

ISO 25964-2를 기반으로 한 시소러스와 온톨로지 상호운용성 연구*

- 행정부처명을 중심으로 -

Thesaurus and Ontology Based on ISO 25964-2: With a Focus on Administrative Departments Names

이 혜 원 (Hyewon Lee)**

목 차

- | | |
|-------------------------|--------------------------|
| 1. 서론 | 4. 매핑 유형을 고려한 온톨로지 모형 구축 |
| 2. 이론적 배경 | 및 활용 |
| 3. 시소러스 상호운용성을 위한 관계 확장 | 5. 결론 |

<초 록>

2013년에 ISO 25964-2가 소개되었지만 지금까지 이에 대한 심층적인 분석이나 구체적인 활용 방안을 제시한 연구는 아주 미미하다. 본 연구에서는 ISO 25964-2에서 강조하고 있는 시소러스 상호운용성을 분석하고, 그 활용가능성을 살펴보았다. 구체적으로, 본 연구에서는 ISO 25964-2에서 제안한 개념 통합 방식인 매핑 유형을 분석하였다. 또한 ISO 25964-2의 매핑 유형을 확인하기 위해 온톨로지와의 연계 방안을 모색하고, 마지막으로 시소러스와 온톨로지의 상호운용성을 확인하기 위해 우리나라 행정부처의 기관명을 대상으로 한 온톨로지 모형을 시범적으로 설계하고 구축하였다.

주제어: ISO 25964, 시소러스, 온톨로지, 기록시소러스, 시소러스 상호운용성, 행정부처 시소러스

<ABSTRACT>

Although ISO 25964-2 was introduced in 2013, existing research about it that can be used for analysis and basis for its application is scarce. This study analyzed the interoperability of vocabularies emphasized in ISO 25964-2. First, this study examined the mapping types of the concept sharing method. In addition, based on the mapping types, this study sought for and examined the interoperability method between a thesaurus and ontology. Lastly, this study has suggested a thesaurus and ontology interoperability model and designed and constructed an ontology model for administrative department names in Korea to verify the possible of the model.

Keywords: ISO 25964, thesaurus, ontology, archival thesaurus, thesaurus interoperability, administrative department thesaurus

* 이 논문은 2015년도 서울여자대학교 연구년 연구비의 지원을 받았음.

** 서울여자대학교 문헌정보학과 부교수(hwlee@swu.ac.kr)

■ 접수일: 2017년 7월 25일 ■ 최초심사일: 2017년 8월 8일 ■ 게재확정일: 2017년 8월 19일

■ 한국기록관리학회지 17(3), 71-96, 2017. <<http://dx.doi.org/10.14404/JKSARM.2017.17.3.071>>

1. 서론

W3C에서 제안한 시맨틱 웹을 지원하기 위해 웹 포털뿐만 아니라 도서관, 박물관, 기록관 등과 같은 데이터 관리 기관에서도 발전된 정보 기술을 바탕으로 데이터 수집, 활용 및 공유 방안을 모색하고 있다. 시맨틱 웹은 말 그대로 의미를 전달하는 정보 환경을 말하는데 이를 위해 웹을 구성하는 데이터를 구조화하고 내포된 의미를 일정한 규칙 안에서 전달해야 한다. 의미에 대한 일정한 규칙을 정보학에서는 시소러스라는 용어로 칭하고 있다. 실제 시소러스는 이용자의 정보검색 행위를 지원하기 위한 도구로 개발되었으나, 그 외연을 넓혀 데이터 자체에 대한 풍부하고 정확한 표현을 제공한다.

Brownson은 1957년에 시소러스라는 용어를 처음 사용하여, 개념의 모호성을 해소하고 개념 간의 관계 표현을 시도하였다(Brownson, 1957). 시소러스는 개념에 집중하여, 개념에 대한 다양한 표현을 구별하고 개념 간의 관계 유형을 정의한 것이다. 시소러스가 자동화된 정보검색 시스템에 매우 중요한 기능을 제공하는 것은 틀림없으나, 정보기술의 발전과 이용자 요구의 복잡성을 충족시키기에는 많은 문제점들이 드러나고 있다. 이러한 문제점을 해결하기 위해 국제표준기관 ISO에서는 2011년과 2013년에 시소러스 제작 및 개발 지침을 개정하였다. ISO 25964-1(2011)에서는 시소러스의 형식과 데이터 교환 프로토콜을 포함한 개발과 관리를 포함하며, 단일언어와 다국어 시소러스에 관한 지침을 모두 다루고 있다. ISO 25964-2(2013)에서는 분류체계나 이름전거파일, 온톨

로지와 같은 유사한 다른 제어어휘 표준과 시소러스의 상호운용성에 관한 지침을 제공한다(ISO, 2011; 박지영, 윤소영, 이혜원, 2017, p. 121 재인용).

ISO 25964-2의 명칭은 'Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies'이다. ISO 25964-2의 키워드는 '시소러스와 다른 어휘집간의 상호운용성'이다. ISO 25964-2는 단순하지만 강력하게 개념을 표현하는 시소러스를 기준으로 분류체계, 기록관리 분류체계, 택소노미, 주제표목표, 온톨로지, 용어집, 이름전거리스트, 동의어리스트 등과 같은 다양한 지식시스템인 어휘집 등도 포함한다. 또한 ISO 25964-2에서는 ISO 25964-1에서 설명하는 기본적인 시소러스 관계를 확장시켜, 2개 이상의 어휘집을 통합하기 위한 매핑 유형 방식을 소개하였다.

2013년에 ISO 25964-2가 소개되었지만 지금까지 이에 대해 심층적인 분석이나 구체적인 활용 방안을 제시한 연구는 아주 미미하다. 본 연구에서는 ISO 25964-2에서 강조하고 있는 시소러스 상호운용성을 분석하고, 그 활용가능성을 제시하고자 한다. 좀 더 구체적으로, 본 연구에서는 ISO 25964-2에서 강조하는 대등관계를 심층적으로 분석하고, 이에 초점을 맞춰 시소러스와 온톨로지 연계 방안을 모색하고자 한다. 또한 본 연구에서는 다양한 대등관계를 적용할 수 있는 우리나라 행정부처의 명칭을 대상으로 온톨로지 모형을 시범적으로 설계 및 구축하고자 하며, 이는 시소러스에서 제시한 개념이 확장되어 온톨로지라는 지식시스템으로 연계될 수 있음을 확인하는 것이다.

2. 이론적 배경

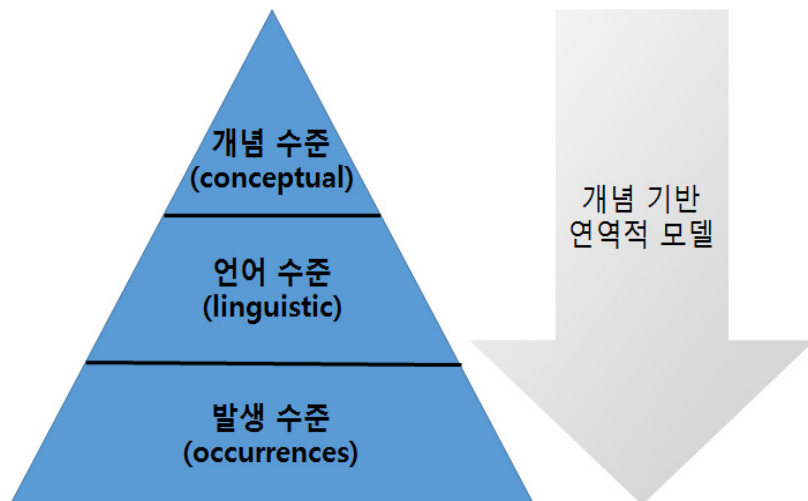
2.1 개념 모형

시소러스의 고도화 작업은 전자적인 정보 환경의 발전에 따라 진행되었다. 그 발전에는 복잡한 개념을 체계적으로 설명하고 다양한 언어 간의 변환이 자유로우며 이중의 시스템에서 관리되는 용어들을 통합하기 위한 노력이 필요하였다. 1990년대 후반부터 시소러스 연구자 및 실무자들은 시소러스의 개념 모형 체계를 갖추고 전자적인 환경에도 잘 적용할 수 있는 개념 공유 시스템을 제안하였다.

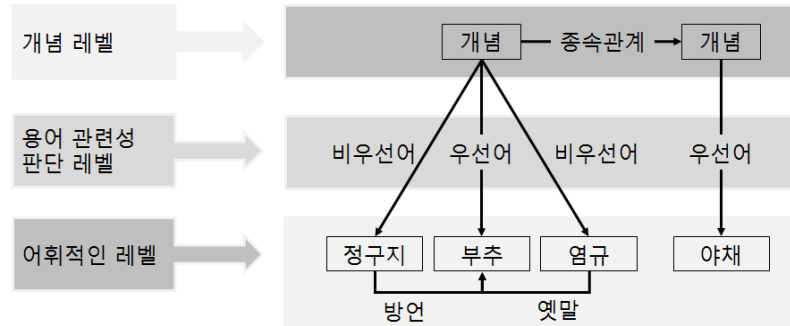
Järvelin et al.의 연구(1996)에서는 질의어 확장을 위한 연역적 모델을 제안하였으며, <그림 1>에서 제시된 바와 같이 세 단계로 표현하였다(Pastor-Sanchez, Mendez, & Rodríguez-Muñoz, 2009, p. 2 재인용). 첫 번째의 개념 (conceptual) 수준에서는 개념과 개념간의 관

계를 표시하였고, 두 번째의 언어(linguistic) 수준에서는 개념을 자연언어로 표현하였으며, 마지막 수준인 발생(occurrences)은 개념의 다양한 표현들 중 실제 상황에 맞는 표현만을 관리하였다.

Pastor-Sanchez, Mendez, & Rodríguez-Muñoz의 연구(2009, p. 4)에서는 개념 기반 시소러스의 세 단계를 제안하였다(<그림 2> 참고). 전통적인 시소러스에서 제공하지 않은 어휘적인 관계를 제공하는 것이 정확한 지식 전달의 밑거름이 될 수 있음을 확인하였다. 또한 시소러스 구축 시 개념에 대한 설명 그리고 개념간의 관계를 표현하는 것도 중요하지만, 실제 사용되는 어휘에 대한 설명이 추가되어야 함을 알 수 있었다. 이는 온톨로지 구축 시 인스턴스 입력을 통해 인스턴스간의 관계를 정보 시스템 안에 포함시키는 방식과 동일하다고 할 수 있다.



