

표준 의학 용어체계의 매핑을 위한 시스템의 설계 및 개발

Design and Development of a System for Mapping of Medical Standard Terminologies

이인근* · 김화선** · 조훈**

In Keun Lee*, Hwa Sun Kim** and Hune Cho**

* 경북대학교 의료정보학과

** 대구한의대학교 IT의료산업학과

요 약

의학 분야에서의 다양한 표준 용어체계는 각기 다른 형태로 구성되어 있다. 따라서 이들을 통합하여 활용하기 위해서는 용어체계 사이의 연결 정보가 필요하다. 이를 위해 여러 통합 도구들이 개발되어 사용되고 있으나, 이들 도구들은 특정 용어체계에 국한하기 때문에, 매핑 데이터의 생성 범위가 제약적이다. 이를 극복하기 위해 여러 용어체계를 통합하여 매핑작업을 수행할 수 있는 도구도 개발되었다. 그러나 의학용어체계는 각각 독특한 형태로 구성되어 있어 이들의 확실적인 통합이 어려운 문제가 있다. 따라서 본 논문에서는 기존용어체계의 형태를 유지하면서 매핑 시스템에서의 통합 및 활용이 가능한 방법을 제안한다. 제안한 방법에서는 용어체계의 위치와 형태에 대한 메타데이터를 작성함으로써 새로운 용어체계를 쉽게 시스템에 추가하여 사용할 수 있도록 하였으며, 기존 용어체계의 수정 및 구조 변경에도 유동적으로 대처할 수 있다. 또한 본 논문에서 생성한 매핑 데이터는 온톨로지에서의 트리플릿 형태로 구성함으로써 다양한 매핑 정보를 생성할 수 있다. 따라서 생성한 정보는 OWL, RDF, Excel 등의 다양한 형태로 변형하여 배포할 수 있다. 제안한 방법에 기반하여 매핑 시스템을 이용한 매핑 데이터 생성 실험을 통해 개발한 시스템의 효용성을 확인하였다.

키워드 : 표준 의학 용어체계, 매핑 시스템, 온톨로지, SNOMED-CT, ICD, NANDA, NIC, NOC

Abstract

Various standard terminologies in medical field are composed individually to different structure. Therefore, information on crosswalking between the terminologies is needed to combine and use the terminologies. Lots of mapping tools have been developed and used to create the information. However, since those tools deal with specific terminologies, the information is restrictly created. To overcome this problem, some tools have been developed, which perform mapping tasks by composing various terminologies. However, the tools also have difficulty of composing automatically the terminologies because the terminologies have different structures. Therefore, in this paper, we propose a method for composing and using the terminologies in the developed mapping system with keeping the original structure of the terminologies. In the proposed method, additional terminologies could be added on the mapping system and used by making metadata involving information on location and structure of the terminologies. And the mapping system could cope flexibly with the changes of the structure or context of the terminologies. Moreover, various types of mapping information could be defined and created in the system because mapping data are constructed as triplets in ontology. Therefore, the mapping data can be transformed and distributed in different formats such as OWL, RDF, and Excel. Finally, we confirmed the usefulness of the mapping system based on the proposed method through the experiments about creating mapping data.

Key Words : Standard medical terminology, Mapping system, Ontology, SNOMED-CT, ICD, NANDA, NIC, NOC

1. 서 론

전 세계적으로 의료 정보 및 의료 기록의 전산화를 위해

접수일자 : 2010년 12월 9일

완료일자 : 2011년 3월 30일

본 연구는 2010년도 교육과학기술부의 재원으로 한국연구재단과 (No.2010-0024256), 지식경제부 및 한국산업기술평가관리원의 지식서비스 산업원천기술개발사업 [KI10033545, KI10033576]의 지원을 받아 수행되었음.

+corresponding author

EHR (Electronic Health Record)과 같은 의료정보화 시스템을 개발하고 이용하려고 노력하고 있다 [1]. 그리고 이러한 시스템 간의 통신을 위해 HL7 (Health Level Seven) [2]과 같은 통신 표준이 제안되었다. 그러나 이들 시스템에서 저장, 활용, 전송되는 정보들은 각 의학 관련 기관이나 단체에서 주로 사용되는 용어들로써 전문분야 혹은 개인에 따라서 그 의미가 다를 수 있다. 따라서 정확한 의사소통을 위한 의학 용어의 표준화에 대한 필요성이 대두되었고, 각종 기구에서 SNOMED-CT [3], ICD-9-CM [4], CPT [5], ABC [6], NANDA [7], NIC [8], NOC [9], CCC [10], Omaha System [11], LOINC [12] 등과 같은 용어를 선정

하였다. 그리고 이들 용어 각각에게 인식 코드 (이하 ‘코드’) 를 부여한 표준의학용어체계 (이하 ‘용어체계’)가 제시되었고, 실제 많은 기관에서 이들을 사용하려 노력하고 있다.

