea as the infrastructure of future biomedical science in Korea. *Osong Public Health and Research Perspectives*. (2012), Vol.3, No.3, pp.177-184.
- [10] J. J. Cimino, Auditing the unified medical language system with semantic methods. *Journal of the American Medical Informatics Association*. (1998), Vol.5, No.1, pp.41-51.
- [11] J. D. Tenenbaum, P. L. Whetzel, K. Anderson, C. D. Borromeo, I. D. Dinov, D. Gabriel and P. Lyster, The Biomedical Resource Ontology (BRO) to enable resource discovery in clinical and translational research. *Journal of biomedical informatics*. (2011), Vol.44, No.1, pp.137-145.
- [12] C. G. Hwang, H. Y. Shin, S. H. Lee and K. D. Jung, The Study of Service Integration using Ontology-based Service Association. *The Society of Digital Policy and Management*. (2014), Vol.12, No.2, pp.327-333.
- [13] S. K. Rho, J. S. Park, Editor, *The key of the internet evolution: Ontology, gods Toy business*, Seoul (2007).
- [14] G. Olund, P. Lindqvist and J. E. Litton, BIMS: An information management system for biobanking in the 21st century. *IBM Systems Journal*. (2007), Vol.46, No.1, pp.171-182.
- [15] C. Jonquet, P. LePendu, S. Falconer, A. Coulet, N. F. Noy, M. A. Musen and N. H. Shah, NCBO Resource Index: Ontology-based search and mining of biomedical resources. *Web Semantics: Science, Services and Agents on the World Wide Web*. (2011), Vol.9, No.3, pp.316-324.
- [16] J. J. Chelsom, I. Pande, R. Summers, I. Gaywood, Editors. *Ontology-driven development of a clinical research information system*. *Computer-Based Medical Systems (CBMS)*, 2011 24th International Symposium on. IEEE, (2011) June 27-30; Bristol, United Kingdom