<그림 1> Järvelin et al. 연구(1996)에서 제안한 시소러스 모형



〈그림 2〉 Pastor-Sanchez, Mendez, & Rodríguez-Muñoz의 연구(2009)에서 제안된 시소러스 모형

출처: Pastor-Sanchez, Mendez, & Rodríguez-Muñoz의 연구(2009, p. 4) 〈그림 2〉를 재구성함

2.2 선행연구

시소러스 지침에 대한 국제표준은 ISO에서 2011년과 2013년에 1부와 2부로 나누어서 제안되었다. 2013년에 선언된 ISO 25964-2는 시소러스 상호운용성을 강조하며, 분류체계, 이름 전거리스트, 온톨로지 등과 같은 다른 어휘체계들과의 연계 방안을 제안하였다. ISO 25964-2에 대한 국내 연구가 활발하게 진행되지 않아서, 본 연구에서는 ISO 25964에 대한 전반적인 연구를 중심으로 살펴보았다.

박옥남의 연구(2011)와 김성원과 김정우의 연구(2011)는 ISO 25964-1이 발표되는 시점에서 시소러스 표준의 변화와 그에 대한 시사점을 제안하였다. 먼저 박옥남의 연구(2011)에서는 시소러스 제작과 개발을 위한 지침 변화의 의미를 살피고, 그 내용을 분석하였다. 해당 연구에서는 ISO 25964를 소개하고 ISO 25964의 범용화를 위하여 향상되어야 할 점을 제시하였다. 이를 위해 시소러스 표준 ISO 2788: 1986, ANSI/NISO Z39.19-2005와의 비교분석이 이루어졌다(박옥남, 2011).

김성원과 김정우의 연구(2011)는 ISO 25964의 개정 현황을 살펴보고 변화된 내용과 그 시사점을 도출하였다. 그러한 고찰 결과를 기반으로 오늘날의 새로운 정보환경 아래에서 시소러스가 제공할 수 있는 새로운 기능과 그 기능을 수행하기 위해 필요한 요소들을 제시하였다.

ISO 25964-2 발표 이후의 대표적인 연구로는 박지영, 윤소영, 이혜원의 연구(2017)를 들 수 있는데, 그들은 최근의 정보환경 변화와 시소러스 구축 및 활용 시스템의 발전에 맞춰 개정된 시소러스 구축 국제표준인 ISO 25964를 반영한 기록시소러스 구축 지침을 제안하였다.

시소러스 구축 및 개발에 관한 최근 논문으로는 이해영 외의 연구(2014)를 들 수 있다. 해당 연구는 2013년에 수행된 대통령기록관의 대통령 기록물 주제분류체계에 사용된 용어를 중심으로, 2009년에 제정된 '국가기록원 시소러스 지침'을 준수하여 대통령기록물 건명을 형태소 분석하고, 관련 시소러스와 이용자가 웹사이트에 입력한 검색 용어 및 정보공개 청구용어 등의 현황을 면밀히 분석한 후, 주제 분류체계와의 매핑을 통해 주제 시소러스를 개발하였다.

3. 시소러스 상호운용성을 위한 관계 확장

ISO 25964는 1부와 2부로 구성되어 있으며, 제1부를 기존 표준과 비교하면 1장부터 13장까지는 기존의 시소러스 구축 표준을 종합한 것이고, 14장부터 18장까지는 시소러스 구축을 위한 S/W와 데이터 모형에 관한 지침을 추가한 새로운 부분에 해당된다. 제2부도 두 가지로 구분되어 있는데, 1장부터 16장까지는 상이한 시소러스 간의 매핑 유형과 방법에 대해 다루고 있고, 17장부터 24장까지는 제어어휘 목록을 유형별로 정리하고 있다(ISO, 2011; 박지영, 윤소영, 이혜원, 2017, p. 121 재인용). <그림 3>은 기존의 표준과 개정된 표준을 비교하여 정리하였고, 그 중 본 연구에서 심층적으로 분석할 부분을 표현하였다.

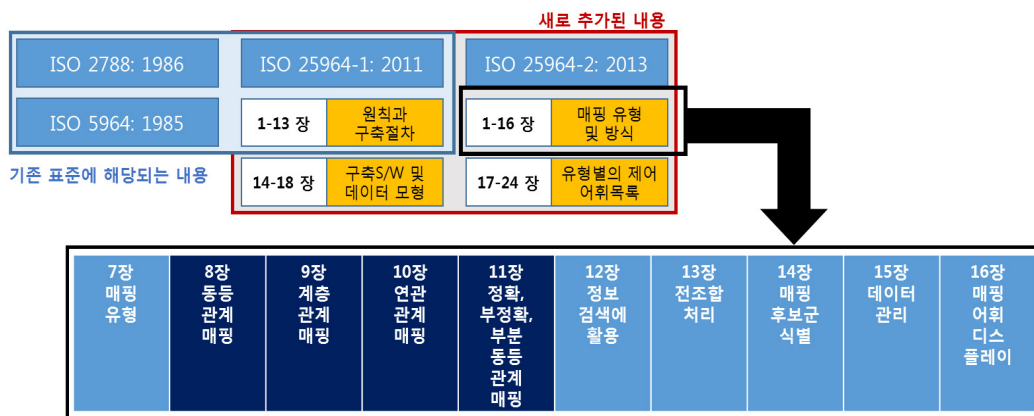
<그림 3>에서 확인할 수 있듯이 ISO 25964-2에서는 기존에 정의된 시소러스 관계를 큰 틀 안에서 유지하면서, 대등관계 매핑을 좀 더 확

장하였다. 즉 시소러스의 기본적인 관계 표현은 제1부에서 제시하였으며, 제2부는 1부를 토대로 다른 체계와의 상호운용성을 위한 매핑 방식을 확장하였다.

3.1 매핑 모델

박지영, 윤소영, 이혜원의 연구(2017)에서는 매핑 모델에 대해 다음과 같이 분석하였다. ISO 25964-2에서는 상이한 시소러스 및 관련 제어어휘 도구에 대한 상세한 매핑 지침을 제공하고 있다. ISO 25964에서 제시한 매핑 방식은 SKOS의 관계 어휘를 이용한 것으로서, 복수의 시소러스에 존재하는 유사한 용어를 1:1로 연계하는 기존의 방식보다 명확하다(박지영, 윤소영, 이혜원, 2017, p. 123).

ISO 25964에서 ‘어휘집’의 범위는 분류체계부터 다양한 지식체계를 다 포함하는 개념이다. <표 1>에서는 상이한 어휘집을 매핑 할 때 참고해야 하는 개념 표현 방식을 정리하였다.



<그림 3> ISO 25964 구성 및 본 연구에서 다루는 영역 표시

* 박지영, 윤소영, 이혜원(2017, p. 122)의 <그림 3>을 확장시킴

〈표 1〉 매핑 시 개념에 대한 표현 방식

어휘집 범위	개념 표현 방식
시소러스(Thesaurus)	우선어(preferred terms)
분류체계(Classification scheme)	기호(notations)
택소노미(Taxonomy)	카테고리 명칭(category label) 또는 기호
주제표목표(Subject heading scheme)	표목(headings)
이름전거리스트(Name authority list)	이름(names)
온톨로지(Ontology)	명칭(labels)
용어집(Terminology)	용어(term) 또는 다양한 표시(designation)

* 출처: ISO(2013, p. 16)

ISO 25964-2에서는 다양한 어휘집들의 매핑을 통해 개념 매핑을 시도하였으며, 그에 대한 방식은 다음과 같다.¹⁾

3.1.1 구조적 통합(Structural unity) 모델

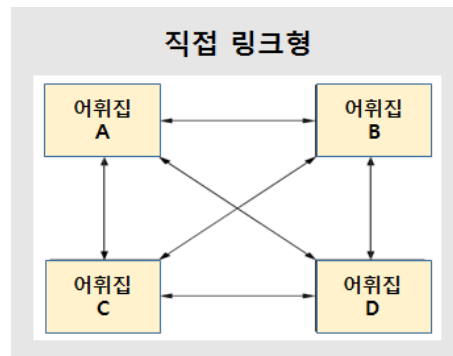
구조적 통합 모델에서는 이종의 시스템에 속한 특정 어휘들이 그 개념 안에서 정확하게 의미를 공유하고 계층 및 연관관계도 동일하게 표현된다. 즉 본 모델은 해당 어휘에 대한 기술과 표현을 다른 언어, 기호, 코드 등을 사용하여 구조화하는 것이다. 대칭 구조로 이루어져 있기 때문에 다국어 시소러스 매핑에 가장 적절한 모델이며, 하나의 시스템 안에서 개념, 용어, 기호, 캡션, 관계를 관리할 수 있는 현실적인 통합 모델이다.

3.1.2 직접 링크(Direct-linked) 모델

직접 링크 모델은 개념 구조가 다른 이종의 어휘집들을 직접 연결하는 것이다. 본 모델은 분류체계, 이름전거리스트 등 범위, 언어, 구조가 다른 어휘집을 통합하기 위한 것이며, 하나 이상의 시소러스 통합에도 적용될 수 있다. 〈그림 4〉를 살펴보면, 노드는 개념을 표현하는 어휘로 다른 어휘와 직접적으로 연결이 되며, 모든

아크는 쌍방향으로 표시되고 있다. 직접 링크에서 4개의 어휘를 연결하기 위해서는 6쌍 즉 12번의 매핑 작업이 필요하다.

만약 매핑 작업의 양을 줄이고자 한다면, 한쪽 방향으로만 색인어나 검색어를 변환하여 적용해야 하며, 이를 〈그림 4〉에 적용하면 아크는 한쪽 방향으로만 향하게 된다.