그러나 이러한 용어체계는 실제 임상에서 직접적 활용의 목적으로 구성된 것이 아니라, 참조용어체계 혹은 용어 간 개념 관계를 정의하여 용어간의 통합을 목적으로 개발된 것이다 [13, 14]. 그러므로 이들 용어체계를 이용하여 단순히 기존의 불명확한 용어를 대체하여 정보를 표현하는 것에서 그치는 것이 아니라, 용어체계를 효율적으로 활용하기 위해서 용어체계에 기반한 지식이 구축되어야 한다. 즉, 용어체계의 효율적인 활용을 위해서는 (1)목적에 따라 용어체계의 사용 영역이 선정되어야 한다. 그리고 영역 내에서의 전문 지식을 표현하기 위해서는 (2)빈번히 사용되는 용어나 개념들을 추출하며, 또한 (3)용어와 개념 사이의 추가적인 연결 관계¹⁾ (이하 ‘매핑’)를 표현해야 한다.

많은 연구 기관에서 용어체계 간의 사전적, 의미적 매핑에 관한 표준 용어 간 연결을 연구하고 수행해 왔다. 그 성과물로, SNOMED-CT와 LOINC와의 매핑 [15], SNOMED-CT와 CPT간의 매핑 [16], 3N (NANDA, NIC, NOC)에서의 간호과정에 관한 결합 (linkage information) [17] 정보 등이 배포되었다. 그러나 이러한 매핑 정보는 넓은 범위에서 객관적으로 구성된 것으로써 각 기관의 모든 요구를 충족하지는 못한다. 즉, 다양한 의학 분야의 요구를 모두 만족하는 표준 용어 간 매핑 데이터를 일일이 생성하여 제공하기에는 무리가 있다. 예를 들어, 실제 ICD와 CPT 간의 매핑 정보는 다양한 기관에서 개발되어 제공되고 있다 [18]. 따라서 각 의학 분야 및 기관별로 그들만의 매핑 정보를 직접 생성하여 활용하는 것이 더 효과적이라 할 수 있겠다. 이런 요구에 따라 용어체계들을 통합하고 관리하며, 표준용어들 사이의 매핑 데이터를 생성하기 위한 다양한 연구가 수행되고, 그 결과로써 다양한 도구도 개발되었다 [19-24]. Lexical Grid (LexGrid) Project [19, 20]에서는 공통의 데이터 모델을 이용하여 다양한 표준용어체계와 온톨로지를 저장, 배포, 서비스하기 위한 도구를 개발하였고, Lexical Care (LexCare) Suit [21, 22]는 LexGrid의 공통모델을 기반으로 개발된 도구로써 다수의 용어체계의 검색 및 매핑 기능을 제공한다. 이들 도구에서는 공통 모델을 이용하여 다양한 표준용어의 통합, 생성, 관리, 배포 등의 다양한 작업을 수행할 수 있다. 이들 도구가 다양한 기능을 제공하는 강력한 도구임에는 틀림없다. 그러나 (1)기존 용어체계를 공통 모델로 변형하는 과정은 고도의 전문적인 지식이 필요하며, 또한 그 과정이 매우 복잡하다. 그리고 (2)용어체계는 특정 용어나 개념의 의미를 표현하기 위해 각자 독특한 구조를 가지고 있어 획일적인 형태로 변환된 용어체계에서는 이들 용어나 개념의 의미를 모두 표현하기가 어렵다. 또한 (3)기존의 용어체계를 시스템에 저장하여 사용하기 위해서는 각 용어체계를 시스템에 저장하기 위해 별도의 파서 (parser)를 개발하여야 하며, (4)기존의 용어체계의 구조가 변하거나 데이터가 업데이트 될 경우 파서를 일일이 수정하거나, 시스템에 저장된 용어체계를 수정해야 하는 번거로움이 있다. 즉, 용어체계간의 단순 매핑을 위해서는 용어체계의 편리한 검색과 그들 사이의 연결정보를 생성하는

1) 용어 및 개념 사이의 연결 정보는 ‘매핑 (mapping)’ 이나 ‘결합 (linkage)’ 이라는 용어로써 표현한다. 그리고 온톨로지에서는 흔히 속성 (property)을 이용하여 나타낸 개념(concept)사이의 연결 정보를 ‘관계 (relationship)’로 표현한다. 본 논문에서는 이러한 연결 정보를 ‘매핑’으로 표현하기로 한다.

것이 주요한 내용이다.

따라서 본 연구에서는 특정 영역에서의 전문가들이 용어체계를 이용하여 지식을 효과적으로 표현하기 위한 매핑 시스템을 설계하고 개발한다. 설계한 시스템에서는 기존 용어체계의 원형을 그대로 유지하면서 단지 매핑 데이터의 생성에 필요한 코드와, 검색에 필요한 개념의 기본 이름이나 의미만으로 일괄 검색이 가능하도록 한다. 또한 용어체계가 저장된 형태에 따라 분석 가능한 최소한의 과서를 생성하고, 이들 용어체계의 저장 형태, 위치, 코드 및 정의에 관한 속성에 관한 메타데이터를 작성함으로써, 메타데이터의 추가 및 수정만으로 시스템에서 용어체계를 검색하고 매핑 작업을 수행할 수 있도록 한다.

2. 의학 용어체계의 구조

용어체계는 그림 1에서 보는 바와 같이 종류에 따라 그 구조가 다르다. 예를 들어 그림 1(a)의 SNOMED-CT의 경우, 개념과 코드, 그리고 개념 사이의 관계를 나타내기 위한 속성들을 모두 ‘Concepts’ 테이블에 정의되어 있고, 각 개념과 유사한 표현들은 ‘Descriptions’ 테이블에 정의되어 있다. 그리고 개념들 사이의 관계는 ‘Relationships’ 테이블에 표현되어 있다. 그림 1(c)의 NANDA의 경우, 진단에 관한 상세 내용 및 코드는 ‘Diagnosis’ 테이블에 표현되어 있고, 이들의 영역 (domain)과 분류 (class)는 각각 ‘Domain’, ‘Class’ 테이블에 정의되어 있다. SNOMED-CT와 NANDA가 복잡한 구조로 표현되어 있는 반면, ICD의 경우에는 그림 1(b)와 같이 단순한 하나의 테이블로 표현되어 있다.