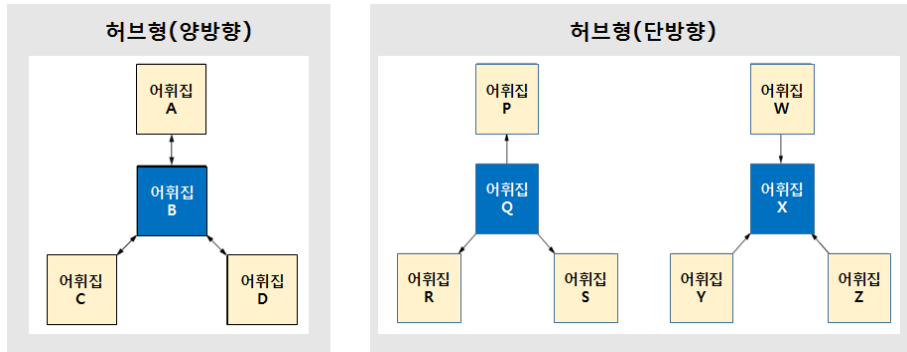


〈그림 4〉 매핑 유형_직접 링크형

3.1.3 허브 구조(Hub structure) 모델

허브 구조는 크게 양방향형과 단방향형으로 나뉜다. 먼저, 허브 구조는 중심이 되는 '허브'와 이를 연결하는 '위성'으로 구성된다(〈그림 5〉참고). 양방향형에서는 허브의 역할을 하는 어휘가

1) ISO 25964-2 p. 17-20 부분을 정리한 것임.



〈그림 5〉 매핑 유형_허브형

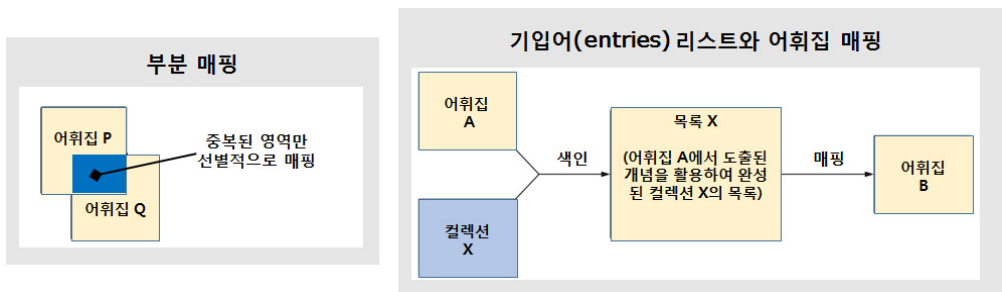
위성의 어휘들과 직접적인 의미 매핑을 실행하며, 허브인 '어휘 B'를 중심으로 '어휘 A', '어휘 C', '어휘 D'가 쌍방향으로 연결된다. 허브 어휘와 위성 어휘간의 매핑은 허브 어휘를 기준으로 색인어 변환이 가능하고, 위성 어휘간의 매핑은 '어휘 A'에서 '어휘 B'로, '어휘 B'에서 '어휘 D'로 허브를 이용한 변환이 이루어진다.

〈그림 5〉의 오른쪽은 허브형 중 단방향성을 표현한 것으로, 왼쪽에 있는 양방향성을 두 가지 형태로 변형한 것이다. 단방향 허브형의 왼쪽 매핑에서는 '어휘 P', '어휘 R', '어휘 S'로 색인된 자원 검색에 '어휘 Q'를 활용한다. 단방향 허브의 오른쪽 매핑은 '어휘 X'로 색인된 자원 검색에 '어휘 W', '어휘 Y', '어휘 Z'를 활용한다.

3.1.4 선택적 매핑(Selective mapping) 모델

지금까지 설명한 구조적 통합, 직접 링크, 허브 구조 모델은 구축 및 유지에 많은 작업이 필요하다. 그러나 매핑 과정에서 통합 대상이 되는 모든 어휘들을 전체적으로 매핑 할 필요가 없을 때도 있다. 이러한 경우에는 선택적 매핑이 시도되며, 그 방법은 〈그림 6〉과 같이, 겹치는 부분만을 선택하여 매핑하는 방식과 소장 자원에만 적용할 수 있는 소규모 색인어 리스트 및 목록을 활용하는 방안이다.

〈그림 6〉 왼쪽의 부분 매핑은 중복된 영역이 적은 경우에 적용된다. 오른쪽의 부분 매핑은 완성된 '어휘 A'를 모두 사용하지 않고 '컬렉션 X'와 관련된 개념만을 추출하여 색인 및 기입어



〈그림 6〉 매핑 유형_선택적 매핑형

(entries) 리스트를 작성하는 방식이다. 즉 '어휘 A'는 컬렉션 X에 속한 개념만이 '어휘 B'와 매핑된다. 본 모델은 비교적 간단한 초기 작업으로 구축될 수 있으나 컬렉션 X의 변화에 민감하게 반응해야 하므로 관리 및 유지가 쉽지가 않다.

3.2 매핑 유형

전통적으로 시소러스 관계는 대등, 계층, 연관 관계로 구별된다. ISO 25964-2에서는 이러한 관계를 그대로 유지하고 그 중 대등관계만을 확장하여 설명하고 있다. ISO 25964-2는

ISO 25964-1에서 제안한 태그 및 표시기호를 그대로 수용하고, <표 2>와 같이 추가된 매핑 유형을 제시하였다.

<표 2>의 표현 중 원천 어휘(source vocabulary)는 검색의 시작점(starting point)으로, 다른 어휘집을 활용 때 사용되는 용어나 개념을 의미하며, 목표 어휘(target vocabulary)는 원천 어휘 안에 존재하는 용어나 개념과 상응하는 것으로, 활용하고자 하는 어휘집의 용어나 개념을 뜻한다.

<표 2>를 중심으로 ISO 25964-1의 시소러스 관계와 ISO 25964-2의 태그를 정리하면 <표 3>과 같다.

<표 2> ISO 25964-2에서 추가된 매핑 유형

태그 및 표시기호	매핑 유형	의미
EQ	대등관계 매핑	<ul style="list-style-type: none"> ■ 대등하게 연결 ■ 'EQ' 태그를 중심으로, 앞의 용어(원천 어휘 안에 존재)와 뒤에 있는 용어(목표 어휘 안에 존재)가 의미적으로 가장 가깝다는 것을 표시
=	대등관계 매핑	<ul style="list-style-type: none"> ■ 등호 표시(유니코드 U+003D에 해당) ■ '=EQ'의 의미는 용어간의 관계가 정확하게 대등함을 표시
~	대등관계 매핑	<ul style="list-style-type: none"> ■ 물결 표시(유니코드 U+007E에 해당) ■ '~EQ'의 의미는 용어간의 관계가 부정확하게 대등함을 표시
BM	계층관계 매핑	<ul style="list-style-type: none"> ■ 큰 개념으로 매핑 ■ 'BM' 태그 뒤에 오는 용어가 더 넓은 의미를 가지고 있음을 표시
NM	계층관계 매핑	<ul style="list-style-type: none"> ■ 좁은 개념으로 매핑 ■ 'NM' 태그 뒤에 오는 용어가 더 명확한 의미를 가지고 있음을 표시
RM	연관관계 매핑	<ul style="list-style-type: none"> ■ 관련 개념으로 매핑 ■ 'RM' 태그 뒤에 오는 용어와 관련이 있음을 표시 ■ 동의어, 유사동의어, 상위어, 하위어 등을 제외
	대등관계-OR 매핑	<ul style="list-style-type: none"> ■ 수직선, 파이프문자 표시(유니코드 U+007C에 해당) ■ 목표 어휘를 활용할 때 두 개 이상의 우선어를 가질 수 있음을 표시 ■ 원천 어휘집의 상위어(상위 개념의 우선어) 범위를 가장 잘 표현하는 목표 어휘집 우선어들의 합 ■ 예시 원천 어휘 가축 목표 어휘 양 소 돼지 가금 ■ 불리언 'OR' 조합을 의미
+	대등관계-AND 매핑	<ul style="list-style-type: none"> ■ 플러스 표시(유니코드 U+002B에 해당) ■ 목표 어휘를 활용할 때 두 개 이상의 우선어를 가질 수 있음을 표시 ■ 원천 어휘의 복합 개념을 표현하기 위한 목표 어휘집 우선어들의 결합 ■ 예시 원천 어휘 여성 경영진 목표 어휘 여성 + 경영진 ■ 불리언 'AND' 조합을 의미

* 출처: ISO(2013), p. 15 <표 1>을 재정리함

〈표 3〉 ISO 25964-1의 시소러스 관계와 ISO 25964-2의 태그 비교

ISO 25964-1 시소러스 관계	ISO 25964-2 관계 매핑(태그 및 표시기호)	
	주요 관계 매핑	선택적인 하위 매핑
USE	EQ	=EQ ~EQ
UF		
USE [A] + [B]*	EQ [A] [B] EQ [A] + [B]	
UF+		
BT	BM	
NT	NM	
RT	RM	

* 표시는 실제 존재하지 않음

3.2.1 대등관계 매핑

대등관계 매핑은 가장 이상적인 통합으로 어휘가 일대일로 대응되는 경우이다. 〈표 4〉가 대표적인 예시이며, 개념에 대한 의미 공유가 바로 이루어진다.

〈표 4〉 대등관계 매핑 예시

어휘집 1	어휘집 2
핸드폰	휴대폰
자두	오얏

어휘집 간의 아크가 쌍방향으로 표시되는 일대일 대등관계 매핑은 매우 이상적이다. 그러나 실제적인 정보 상황에서 아크는 한쪽 방향씩 처리되어야 한다. 예를 들면, '어휘집 A'에서 사용하는 '선생님'과 '어휘집 B'에서 사용하는 '선생님'의 의미는 다를 수 있다. '어휘집 A'의 '선생님'은 학교와 대학에서 가르치는 사람을 통칭하는 의미로 사용되고, 반면 '어휘집 B'의 '선생님'은 학교에서 학생을 가르치는 사람만을 국한시켜 대학에서 학생을 가르치는 '강사'

와 그 의미가 구별된다.

1) 복합 대등관계

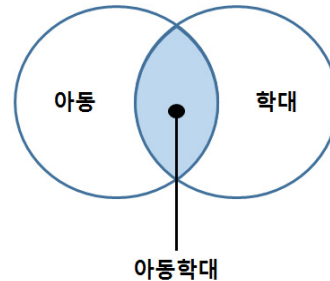
ISO 25964-2에서는 복합어의 대등관계를 표현하기 위해 복합 대등관계(compound equivalence)라는 장치를 도입하였다. 복합 개념은 하나의 어휘로 표현되며, 해당 어휘는 다른 어휘집에 존재하는 두 개 이상의 우선어의 조합으로 이루어져 있다. 이러한 관계는 〈표 2〉와 〈표 3〉에서 살펴본 'EQ+', 'EQ |'와도 연결된다.

'EQ+'로 표시되는 교차 복합 대등관계(intersecting compound equivalence)는 특정 어휘의 의미가 다른 어휘집에 속해 있는 두 개 이상의 어휘들의 의미 교집합으로 설명되는 관계를 말한다(〈표 5〉와 〈그림 7〉 참고). 〈그림 7〉을 살펴보면, '아동 학대' 세트는 '아동'과 '학대' 세트의 교차점인 세브 세트와 동일하다는 것을 알 수 있다.

색인어 매핑 시 불리언 연산자를 사용하지는 않지만, 교차 복합 대등관계 매핑방식은 불리언 연산자 'AND'와 동일한 개념이다.

〈표 5〉 교차 복합 대등관계 매핑 예시

어휘집 1	어휘집 2
아동 학대	아동 학대
철도승객안전	철도 승객 안전
핸드폰 거치대	핸드폰 거치대



〈그림 7〉 교차 복합 대등관계 매핑

‘EQ 1’로 표시되는 누적 복합 대등관계 (intersecting compound equivalence)는 특정 어휘의 의미가 다른 어휘집에 속해 있는 두 개 이상의 어휘들의 의미 합으로 설명되는 관계를 말한다. 〈그림 8〉은 ‘화석연료’가 ‘석탄’, ‘석유’, ‘천연가스’의 합으로 이루어져 있음을 보여준다.

누적 복합 대등관계에서 단순히 시도되는 개념 결합은 양쪽 방향으로 모두 이루어지지 않아도 되지만, 복잡한 개념 표현을 위한 지속적인 노력은 요구된다. 〈표 6〉을 살펴보면, 누적 복합 대등관계의 두 가지 특징을 확인할 수 있다. 첫째, ‘어휘집 1’에서 제시한 어휘의 의미를 ‘어휘집 2’의 어휘들을 통해 표현할 수는 있지만, 그 역의 관계는 정확하지 않다는 것이다. 역의 관계는 또 다른 검증의 절차를 통해 확인되어야 한다. 둘째, 누적 복합 대등관계는 전통

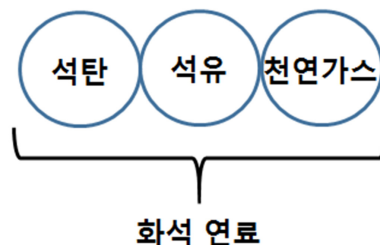
적인 시소러스의 계층관계와 유사하다. 〈표 7〉의 예를 활용하면, ‘동물성 지방’과 ‘식물성 지방’은 ‘지방’의 하위어가 될 수 있으며, ‘지방’은 ‘동물성 지방’과 ‘식물성 지방’의 상위어가 될 수 있다. 즉 매핑 시, 단순히 누적 복합 대등관계로 처리할지 아니면 계층관계로 표현할지를 결정해야 한다.

색인어 매핑 시 불리언 연산자를 사용하지는 않지만, 누적 복합 대등관계 매핑방식은 불리언 연산자 ‘OR’과 동일한 개념이다.

복합 대등관계 매핑 방식에서는 어휘 출처를 확인할 수 있는 한정어(qualifier vocabulary) 활용도 고려해야 한다. ‘어휘집 A’에서 나온 어휘를 사용한다고 가정하면, 해당 어휘 앞에 ‘어휘집 A’를 확인할 수 있는 한정어 ‘VocA’를 함께 기입하는 것이다.

〈표 6〉 누적 복합 대등관계 매핑 예시

어휘집 1	어휘집 2
필기도구	종이 볼펜 연필
화석 연료	석탄 석유 천연가스
지방	동물성 지방 식물성 지방



〈그림 8〉 누적 복합 대등관계 매핑

2) 정확한, 부정확한, 부분적 대등관계

ISO 25964-2에서는 대등관계의 정도를 표현할 수 있는 장치를 제안하였다. <표 7>과 같이, 정확한, 부정확한, 부분적인 대등관계 매핑을 정의하고 있다.

'~EQ'로 표시되는 매핑은 두 어휘간의 의미가 정확하고 대등하게 연결된다는 것을 보여준다.

<표 7> 정확한 대등관계 매핑 예시

어휘집 1	어휘집 2
랩탑	노트북
부추	술

정확한 대등관계 매핑은 쌍방향 변환은 가능하지만, 이를 위해서는 양쪽 방향을 모두 확인하는 이중(two-way) 변환 방식으로 작업되어야 한다.

'~EQ'로 표시되는 매핑은 두 어휘간의 의미가 부정확하거나 부분적으로 연결되는 것이다. 먼저 부정확한 대등관계 매핑을 살펴보고자 한다. 두 개 이상의 어휘집을 매핑하는 것, 즉 상이한 어휘집이 가지고 있는 개념을 정확하게 연결하는 것은 불가능할 때가 있다. 특히 문화가 다른 집단들의 어휘집을 개념적으로 변환할 때 그 어려움은 가중된다.

<표 8>에 제시한 예를 살펴보면, 이용자에 따라 암과 종양은 같은 의미로 인식될 수 있다. 그러나 학문적으로는 암은 악성종양만을 의미한다. 다른 예시인 '문화체육관광부'는 2008년도에 '문화관광부'와 '국정홍보처'를 통합하여 만들어진 행정부처이다. 그러므로 '문화체육관광부'와 '문화관광부'가 동일한 기관이라고 인식될 수도 있고, 다른 기관이라고도 판단될 수도 있다.

<표 8> 부정확한 대등관계 매핑 예시

어휘집 1	어휘집 2
암 (cancer)	종양 (tumor)
동무	친구
문화체육관광부	문화관광부

부정확한 대등관계 매핑은 두 어휘간의 의미가 얼마나 겹치는가를 판단하는 것이 중요하다. 이러한 판단은 검색 결과에 적합문서가 얼마나 포함되는가에 의해 결정된다.

'~EQ'로 표시되는 또 다른 대등관계 매핑은 두 어휘간의 의미가 부분적으로 연결되는 것이다. 두 개 이상의 어휘집을 대등적으로 매핑할 때, 가끔 발생할 수 있는 상황은 하나의 어휘가 다른 어휘보다 큰 개념을 가지고 있을 때이다. 즉 큰 개념인 어휘에 작은 개념의 어휘가 부분적으로 대등관계로 연결되어 의미 변환이 일어나는 것이다. <표 9>를 살펴보면, '군주', '항공기', '학교 경내'가 상위 개념이면서 더 포괄적 의미를 담고 있으며, '왕', '여객기', '학교 건물'은 앞의 어휘에 비해 하위 개념이면서 더 상세한 의미를 표현하고 있지만, 상황에 따라서는 두 어휘가 같은 의미로 쓰이기도 한다.

<표 9> 부분적 대등관계 매핑 예시

어휘집 1	어휘집 2
군주	왕
항공기	여객기
학교 경내	학교 건물

부분적 대등관계 매핑임을 결정하는 확실한 방법을 없으나, 다음과 같은 상황을 고려한다면 그 답을 찾을 수도 있다. 첫 번째로는 계층관계 매핑으로 선언할 가능성에 대해 고민해야 한다.

두 번째로는 계층관계 매핑과 다음에서 설명하는 세 번째 매핑 유형에 중복적으로 적용할 수 없음을 확인해야 한다. 세 번째로는 누적 복합 대등관계 매핑으로 선언할 가능성에 대해 고민해야 한다. 누적 복합 대등관계 매핑은 특정 어휘의 의미가 다른 어휘집에 속해 있는 두 개 이상의 어휘들의 의미 합으로 설명되는 관계이므로, 특정 어휘의 의미를 전달할 수 있는 어휘들을 다 포함하고 있어야 한다. 예를 들어, '항공기 EQ 여행기 | 전투기'로 정의한다면 항공기와 관련된 다른 개념(경비행기, 군용기 등)을 추가하는 것이 불가능하다. 이러한 경우에는 계층관계 매핑을 통해 관계 정의가 완전하게 이루어진 어휘만을 표현하는 것이 바람직하다. '항공기 NM 여행기', '항공기 NM 전투기' 등으로 나눈 계층관계를 표현해야 한다.

3.2.2 계층관계 매핑

ISO 25964-2 계층관계 매핑은 ISO 25964-1에서 제시한 시소러스 계층 관계를 그대로 수용하였다. 또한 시소러스 계층 관계의 종속, 사례, 전체-부분 등을 적용하여, 종속매핑(BMG/NMG), 사례매핑(BMI/NMI), 전체-부분매핑(BMP/NMP) 유형을 제안하였다. 계층관계 매핑은 상호간의 계층 연결이 가능하다.

3.2.3 연관관계 매핑

ISO 25964-2 연관관계 매핑은 ISO 25964-1에서 제시한 시소러스 연관 관계를 그대로 수용하였다. 연관관계 매핑은 쌍방향으로 의미 연결이 가능하다.

지금까지 이종의 어휘집의 의미 공유 방식을 살펴본 바에 따르면 ISO 25964-2의 매핑 유형

은 대등관계 매핑에 집중하는 것을 확인할 수 있었다(〈표 2〉와 〈표 3〉 참고).

4. 매핑 유형을 고려한 온톨로지 모형 구축 및 활용

4장에서는 ISO 25964-2의 상호운용성에 초점을 두고 시소러스 관계와 온톨로지 변환 시 고려해야 할 사항들을 확인하고, 실제 데이터를 활용하여 그 효용성을 증명하였다.