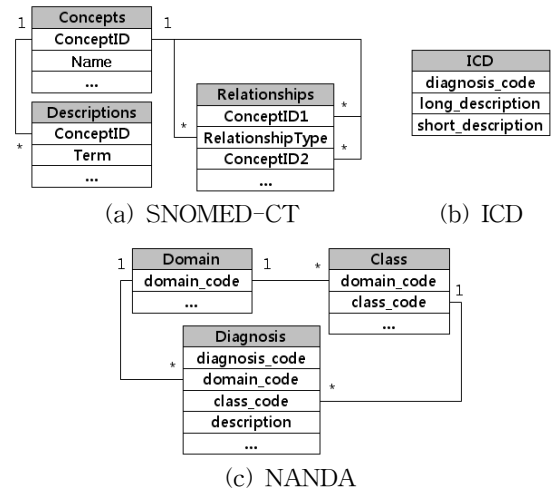


그림 1. 용어체계의 데이터 구조
Fig. 1. Data structure of terminologies

이들 용어체계를 이용하기 위해서는 표 1과 같이 코드, 대표이름이나 개념의 설명문을 기준으로 검색이 수행된다. 그리고 이들 사이의 매핑 결과는 용어체계의 코드 정보만을 필요로 한다. 따라서 시스템에서 코드와 용어체계의 이름만으로 매핑 데이터의 생성이 가능하다.

용어체계의 매핑은 전문가의 주관적 판단에 의해 수행되어야 한다. 그러나 용어체계의 정의문 자체의 애매성으로 인해 단순한 개념의 이름이나 설명문으로는 판단을 할 수 없거나, 복잡한 용어체계의 구조로 인해 용어체계 자체의

추가적인 정보가 필요할 수 있다. 이런 경우에는 용어체계에서 제공하는 정보를 최대한 활용해야 한다.

표 1. 용어체계별 코드 및 정의문 예

Table 1. Examples of code and description of each terminologies

용어체계	테이블	코드	개념 및 용어 정의
SNOMED-CT	Concept	170312000	Second cholera vaccination
CPT	-	90725	Cholera vaccine for injectable use
ICD-9-CM	Diagnosis	E871.3	Foreign object left in body during injection or vaccination
	Procedure	991.7	Injection of insulin
ABC	-	NAASV	Insulin injection assessment, Common nursing interventions, Nursing
CCC	Diagnosis	60.2	labor risk
	Intervention	23.1	insulin injection
Omaha System	Problem	06	Communication with community resources
	Scheme	(34)	coordination among providers
NANDA	Diagnosis	00076	Readiness for enhanced community coping
NIC	Intervention	8500	community health development
NOC	Outcome	2800	community health status:immunity

3. 매핑 시스템의 설계

3.1 용어체계의 활용을 위한 매핑 시스템의 설계

다양한 형태로 분산되어 저장되어 있는 용어체계들을 시스템의 DB에 저장하여 활용할 경우, 네트워크 상황이나 전송속도에 영향을 받지 않고 자원을 원하는 대로 사용할 수 있는 장점이 있다. 그러나 시스템에 새로운 용어체계를 추가하여 작업을 하려 하거나, 용어체계 자체가 변형되는 경우에는 시스템 구조를 변경하거나, 새로운 용어체계를 업데이트 하여야 하는 번거로움이 있다.

여러 곳에 흩어져 있는 용어체계를 통합하여 활용하기 위한 방법으로는 일반적으로, (1)용어체계의 구조를 시스템에 맞게 변형하여 시스템에 저장하여 사용하는 방법, (2)용

어체계의 원형 변형하지 않고 시스템에 저장하여 활용하는 방법, 그리고 (3)용어체계 자체의 변형이 불가하거나, 용어체계에 제한적으로 접근할 수 있는 경우에는 제공되는 API를 이용하는 방법이 있다. (1)의 경우에는 용어체계의 구조를 원하는 대로 수정하여 시스템에 저장함으로써, 용어체계의 활용 면에서 최적의 성능을 얻을 수 있는 장점이 있다. 그러나 용어체계의 구조를 임의로 수정할 경우, 그 원형이 손상되거나 왜곡 될 수 있는 단점이 있다. (2)의 경우에는 용어체계의 원형을 유지함으로써 용어체계의 의미를 그대로 활용할 수 있는 장점이 있다. 그러나 이 경우는 각 용어체계를 시스템에 저장하여 운영해야 하므로 용어체계 원본의 업데이트 발생에 따라 시스템에 저장되어 있는 용어체계를 관리해 주어야 한다. (3)의 경우에는 용어체계의 기본 정보와 위치만을 이용하여 용어체계를 네트워크를 통해 활용할 수 있기 때문에, 단순히 용어체계의 참조 정보만을 수정함으로써 자원의 업데이트나 구조적 변형이 발생하여도 유동적으로 대처할 수 있다. 그러나 네트워크의 문제가 발생할 경우 용어체계를 이용할 수 없고, 또한 용어체계 제공자의 정책에 따라야 하는 단점이 있다.

그림 2는 다양한 형태로 저장되어 있는 분산된 환경의 용어체계를 활용하기 위한 매핑 시스템의 구조를 보인다. 즉, 용어체계는 매핑 시스템과 같은 로컬 환경에 파일이나 데이터베이스 형태로 저장되어 있을 수도 있고, 인터넷에 공개된 파일이나 외부 시스템에 설정되어 있는 데이터베이스에 저장되어 있을 수도 있다. 또한 외부에서의 데이터베이스 접속은 API를 통해 가능할 수도 있다. 따라서 매핑 시스템은 각 용어체계의 상태에 관한 정보를 메타데이터로 표현하고, 시스템은 선택기에서 메타데이터의 정보만으로 용어체계를 어떠한 형태로 불러올지를 결정하게 된다. 즉, 용어체계가 공개된 데이터베이스일 경우에는 DAO (Database Access Object)로, 비공개된 데이터베이스일 경우에는 API (Application Programming Interface)로, 그리고 파일 형태로 공개된 경우에는 FAO (File Access Object)로 전달한다. 연결된 용어체계는 'I/O Manager'에서 코드와 설명문을 추출하여 'Data Manager' 모듈로 전달한다. 사용자는 이 정보를 기반으로 매핑 작업을 수행한다. 사용자가 용어체계의 검색을 요청하면, 검색 엔진은 메타데이터에서 설정된 방법을 통해 각 용어체계의 검색을 수행한다. 예를 들어, DAO를 통한 접속일 경우에는 검색 질의를 생성하고, API를 통

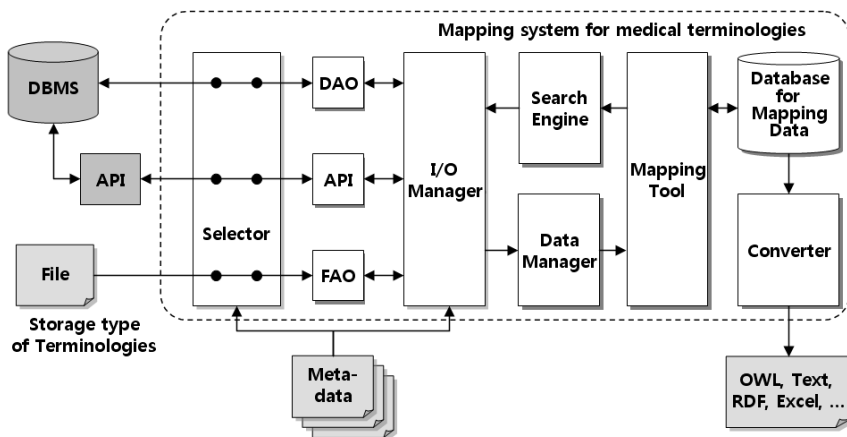


그림 2. 분산된 용어체계의 활용을 위한 매핑 시스템의 구조

Fig. 2. Architecture of a mapping system for using distributed medical terminologies

한 접속일 경우에는 API에서 제공하는 검색방법을 이용하며, FAO 에서는 시스템에서 별도의 검색 모듈에서 검색을 수행한다. 단, 용어체계와의 접속 방법이 API의 경우에는 제공되는 API로부터 코드 및 설명문을 추출하거나, 검색을 수행하기 위한 별도의 변환 모듈을 생성하여야 한다. 따라서 이 모듈의 위치를 메타데이터에 표시한다.

3.2 용어체계의 연결을 위한 메타 데이터의 설계

분산되어 저장되어 있는 다양한 형태의 용어체계를 매핑 시스템에 연결하기 위해 각 용어체계의 특성에 따른 메타데이터를 작성한다. 이 경우, 시스템에서는 용어체계별로 별도의 모듈을 제작할 필요 없이 단순히 메타데이터만 추가하거나 수정함으로써 용어체계를 이용할 수 있다. 용어체계의 연결을 위해 필요한 메타데이터는 표 2와 같은 정보들로 구성되어 있다. 그리고 이들 정보를 이용하여 XML 형태의 메타데이터를 구성하기 위한 XML Schema를 그림 3에서 보인다.

표 2. 메타데이터 구성 요소
Table 2. Elements of metadata

요소 이름	정의
TerminologyCode	용어체계 인식 코드
TerminologyName	용어체계 이름
TerminologyType	용어체계 종류 (DB, File, API)
TerminologyLocation	용어체계 위치 (DB, File, API 접속 IP)
Type	DB/File/API 종류
Name	DB/File/API 이름
ID	DB/API 접속 ID
Password	DB/API 접속 비밀번호
CodeFieldName	DB/File 에서의 코드 필드명
DescFieldName	DB/File 에서의 설명문 필드명
ApiIoModule	API I/O 변환 모듈
ViewerType	용어체계 상세정보 뷰어 종류
ViewerLocation	용어체계 상세정보 뷰어 위치
...	...