4.1 온톨로지 변환 시 고려점

ISO 25964-2에서는 상호운용성을 적용해야 할 다양한 어휘집의 유형으로 시소러스뿐만 아니라 분류체계, 기록관리 분류체계, 택소노미, 주제표목표, 온톨로지, 용어집, 이름전거리스트, 동의어 리스트 등을 포함시켰다. 그 중 온톨로지 모형은 현재의 시맨틱 웹 정보환경을 수용하며, 개념의 의미를 공유하는 지식 틀로 많은 강점을 가지고 있다. 첫 번째의 강점은 개념을 클래스, 속성, 관계, 인스턴스 등으로 구별하여, 개념 차원에 대한 설명을 다차원적으로 표현할 수 있다는 것이다. 두 번째의 강점은 개념 확장이 용이하다는 것이다. European Data Model(EDM), Open Archives Initiative Object Reuse and Exchange(OAI-ORE), Simple Knowledge Organization System(SKOS), Friend-of-A-Friend(FOAF) 등과 같은 지식 체계나 더블링크어와 같은 메타데이터 체계와의 연계도 가능하다. 또한 인스턴스를 지속적으로 추가함으로써 개념 표현의 질을 높이고, 범

위도 확장할 수 있다. 세 번째의 강점은 웹 온톨로지 언어(OWL: Web Ontology Language)를 사용하여 개념의 표현을 상세하게 기술한다는 것이다. 마지막 강점으로는 특정 주제영역에 대한 상세한 표현들을 수집할 수 있다. 특정 주제 영역을 공유하는 복수의 어휘집의 개념을 연결함으로써 의미의 상호운용성을 구축하고 확장된 개념을 공유할 수 있다.

시소러스와 다양한 지식체계와의 상호운용성을 살피기 위해, 본 연구에서는 시소러스와 온톨로지와의 상호운용성을 구체적으로 분석하였다. 본 연구에서 다양한 어휘집 중 온톨로지 모형으로의 변환을 시도한 근거와 시소러스와 온톨로지와의 차이점을 함께 알아보았다.

위에서 제시한 온톨로지 모형의 강점을 중심으로 첫 번째 근거 및 차이점을 설명하였다. 첫째, 시소러스의 개념과 온톨로지의 클래스의 차이이다. 시소러스는 개념의 의미를 용어로 표현하고, 관계 개념, 한정어, 범위주기, 정의 등을 정보검색 이용자에게 제공하고자 하지만, 온톨로지는 클래스의 의미를 기계가 이해할 수 있는 정보 환경에서 전달하고자 한다(ISO, 2013, p. 76). 시소러스는 이용자 중심으로 정보를 표현하고자 하였으나, 디지털 환경에서는 적합하지 않을 수도 있다. 정보 전달이 거의 전자적인 환경에서 이루어지기 때문에, 기계가 이해하는 언어로 표현하는 것이 매우 중요하다. 온톨로지는 기계와 인간이 모두 이해할 수 있는 언어를 사용하여 개념공유나 의미 전달에 바로 활용될 수 있다는 강점이 있다. 둘째, 시소러스는 정보검색에 초점을 맞추어, 우선어 및 비우선어를 표시하는데, 온톨로지는 클래스와 인스턴스의 명칭을 통해 표현한다(ISO, 2013, p. 76). 정보검색에서 색인

및 검색의 접근점을 제시하는 것은 매우 중요하며, 이는 시소러스의 근본적인 목표이다. 반면, 접근점 처리를 통해 이용자의 검색을 지원하는 것은 온톨로지의 근본적인 목표가 아니다. 온톨로지에서는 우선어와 비우선어를 클래스와 인스턴스의 다양한 명칭으로 표현하고, 속성으로 명칭의 특성을 확인할 수 있으며, 그에 대한 추가 및 수정 작업도 용이하도록 설계된다. 마지막으로, 시소러스에서는 자연언어를 정확하게 표현하기 위해 온톨로지보다 더 많은 노력을 해야 한다. 예를 들면, 시소러스에서는 두 개 이상의 의미를 가진 'Moon'이라는 용어에 대한 처리를 복수형인 'Moons'로 하게 된다. 이는 애매성을 해소하기 위한 것으로, 특정 'Moon'을 사용할 수 있는 가능성을 남겨두는 것이다. 반면 온톨로지는 다양한 개념은 식별자로 구별되어지며, 개념에 대한 명칭은 필수적인 요소가 아니므로, 명칭의 표시 유무는 중요하지 않다.

두 번째 특징은 온톨로지 모형이 이용자 검색 요구를 수용할 수 있는 유연한 지식 체계라는 것이다. 이용자의 검색 내비게이션은 핵심형과 참고형으로 구별된다. 핵심 검색 내비게이션은 수직적인 방식과 수평적인 방식으로 다시 구분된다. 수직적인 방식은 특정 용어를 중심으로 상위어와 하위어를 함께 고려하는데, 수평적인 방식은 특정 용어의 개념과 유사한 의미를 탐색하게 된다. 이용자의 다양한 정보요구를 수용하기 위해, 수직적인 방식에서는 계층관계의 기준점을 제공해야 하고, 수평적인 방식에서는 개념의 유사성의 근거를 제공해야 한다. 참고형 검색 내비게이션은 대등 및 계층관계에서 다루지 않은 모든 개념을 연관관계로 탐색하는 방식이다. 이용자 내비게이션 방식에 최적화된 개념

표현은 온톨로지 모형에서 구현될 수 있다.

세 번째 특징은 온톨로지 모형의 구성요소이다. 온톨로지는 클래스, 속성, 인스턴스로 구성된다. 클래스는 개념을 의미하며, 속성은 그 개념에 대한 설명 및 표현이다. 속성은 데이터타입과 오브젝트타입으로 구별되는데, 데이터타입 속성은 개념 자체에 대한 설명으로, 대표적으로 식별자, 명칭 등을 들 수 있다. 오브젝트타입 속성은 클래스 간의 관계 즉 개념 간의 관계를 표현한다. 인스턴스는 클래스의 사례이다. 인스턴스를 수집하고 관리하는 작업은 온톨로지 구축 과정에서 매우 중요하다. 왜냐하면 특정 주제 영역의 시소러스가 관련 용어의 리스트를 완성하듯이, 특정 주제 영역의 온톨로지는 클래스의 관련 인스턴스를 입력해야 완성되는 것이다.

네 번째의 특징은 지식시스템과의 연계가 가능하다는 것이다. ISO 25964에서는 기본 구조를 계승하면서도 부가적인 기능을 담을 수 있는 데이터 모형을 제안하였다. 이 표준에서는 시소러스를 일종의 지식조직시스템(Knowledge

Organization System, KOS)으로 간주하고, 특정 KOS를 기계 처리가 가능한 형태로 발행하며 시맨틱 웹 환경에 적용 가능하도록 설계하기 위해 도입하였다(Will, 2012; 박지영, 윤소영, 이혜원, 2017, p. 122 재인용).

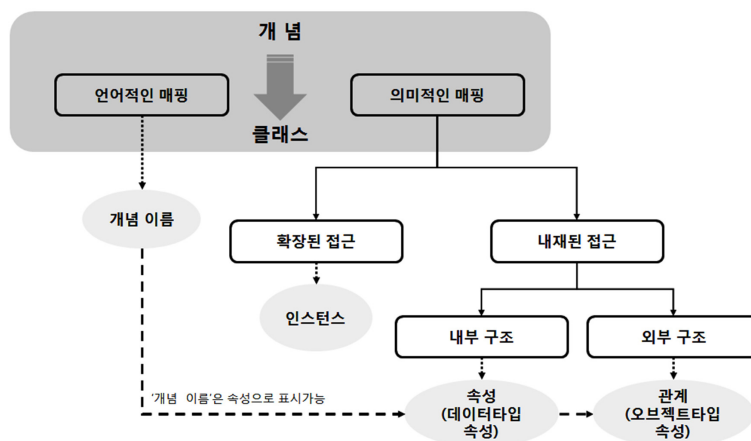
〈표 10〉은 SKOS와 ISO25964-2 매핑 유형을 비교한 것이다. 본 연구에서도 SKOS의 관계유형을 참고하였다.

〈표 10〉 SKOS와 ISO25964-2 매핑 유형 비교

SKOS	ISO 25964-2
skos:closeMatch	Equivalence
skos:exactMatch	Exact equivalence
-	Compound equivalence
skos:relatedMatch	Associative
skos:broadMatch	Hierarchical (broader)
skos:narrowMatch	Hierarchical (narrower)

출처: Clarke(2011)

본 연구에서 적용한 시소러스와 온톨로지의 매핑 방식은 Roche, Damas, & Roche의 연구(2014)를 바탕으로 재정리하였다. 〈그림 9〉를



〈그림 9〉 온톨로지 매핑 방식

출처: Roche, Damas, & Roche(2014), p. 10 〈그림 7〉을 바탕으로 재작성함

살펴보면 다음과 같다. 개념은 온톨로지 모형에서 클래스로 표현되며, 개념에 대한 의미 공유는 언어적인 매핑과 의미적인 매핑을 통해 이루어진다. 언어적인 매핑은 개념을 대표하는 이름으로 속성으로 표현될 수 있다. 의미적인 매핑은 클래스의 확장된 접근과 내재된 접근으로 나눌 수 있는데, 확장된 접근은 클래스의 사례인 인스턴스를 추가하는 것으로 이루어지며, 내재적인 접근은 클래스의 속성과 관계를 표현하는 것을 의미한다. 특히 클래스 즉 개념 자체에 대한 설명을 데이터타입 속성으로 기술하고, 클래스 간의 구조 즉 외부와의 구조는 오브젝트타입 관계속성으로 표현한다.

4.2 매핑 유형을 고려한 온톨로지 변환 모형 설계

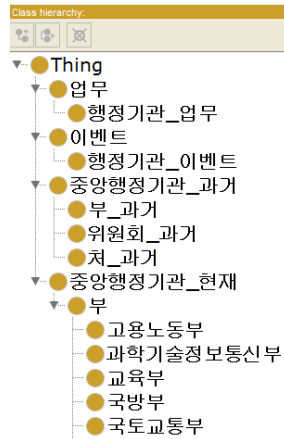
본 연구에서 제안한 온톨로지 변환 모형 설계 원칙은 다음과 같다. 또한 웹 온톨로지 언어 OWL의 특성을 고려하여 설계하였다.