3.3 매핑 데이터 생성 개념

매핑 시스템에서 매핑에 필요한 최소한의 정보는 매핑 데이터를 생성할 코드와 코드의 개념을 인식할 정의문이다. 코드의 개념 글은 의미가 변하지 않는 범위에서 일부 변형의 가능성이 있으나, 코드는 그 개념을 대표하는 것이므로 변형될 수 없다. 경우에 따라서는 사용하지 않는 개념으로써 삭제되거나, 새로운 개념이 생성될 수도 있다. 연결 정보를 나타낼 때, 이중 용어체계의 요소간의 연결정보를 나타내는 매핑(mapping)과, 유사한 목적으로 생성된 용어체계들에서 요소간의 연결 정보를 나타내는 결합(linkage)으로 구분할 수 있다. 매핑의 대표적인 예는 SNOMED-CT와 ICD-9-CM에서의 동일 개념을 연결한 것이며²⁾, 결합의 대표적인 예는 3N 사이의 간호과정 및 절차와 같은 연결³⁾을 들 수 있다. 그러나 여기서는 이들 두 가지 개념을 통틀어 매핑이라 하고, 매핑 작업의 결과를 매핑 데이터라고 한다. 표 3은 이들 두 가지 매핑 데이터의 예를 보인다.

- 2) 의학 분야의 참조 용어체계인 SNOMED-CT와 국제질병분류 코드인 ICD-9-CM 간에 개념은 동일하나 그 표현이 다른 것들을 연결해 놓은 것을 'cross mapping table'이라고도 한다.
- 3) 간호 과정은 간호진단을 포함하는 NANDA, 간호결과를 포함하는 NOC, 그리고 간호중재를 포함하는 NIC에서의 코드들의 결합으로 하나의 간호 과정을 표현한다.

```
<schema xmlns="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://med.knu.ac.kr/mapping"
xmlns:med="http://med.knu.ac.kr/mapping">
<element name="Terminologies" type="med:Terminology"/>
<element name="Terminology">
<complexType>
<sequence>
<element ref="med:TerminologyType"/>
</sequence>
</complexType>
<attribute name="code" type="string"/>
<attribute name="name" type="string"/>
</element>
<element name="TerminologyType">
<complexType>
<choice>
<element ref="med:Database"/>
<element ref="med:API"/>
<element ref="med:File"/>
</choice>
</complexType>
</element>
<element name="Database">
<complexType>
<attribute name="type" type="med:DBType"/>
<attribute name="name" type="string"/>
<attribute name="location" type="string"/>
<attribute name="id" type="string"/>
<attribute name="password" type="string"/>
<attribute name="code_field_name" type="string"/>
<attribute name="desc_field_name" type="string"/>
</complexType>
</element>
<element name="API">
<complexType>
<attribute name="type" type="med:APIType"/>
<attribute name="id" type="string"/>
<attribute name="password" type="string"/>
<attribute name="api_io_module" type="string"/>
</complexType>
</element>
<element name="File">
<complexType>
<attribute name="type" type="med:APIType"/>
<attribute name="name" type="string"/>
<attribute name="location" type="string"/>
<attribute name="code_field_name" type="string"/>
<attribute name="desc_field_name" type="string"/>
<attribute name="field_line" type="nonNegativeInteger"/>
<attribute name="data_start_line" type="nonNegativeInteger"/>
</complexType>
</element>
<simpleType name="APIType">
<restriction base="string">
<enumeration value="java"/>
</restriction>
</simpleType>
<simpleType name="FileType">
<restriction base="string">
<enumeration value="text"/>
<enumeration value="rdf"/>
<enumeration value="owl"/>
</restriction>
</simpleType>
<simpleType name="DBType">
<restriction base="string">
<enumeration value="mysql"/>
<enumeration value="mssql"/>
<enumeration value="oracle"/>
</restriction>
</simpleType>
</schema>
```

그림 3. 메타 데이터 생성을 위한 XML Schema
Fig. 3. XML Schema for description of metadata

표 3. 매핑 데이터의 예

Table 3. Examples of mapping data

종류	매핑	
의미	equivalent	
데이터	SNOMED-CT 13313007 (Mild bipolar disorder)	ICD-9-CM 296.7 (Bipolar I current NOS)
	13322008 (Postseizure headache)	339.89 (Headache syndrome NEC)

종류	결합		
의미	process		
데이터	NANDA 00092 (Activity intolerance)	NOC 0005 (activity tolerance)	NIC 4310 (activity therapy)
	00011 (Constipation)	0501 (bowel elimination)	0430 (bowel management)

* (·)는 각 코드의 설명문을 나타낸다.

표 4. 매핑 데이터의 Triplet 표현 예

Table 4. Example of triplet expression of mapping data

Domain terminology code	relationship type	Range terminology code
snomed:13313007	equalTo	icd9:296.7
snomed:13322008	equalTo	icd9:339.89
...
nanda:00092	linkedTo	nic:4310
nic:4310	linkedTo	noc:0005
nanda:00011	linkedTo	nic:0501
nic:0501	linkedTo	noc:0430
...

* 각 code 앞의 snomed, icd9, nic, nanda, noc은 용어체계를 구분하기 위해 임의로 부여한 용어체계 인식 코드이다. 이 인식 코드는 OID [25] 코드로 대체할 수 있다.