본 연구에서는 매핑 유형을 고려한 시소러스와 온톨로지 간의 상호운용성 모형을 설계하고, 이에 대한 온톨로지 클래스와 속성을 선언하였다. 온톨로지는 특정 주제 영역을 중심으로 전개되는 정보표현 방식이므로, 본 연구에서도 특정 주제 영역을 선정하였다. 본 연구에서 온톨로지 변환 모형의 효용성을 확인하기 위해 활용한 데이터의 선정 기준은 다음과 같다. 웹 기반 시소러스 확보가 가능하고, 데이터에 대한 설명이 다양한 방식으로 웹 상에 존재하며, 무엇보다 대등관계 매핑을 시도할 수 있는 복잡한 연결 구조를 내포하고 있어야 한다. 이러한 선정 기준을 고려하여, 본 연구에서는 행정부처 시소

러스를 채택하였다. 행정부처 시소러스는 국가 기록원 웹사이트에서 확인할 수 있었으며, 행정부처에 대한 정보는 해당 부처의 홈페이지, 포털사이트, 위키백과 등에서 쉽게 얻을 수 있다. 마지막으로 중앙행정기관은 집권하는 정권에 따라, 사회의 요구에 따라 신설, 통합, 폐지, 부활되기도 한다. 마지막 선정 기준은 본 연구의 범위와도 관련이 있다. 본 연구의 범위는 시소러스와 다양한 어휘집간의 의미 연결 유형을 심층적으로 살펴보고, 이를 확인하기 위해 시소러스와 온톨로지 상호운용성을 분석하고 활용가능성을 제시하는 것이었다. 본 연구에서 활용되는 데이터 범위는 전체 시소러스 셋이 아니고 시소러스 일부 즉 행정부처의 명칭 변화를 잘 보여주는 데이터만으로 국한하였다.

본 연구에서는 ISO 25964-2에서 강조하는 대등관계를 중심으로 한 온톨로지 모형을 설계하였으며, 클래스, 속성, 관계, 인스턴스 등을 구축하는데도 이점을 고려하였다. 구체적인 온톨로지 모형 설계 및 구축 과정을 설명하면 다음과 같다.

먼저, 클래스 선언과 관련하여 본 연구에서는 다음과 같은 사항들을 고려하였다. 첫째, OWL의 클래스 표현은 계층관계를 포함하고 있지만, 계층관계를 중요시하는 시소러스를 고려한다면 속성 등을 통해 그 의미를 구체화해야 한다. 둘째, 클래스는 개념을 표현하는 영역으로 온톨로지 구축 범위인 주제영역과 직접적인 관련이 있다. 추가되는 클래스는 행정부처 명칭의 변경과 관련 정보를 담을 수 있는 개념이 반영되어야 한다. <그림 10>의 '행정기관 이벤트' 클래스는 행정부처 명칭의 변화를 정리할 수 있는 구심점이 된다. '행정기관 이벤트' 클래스는 <그림 11>과 같이 활용된다.



〈그림 10〉 클래스(일부)



〈그림 11〉 이벤트 클래스 활용

〈그림 11〉을 살펴보면, ‘미래창조과학부’는 방송통신위원회, 문화체육관광부, 교육과학기술부 등의 기관과 연계되어 2013년 2월 23일에 개편된 행정부처이다. 관련 기관과의 연계성을 확인하기 위해 본 연구에서는 ‘이벤트’ 클래스의 인스턴스 ‘event_2’를 활용하였다. ‘event_2’는 관련 기관과의 연결을 통합하는 상위인스턴스이며, ‘event_21’, ‘event_31’, ‘event_41’ 등의 하위 인스턴스를 가지고 있다. 인스턴스의 연결은 클래스의 관계 속성으로 결정된다. 즉 ‘이벤트’ 클래스 간의 연계는 오브젝트타입속성 ‘hasEvent’, ‘hasEventUpper’를 통해 표현되었

다. ‘event_21’은 ‘기관이관’이라는 속성의 값으로 구별되어지며 ‘방송통신위원회’와 ‘미래창조과학부’ 간의 연결 관계를 구체화하고 있다. ‘event_31’은 ‘사무이관’이라는 속성의 값으로 구별되어지며 ‘문화체육관광부’와 ‘미래창조과학부’ 간의 연결 관계를 구체화하고 있다. 마지막으로 ‘event_41’은 ‘기관분리’라는 속성의 값으로 구별되어지며 ‘교육과학기술부’와 ‘미래창조과학부’ 간의 연결 관계를 구체화하고 있다.

다음은 데이터타입 속성을 정의할 때의 고려점을 기술하였다. 첫째, 시소러스와의 매핑 시 시소러스 구축 지침(ISO, 2011, p. 13)에서 제안

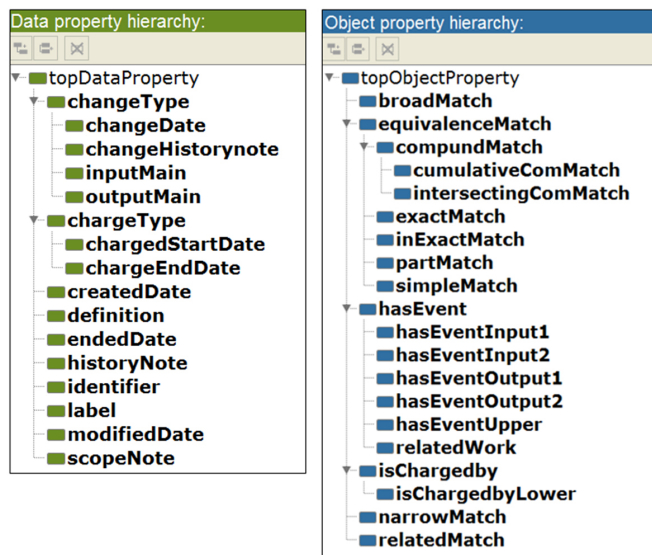
한 기본적인 기술요소(범위주기, 정의, 이력주기)와 코드(주제 카테고리, 개념코드) 등을 활용해야 한다. 둘째, 일부 요소들은 그 값을 부여하는 규칙이 필요하다. 예를 들면, <dc:creator>, <dc:subject> 요소의 값으로 저자명, 기관명, 주제명 등을 기입해야 하는데, 자연언어를 입력할지 또는 다른 어휘집을 활용할지를 결정해야 한다.

<그림 12>에서 제시된 데이터타입 속성 중 본 연구의 특성을 반영하는 'changeType'과 'chargeType'을 살펴보면 다음과 같다. 'changeType' 속성은 기관의 변화를 확인하는 것으로 changeType(변화유형)을 중심으로 변화가 일어난 시점(changeDate), 변화이력에 대한 설명(changeHistorynote), 변화의 중심이 되는 변경 전 기관(inputMain), 변화의 중심이 되는 변경 후 기관(outputMain) 등으로 이루어져 있다. 변화의 중심이 되는 변경 전 기관(inputMain)과 변화의 중심이 되는 변경 후 기

관(outputMain)을 따로 관리한 이유는 다음과 같다. 만약, 두 기관이 통합되어 다른 기관이 만들어진 경우에 통합의 대상이 된 기관 중 업무 및 기능의 비중이 더 많은 기관이 있다. 예를 들면, 위에서 설명한 '문화체육관광부'는 '문화관광부'와 '국정홍보처'와의 '기관통합'으로 이루어진 조직이다. 이때 변화의 중심이 되는 변경 전 기관(inputMain)은 '문화관광부'이며, 변화의 중심이 되는 변경 후 기관(outputMain)은 '문화체육관광부'가 된다.

'chargeType' 속성은 클래스 'work(업무)'와 오브젝트타입 'isChargedby'와 관련이 있다. 행정부처의 변화는 기관 대 기관의 경우도 있고, 기관의 업무 대 기관의 업무의 변화도 있을 수 있다. <그림 11>에서 표현된 '사무기관'이 대표적이다.

오브젝트타입 속성은 본 연구에서 분석한 때 평유형이 가장 잘 반영되는 부분이다. <표 11>에서는 복잡해진 대등관계를 보다 간결하고 정



<그림 12> 데이터타입 및 오브젝트타입 속성(일부)

〈표 11〉 오브젝트타입 속성: ISO25964-2 매핑 유형, SKOS와 비교

오브젝트타입 속성	ISO25964-2 매핑 유형	SKOS
equivalenceMatch	Equivalence	skos:closeMatch
simpleMatch	Simple equivalence	-
exactMatch	Exact equivalence	skos:exactMatch
inExactMatch	Inexact equivalence	-
partMatch	Part equivalence	-
compundMatch	Compound equivalence	-
intersectingComMatch	Intersecting	-
cumulativeComMatch	Cumulative	-
relatedMatch	Associative	skos:relatedMatch
broadMatch	Hierarchical (broader)	skos:broadMatch
narrowMatch	Hierarchical (narrower)	skos:narrowMatch

확하게 정리하였다. 이 외에 추가된 요소들은 <그림 12>에서 제시하였다.

4.3 온톨로지 변환 모형의 활용성 확인

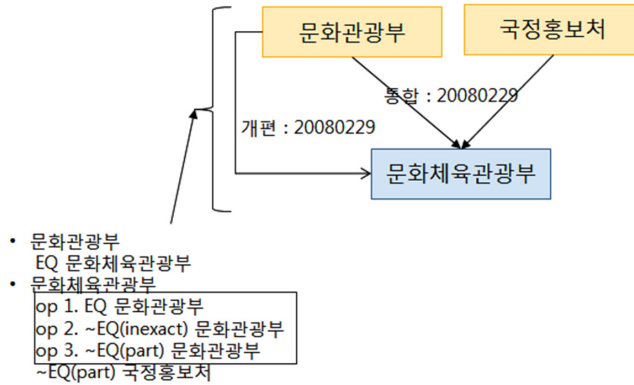
본 연구에서는 온톨로지 구축 틀인 protégé 4.3을 이용하여 행정부처 관련 온톨로지 모형을 구축하였다. 특히 본 연구에서는 대등관계 매핑 유형을 집중하여 개념을 표현하였다. 인스턴스는 행정부처 시소러스를 활용하여 구축하였으며, 본 연구에서 다른 데이터는 행정부처 중 다른 기관과의 변화 이력이 많은 것으로 설정하였다. 본 연구에서는 현재 존재하는 행정부처 중 '과학기술정보통신부', '교육부', '문화체육관광부', '중소벤처기업부'를 중심으로 기술하였다.