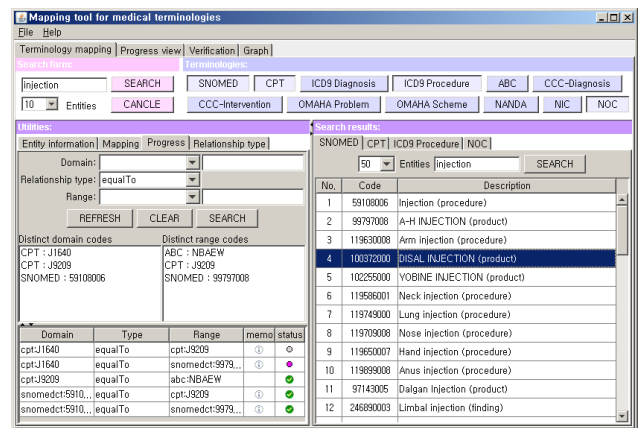
매핑 데이터의 정의가 명확히 구분될 경우에는 표 3과 같이 단순히 두 용어체계의 코드만을 나열함으로써 그 의미를 전달할 수 있다. 그러나 다양한 용어체계를 하나의 시스템에서 다루는 경우에는 매핑 작업의 목표가 사용자마다 다를 수 있다. 그러므로 두 용어체계의 코드를 단순히 나열하는 것으로는 매핑 데이터의 의미가 모호해진다. 이러한 문제를 해결하기 위해, 본 논문에서는 두 가지 매핑 개념(매핑 및 결합) 이외에도 다양한 매핑 개념을 구분하기 위해서 트리플릿⁴⁾(triplet)의 형태로 매핑 데이터를 표현한다. 예를 들어, 'A'와 'a'가 각각 서로 다른 용어체계의 개념이고, 이들 두 개념의 의미가 동일한 경우, 'equalTo'라는 'relationship type'을 이용하여 (A-equalTo-a)와 같이 매핑 데이터를 표현한다. 따라서 다수의 작업자가 동시에 같은 분야의 매핑 작업을 수행할 때, 'relationship type'만으로 같은 그룹에 속한 매핑 데이터라는 것을 알 수 있다. 또한, 매핑 데이터를 'relationship type'을 이용하여 표현할 경우, 매핑 데이터

4) 온톨로지에서의 트리플릿(triplet)은 속성(property)을 이용하여 개념(concept)과 개념 사이의 관계(relationship)를 표현한 단위 지식을 의미한다. 즉, Subject, Predicate, Object 사이의 관계로써 'S-P-O'나 '(P(S, O))'로 표현한다.

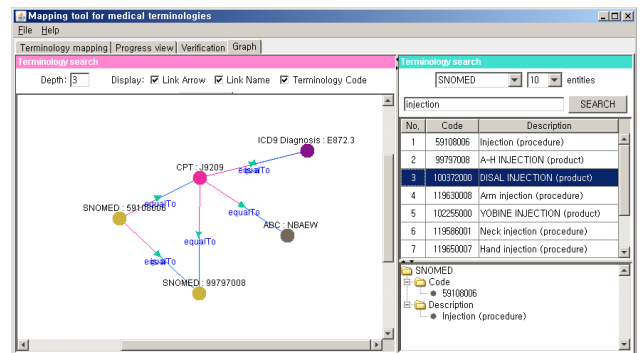
의 형태가 온톨로지에서의 트리플릿과 동일하다. 트리플릿으로 표현된 매핑 데이터는 엑셀(excel), 테이블 형태의 텍스트 파일, RDF(Resource Description Framework), OWL(Web Ontology Language) 등의 다양한 형태로 변형되어 배포된다. 표 4는 표 3의 매핑 데이터를 'relationship type'을 이용하여 트리플릿으로 표현한 예이다.

4. 개발 및 실험

본문에서 제안한 방법에 기반하여 그림 4와 같이 매핑 시스템을 개발하였다. 그림 4(a)는 다수의 용어체계를 선택적으로 검색하고, 트리플릿 형태의 매핑 데이터를 생성하기 위한 사용자 화면을 보인다. 그리고 그림 4(b)는 특정 용어체계의 개념을 중심으로 관련된 매핑 데이터를 2차원 그래프로 표현함으로써 사용자가 매핑 결과를 쉽게 확인할 수 있도록 하기 위한 가시화 도구(visualization tool)를 보인다.



(a) 매핑 시스템의 사용자 화면



(b) 매핑 결과의 2차원 가시화 도구

그림 4. 개발한 매핑 시스템

Fig. 4. Developed mapping system

그림 5는 다음 3가지 상황과 같은 용어체계의 저장 형식에 따라 XML로 구현된 메타데이터를 보인다. 즉, (1)다른 시스템의 DB에 저장되어 있는 용어체계로써 직접 액세스 가능한 경우, (2)다른 시스템의 DB에 저장되어 있는 용어체계를 공개된 API를 통해 액세스 가능한 경우, (3)다른 시스템에 저장되어 있는 테이블 형태로 구성된 파일을 다운로드 받는 경우이다.

```

<?xml version="1.0" encoding="euc-kr"?>
<Terminologies
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="Metadata.xsd"
  xmlns:med="http://medinfo.knu.ac.kr/mapping">
  <Terminology code="snomed" name="SNOMED-CT">
    <TerminologyType>
      <Database
        type="mssql"
        name="Snomed"
        location="155.230.149.144"
        id="****"
        password="****"
        code_field_name="conceptid"
        desc_field_name="fullyspecifiedname"/>
      </TerminologyType>
    </Terminology>
    <Terminology code="nanda" name="NANDA">
      <TerminologyType>
        <API
          type="java"
          id="****"
          password="****"
          api_io_module="com.mipth.mapping.api.NandaApi"/>
        </TerminologyType>
      </Terminology>
    <Terminology code="icd" name="ICD9_Diagnosis">
      <TerminologyType>
        <File
          type="text"
          name="CMS_DX.txt"
          location="155.230.149.144/terminologies"
          code_field_name="DIAGNOSIS CODE"
          desc_field_name="LONG DESCRIPTION"
          field_line="1"
          data_start_line="2"/>
        </TerminologyType>
      </Terminology>
    </Terminologies>
  
```

그림 5. 메타 데이터 생성 예
Fig. 5. Examples of metadata description

개발한 매핑 시스템의 효용성을 확인하기 위해 (1)SNOMED-CT와 ICD-9-CM 간의 'equalTo' 속성에 대한 매핑과 (2) 3N에서의 간호과정에 대한 결합정보의 생성 작업을 수행하였다. 이 두 가지 상황에서 단순히 파일 형태의 용어체계를 참조한 매핑 작업과 매핑 시스템을 이용한 매핑 작업의 시간을 비교하였다. (1)의 실험을 위해 [3]에서 제공한 SNOMED-CT와 ICD-9-CM 간의 매핑 데이터 중 각 용어체계의 코드가 생략된 50개의 정의문만 선택하였고, 정의문에서 3개의 키워드를 추출하였다. 그리고 (2)의 실험에서는 [17]에서 제공하는 50개의 NANDA-NOC-NIC 결합 정보를 이용하였다. 이 역시 각 용어체계의 코드가 없이 오로지 개념명만으로 검색을 수행하도록 하였다. 표 5와 표 6은 총 5회에 걸쳐 10개 용어의 매핑 및 결합에 대한 (1)과 (2)의 매핑데이터 생성에 소요된 시간을 나타낸 것이다.

실험 (1)에서는, 파일로부터 검색을 수행하기 위해 구 (phrase) 형태의 연속된 3개 단어를 이용할 수밖에 없었다. 따라서 검색어의 선택에 제한이 있어 일치하는 내용을 일일이 눈으로 확인하는데 시간이 많이 소요되었다. 반면에 매핑 시스템에서는 중요한 핵심 단어들만 선택하여 검색할 수 있고, 또한 각 용어체계별로 별도로 검색해야 하는 번거로움이 없이 한 번에 검색이 가능하였다. 또한 검색 리스트를 한눈에 볼 수 있어서 일치하는 개념을 빨리 찾을 수 있었다. 따라서 표 5에서 보는 바와 같이, 매핑 시스템을 이용했을 때 매핑 시간이 크게 단축됨을 알 수 있었다. 실험 (2)에서는, 용어체계의 개념 수가 많지 않고, 또한 개념명의 길이가

길지 않아 파일에서의 검색이 상대적으로 쉬웠다. 그러나 파일에서 코드를 복사하고 붙여 넣는데 시간이 많이 소요된 반면, 매핑시스템에서는 코드에 신경을 쓰지 않고 한 번에 여러 개의 매핑이 가능하였다. 따라서 매핑 시스템을 이용하였을 때, 매핑 과정에서의 편리함으로 인해 약간의 시간을 단축할 수 있었다.

표 5. SNOMED-CT와 ICD-9-CM간의 매핑 소요 시간
Table 5. Time spent on creating mapping data between SNOMED-CT and ICD-9-CM

실험차수	소요시간 (초)	
	파일 기반	매핑 시스템 기반
1차	915	446
2차	532	359
3차	570	368
4차	596	401
5차	449	330

표 6. 3N간의 매핑 소요 시간
Table 6. Time spent on creating mapping data among 3N

실험차수	소요시간 (초)	
	파일 기반	매핑 시스템 기반
1차	308	295
2차	134	202
3차	298	262
4차	342	320
5차	246	215

5. 결 론

본 논문에서는 기존용어체계의 형태를 유지하면서 매핑 시스템에서의 통합 및 활용이 가능한 방법을 제안하였다. 제안한 방법에서는 용어체계의 위치와 형태에 대한 메타데이터를 작성함으로써 새로운 용어체계를 쉽게 시스템에 추가하여 사용할 수 있도록 하였으며, 기존 용어체계의 수정 및 구조 변경에도 유동적으로 대처할 수 있는 장점이 있다. 또한 본 논문에서 생성한 매핑 데이터는 온톨로지에서의 트리플릿 형태로 표현함으로써 매핑 이외의 다양한 연결 정보를 생성할 수 있고, 생성한 정보는 OWL, RDF, Excel 등의 다양한 형태로 변형하여 배포할 수 있다. 본 시스템은 다수의 전문가가 정보를 공유하여 구축하는 의학용어체계의 개념 간 연결 정보의 질을 높일 수 있는 환경을 제공하며, 다양한 의학 분야에서 특정 영역의 지식을 통합하고 관리하는데 유용한 도구로 활용될 수 있을 것으로 기대된다.