대등관계 매핑이 쌍방향으로 진행될 때, 두 방향이 같은 매핑 유형으로 표시되는 경우가 드물다. <그림 13>을 살펴보면, '문화관광부'는 2008년 2월 29일에 '문화체육관광부'로 개편되었는데, 그때 '국정홍보처'도 '문화체육관광부'

로 통합되었다. '문화관광부'에서 '문화체육관광부'로의 매핑 유형은 정확한 대등관계이지만, 그 역은 부정확한 대등관계 매핑으로 이해될 수 있다. 또 다른 고려점은 대등관계 유형에서 선택적으로 표시할 수 있는 부정확한 또는 부분적인 대등관계 유형을 결정하는 것이 쉽지 않다는 것이다. '문화체육관광부'의 전신은 '문화관광부'이지만 '문화관광부' 업무 이외에도 '국정홍보처' 업무를 함께 인수받았다. 이 점만을 상기하면, '문화관광부'와 '국정홍보처'는 '문화체육관광부'의 부분적인 대등관계로 표현될 수 있다. 그러나 '문화관광부'가 '문화체육관광부'로 개편된 것만을 고려하면 이들의 관계는 정확한 대등관계 매핑이 될 수도 있다. 그러나 개편에서 다른 기관과의 통합도 이루어졌으니 두 기관이 아주 동일한 기관이라고 할 수 없다는 판단도 가능하다. 다양한 가능성 중에 어떤 것을 선택할지는 이용자의 요구와 정보환경을 고려하여 결정되어야 한다.

<그림 14>는 국가기록원의 웹페이지²⁾에서

2) 국가기록원에서는 시소러스 전용 사이트를 제공하지는 않지만 관련용어검색 메뉴에서 검색도구로 시소러스를 사용하고 있다(<http://www.archives.go.kr/next/search/thesaurusSearch.do>). 국가기록원은 관련용어검색이 검색도구 중 하나인 시소러스를 활용한 검색이라는 점을 사이트 안내를 통해 제공하고 있다.



<그림 13> 대등관계 매핑 예시; 문화체육관광부

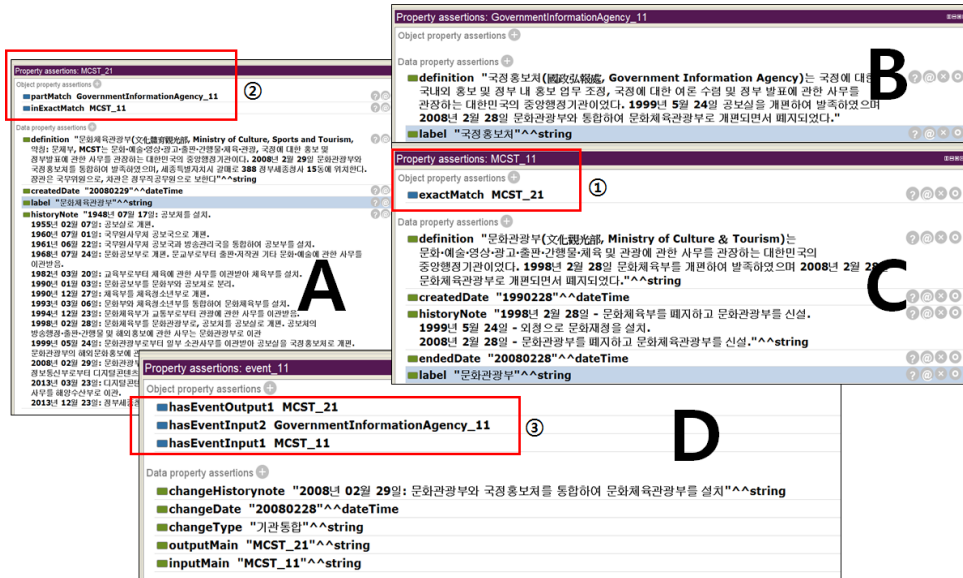
대표어	유의어	상위어	하위어	관련어
<input checked="" type="checkbox"/> 문화 체육 관광부	<input type="checkbox"/> 문공부 <input type="checkbox"/> 문광부 <input type="checkbox"/> 문체부 <input type="checkbox"/> 문화 공보부 <input type="checkbox"/> 문화 관광부 <input type="checkbox"/> 문화 체육부 <input type="checkbox"/> 문화부 <input type="checkbox"/> 체육 청소년부 <input type="checkbox"/> 체육부@정부부처 <input type="checkbox"/> Ministry of Culture and Tourism <input type="checkbox"/> Ministry of culture sports and tourism		<input type="checkbox"/> 국립 국악원 <input type="checkbox"/> 국립 국어원 <input type="checkbox"/> 국립 민속 박물관 <input type="checkbox"/> 국립 중앙 극장 <input type="checkbox"/> 국립 중앙 도서관 <input type="checkbox"/> 국립 중앙 박물관 <input type="checkbox"/> 국립 현대 미술관 <input type="checkbox"/> 문화재청 <input type="checkbox"/> 한국 예술 종합 학교 <input type="checkbox"/> 한국 정책 방송 <input type="checkbox"/> 해외 문화 홍보원	<input type="checkbox"/> 문화 관광 위원회 <input type="checkbox"/> 한국 영상 자료원

<그림 14> 국가기록원 검색화면; 문화관광부 검색 결과

‘문화관광부’를 검색한 것이다. 결과는 현재 행정부처인 ‘문화관광체육부’를 중심으로 선행기관명 및 영문이름 등을 유의어로 표시하고, 관련 하위 기관을 제시하였다. 이보다 행정부처의 변화를 좀 더 정확하게 전달할 방안을 모색할 필요가 있다.

<그림 15>는 네 개의 인스턴스 화면(A-D)을 제시하였다. 화면 A는 문화체육관광부를 기술한 것이고, 화면 B는 국정홍보처, 화면 C는 문화관광부, 화면 D는 기관들의 관계를 설명하는 이벤트 클래스의 인스턴스이다. ①번은 ‘문화관

광부(MCST_11)’를 중심으로 ‘문화체육관광부(MCST_21)’가 동일한 기관이라는 것을 확인한 것이고, ②번은 ‘문화체육관광부(MCST_21)’를 중심으로 ‘문화관광부(MCST_11)’의 연결은 부정확한 대등관계(inExactMatch)로, ‘국정홍보처(Government Information Agency_11)’의 연결은 부분적인 대등관계(partMatch)로 처리한 결과를 보여주고 있다. ③번은 이벤트 클래스의 인스턴스로 세 기관의 관계를 표현하였다. 오브젝트타입 속성을 통해 ‘문화관광부(MCST_11)’와 ‘국정홍보처(Government



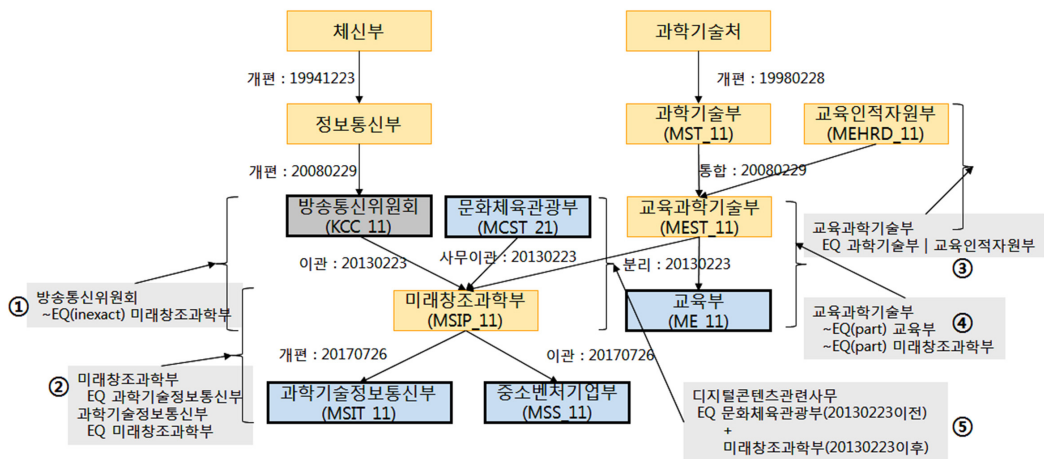
<그림 15> 문화관광부-국정홍보처-문화체육관광부 대등관계 매핑

* 위키백과에서 제시한 변화만을 기준으로 작성됨

Information Agency_11)'가 '문화체육관광부 (MCST_21)'로 통합되었음을 보여주며, 데이터 타입 속성을 통해 통합의 유형(changeType)이나 이벤트 발생 일시(changeDate) 등을 표시

하고, 이 통합의 주요한 기관들을 'inputMain', 'outputMain'으로 표시하여 통합의 적극적인 주체들을 구별하였다.