참 고 문 헌

- [1] 김용욱, *Essential Elements of EHR System*, 군자출판사, 2006.
- [2] "HL7", Available: <http://www.hl7.org>, [Accessed: October 2, 2010]
- [3] "SNOMED CT", Available: <http://www.ihtsdo.org>, [Accessed: November 5, 2010]
- [4] "ICD-9-CM", Available: <http://icd9cm.chrisendres.com>, [Accessed: November 5, 2010]

- [5] "CPT", Available: <http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.shtml>, [Accessed: November 5, 2010]
- [6] "ABC", Available: http://www.abccodes.com/ali/products_services/pro_description.asp, [Accessed: November 5, 2010]
- [7] "NANDA", Available: <http://www.nanda.org>, [Accessed: November 5, 2010]
- [8] "NIC", Available: http://www.nursing.uiowa.edu/excellence/nursing_knowledge/clinical_effectiveness/nic.htm, [Accessed: November 5, 2010]
- [9] "NOC", Available: http://www.nursing.uiowa.edu/excellence/nursing_knowledge/clinical_effectiveness/noc.htm, [Accessed: November 5, 2010]
- [10] "CCC", Available: www.sabacare.com, [Accessed: November 5, 2010]
- [11] "Omaha System", Available: <http://www.omahasystem.org>, [Accessed: November 5, 2010]
- [12] "LOINC", <http://loinc.org>, [Accessed: November 5, 2010]
- [13] G. Wade and S.T. Rosenbloom, "Experiences mapping a legacy interface terminology to SNOMED CT," *BMC Med Inform Decis Mak*, vol. 8, no. 1, p. S3, 2008.
- [14] A.L. Osornio, D. Luna, M.L. Gambarte, A. Gomez, G. Reynoso, and F.G. de Quiros, "Creation of a local interface terminology to SNOMED CT," *Studies in Health Technology and Informatics*, vol. 129, pp. 765-769, 2007.
- [15] National Library of Medicine (US), "LOINC to CPT Mapping," Available: http://www.nlm.nih.gov/research/umls/mapping_projects/loinc_to_cpt_map.html, 2006, [Accessed: November 10, 2010]
- [16] S.Y. Kim, H.H. Kim, I.K. Lee, H.S. Kim, and H. Cho, "Proposed Algorithm with Standard Terminologies(SNOMED and CPT) for Automated Generation of Medical Bill for Laboratory Tests," *Healthcare Informatics Research*, vol. 16, no. 3, pp. 185-190, 2010.
- [17] M. Johnson, G. Bulechek, H. Butcher, J.M. Dochterman, M. Maas, S. Moorhead, and E. Swanson, *NANDA, NOC, and NIC Linkages*, Mosby, 2006.
- [18] Wasserman Medical Publishers, "Cross Coder," Available: <http://www.crosscoder.com>, [Accessed: December 3, 2010]
- [19] J. Pathak, H.R. Solbrig, J.D. Buntrock, T.M. Johnson, and C.G. Chute, "LexGrid: A Framework for Representing, Storing, and Querying Biomedical Terminologies from Simple to Sublime," *Journal of the American Medical Informatics Association*, vol. 16, no. 3, pp. 305-315, 2009.
- [20] Mayo Clinic, "LexGrid", Available: <http://www.lexgrid.org/LexGrid>, [Accessed: November 26, 2010]
- [21] J.-A. Yang, S.-J. Nam, J.-H. Choi, S.-J. Koh, and H.-G. Kim, "LexCare WS-A Web Services Based Platform for Medical Terminologies management," *2010 대한의료정보학회 추계학술대회 포스터 초록집*, p. 5, 2010.
- [22] Center for Interoperable EHR, "LexCare Suite," <http://ehrkorea.org/experience/lexcare.aspx>, [Accessed: November 26, 2010]
- [23] "CliniClue Xplore", Available: http://www.cliniclue.com/cliniclue_xplore, [Accessed: November 26, 2010]
- [24] I.K. Lee, H.S. Kim, S.J. Hong, and H. Cho, "Development of a System for Construction of Linking Information between Concepts in Medical Terminologies," *2010 대한의료정보학회 추계학술대회 포스터 초록집*, p. 6, 2010.
- [25] "Object Identifier Repository", Available: <http://www.oid-info.com>, [Accessed: November 26, 2010]

저 자 소 개



이인근(In Keun Lee)

2001년: 영남대학교 재료금속공학(학사)
2004년: 영남대학교 대학원 전기공학(석사)
2010년: 영남대학교 대학원 전기공학(박사)
2010년~현재: 경북대학교 의료정보원천
기술연구소 수석연구원

관심분야 : 지능시스템, 온톨로지, 의료정보표준, EHR
E-mail : inkeunlee@gmail.com



김화선(Hwa Sun Kim)

2003년: 인제대학교 컴퓨터공학(석사)
2007년: 경북대학교 의료정보학(박사)
2009년~2011년: 경북대학교 의료정보학과
연구교수
2011년~현재: 대구한의대학교 IT의료산
업학과 교수

관심분야 : XML기반 병원정보시스템, 객체지향방법론 기
반 CDA 및 RMI 개발, 임상표준용어코드
E-mail : pulala@paran.com



조 훈(Hune Cho)

1980년: 서울대 수학과(학사)
2004년: 미국 남캘리포니아대학 전산학
(석사)
2010년: 미국 유타주립대학 의료정보학
(박사)
1994년~1999년: 아주대학교 의과대학 조교수
1999년~현재: 경북대학교 의료정보학과 교수

관심분야 : 병원정보시스템, 온톨로지, 적정보상체계, HL7
E-mail : hunecho@mail.knu.ac.kr