<그림 16>에서, 굵게 표시된 네모는 현재의

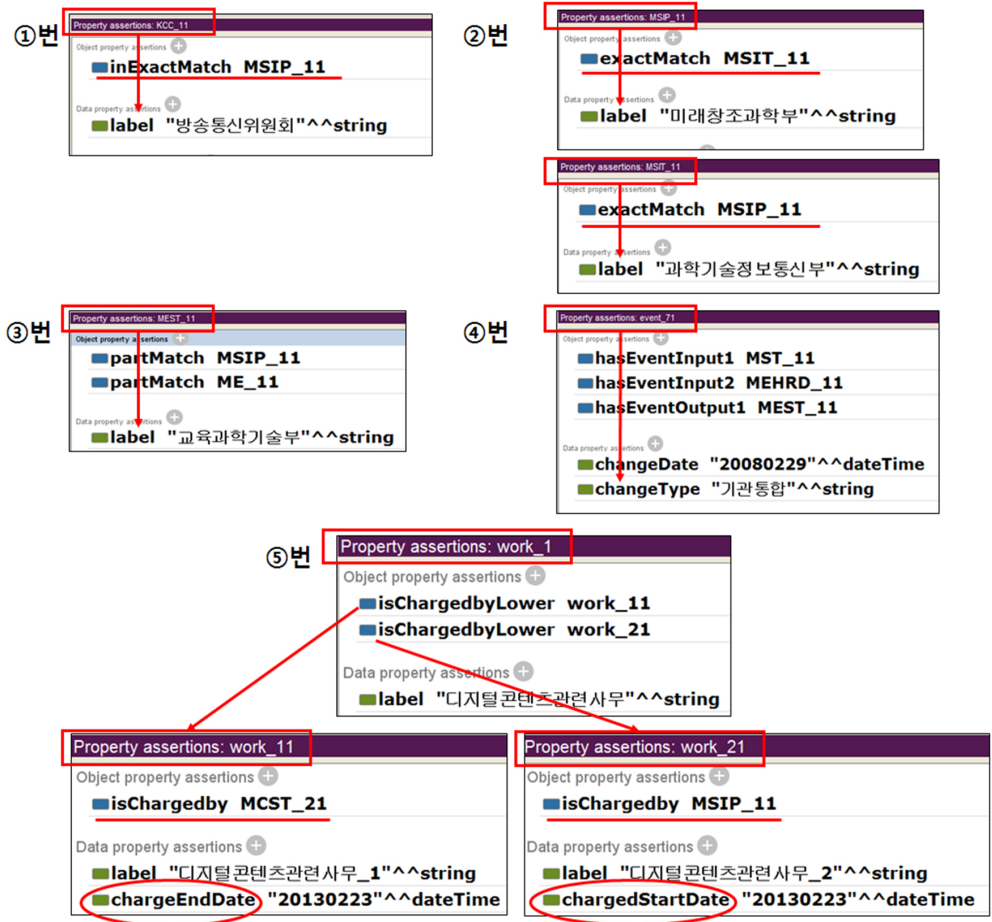


<그림 16> 교육, 과학, 통신 등과 관련된 행정부처의 변화

행정부처를 의미하고 그 외의 네모는 현재는 존재하지 않은 부처를 말한다. 네모 안에 있는 'MCST_21' 등과 같은 기호는 인스턴스 식별자를 의미한다. ①번부터 ⑤번은 대등관계 매핑 유형을 정리하였으며, 온톨로지 표현과도 연관되어 있다.

〈그림 17〉에 대한 설명은 다음과 같으며, 그 이해를 위해 〈그림 16〉을 참고해야 한다. ①번은 '방송통신위원회'를 기준으로 '미래창조과학부'와의 불완전한 대등관계를 표현하였으며, ②

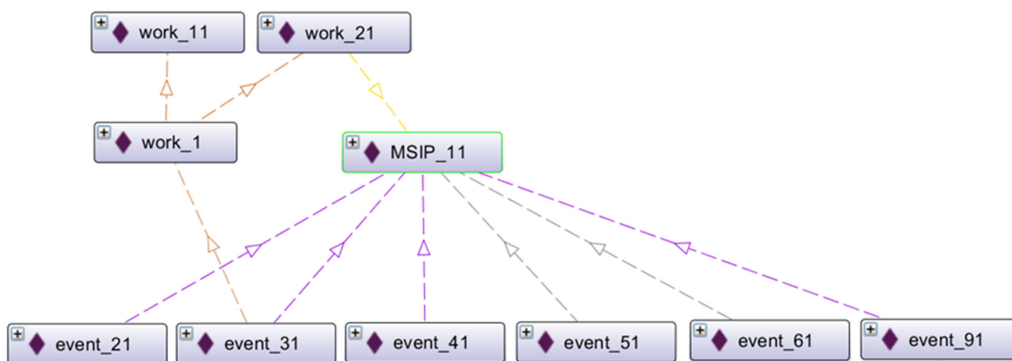
번은 '미래창조과학부'를 기준으로 '과학기술정보통신부'를 완전한 대등관계로 표현하고, 그 역관계도 표시하였다. ③번은 '과학기술부'와 '교육인적자원부'가 통합된 기관인 '교육과학기술부'를 표현하였다. ④번은 ③번의 역관계로 '교육과학기술부'는 '과학기술부'와 '교육인적자원부'가 통합되어 생성된 기관임을 보여주고 있다. ⑤번은 불리언 AND 조합의 의미를 표현한 것으로, '사무이관'이 이루어진 '디지털콘텐츠관련사무'를 책임지는 기관들을 표시한 것이다.



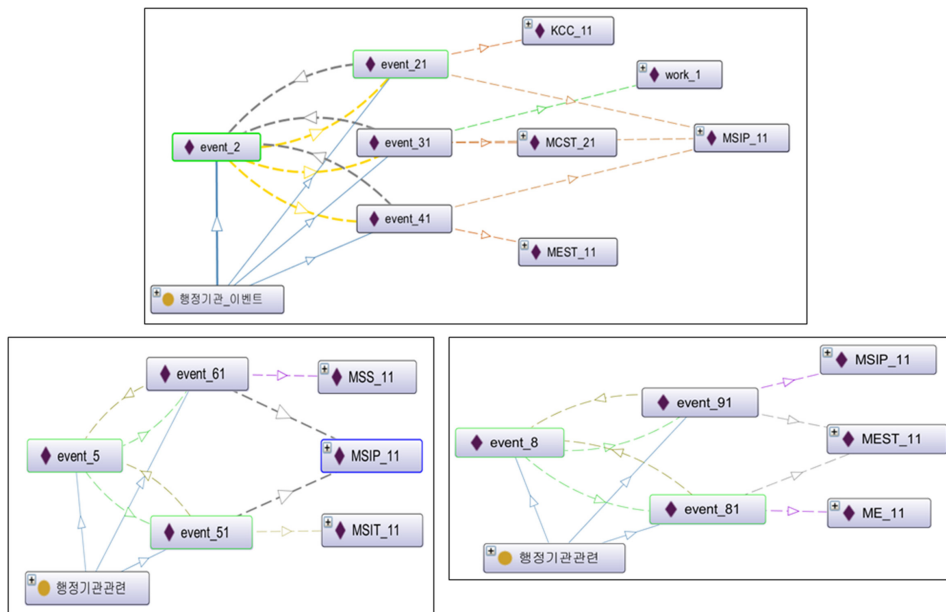
〈그림 17〉 대등관계를 중심으로 확장된 매핑 관계

〈그림 16〉을 살펴보면, 변화가 가장 많이 일어난 기관은 '미래창조과학부(MSIP_11)'이다. '미래창조과학부(MSIP_11)'를 중심으로 인스턴스를 정리하면 다음의 〈그림 18〉과 같으며, 〈그림 19〉는 관련된 인스턴스들을 좀 더 자세히 분석한 것이다.

〈그림 19〉는 각 이벤트들을 통해 다른 기관과의 관계 및 관계 방식을 표현한 것이다. 〈그림 16〉부터 〈그림 18〉까지 제시된 '미래창조과학부(MSIP_11)'와 관련된 관계들은 〈그림 19〉의 이벤트들을 통해 정확하게 개념이 전달되고 공유된다.



〈그림 18〉 '미래창조과학부(MSIP_11)' 인스턴스와 관련된 인스턴스



〈그림 19〉 '미래창조과학부(MSIP_11)' 인스턴스와 관련된 이벤트

본 연구에서는 시소러스를 중심으로 한 온톨로지 모형 변환을 설계하고 구축하였으며, 실제 데이터를 통해 활용가능성을 확인하였다. 또한 ISO 25964-2에서 강조된 대등관계 매핑을 복잡한 개념 구조에 적용하였으며, 이를 체계적으로 분리하는 과정을 수행하였다.

5. 결론

본 연구에서는 시소러스의 구축과 활용, 상호운용성 확보에 관한 최근의 국제 동향을 파악하고 실제 활용가능성을 확인하였다. 시소러스 구축과 활용에 대한 국제 동향으로는 ISO 25964의 개정을 들 수 있으며, 본 연구에서는 제2부에 초점을 두었다.

ISO 25964-2에서는 기존에 정의된 시소러스 관계를 큰 틀 안에서 유지하면서, 대등관계 매핑을 좀 더 확장하였다. 시소러스의 기본적인 관계 표현은 제1부에서 제시하였으며, 제2부는 1부를 토대로 다른 체계와의 상호운용성을 위한 매핑 방식을 확장하였으며, 대등관계 매핑에 집중하였다. 대등관계 매핑(EQ)은 대등관계의 정도를 표시한 것으로, 정확한(exact: =EQ), 부정확한(inexact: ~EQ), 부분적인(part: ~EQ) 대등관계 매핑 유형으로 구별되었다. 그리고 개념 공유를 위해 불리언 연산자의 개념을 차용하여 검색에 활용되는 어휘들의 조합을 설명하였다. 특정 어휘가 둘 개 이상의 어휘의 합으로 표현되어야 하면 'EQ [A] | [B]'를 활용하고, 어휘의 교집합으로 표현되어야 하면 'EQ [A] + [B]'를 사용하는 것을 강조하였다.

본 연구에서는 ISO 25964-2의 상호운용성에

초점을 두고 시소러스 관계와 온톨로지 변환 시 고려해야 할 사항들을 확인하였다. ISO 25964-2에서는 상호운용성을 적용해야 할 다양한 어휘 집의 유형으로 시소러스뿐만 아니라 분류체계, 기록관리 분류체계, 택소노미, 주제표목표, 온톨로지, 용어집, 이름전거리스트, 동의어 리스트 등을 포함시켰다. 그 중 온톨로지 모형은 현재의 시맨틱 웹 정보환경을 수용하며, 개념의 의미를 공유하는 지식 틀로 많은 강점을 가지고 있다.

시소러스와 온톨로지와의 변환 모형은 다음과 같은 특징을 갖는다. 첫 번째는 시소러스와 온톨로지 모형 자체의 차이점을 분석하여, 그것들의 강점을 파악하는 것(ISO, 2013, p. 76)이었다. ① 시소러스의 개념과 온톨로지의 클래스에 대한 차이를 확인하였다. 시소러스는 개념의 의미를 용어로 표현하고, 관계 개념, 한정어, 범위주기, 정의 등을 정보검색 이용자에게 제공하는 반면, 온톨로지는 클래스의 의미를 기계가 이해할 수 있는 정보환경에서 전달하였다. ② 시소러스는 정보검색에 초점을 맞추어, 우선어 및 비우선어를 표시하는데, 온톨로지는 클래스와 인스턴스의 명칭을 통해 표현하였다. ③ 시소러스에서는 자연언어를 정확하게 표현하기 위해 온톨로지보다 더 많은 노력이 요구되었다.

두 번째는 온톨로지 모형이 이용자 검색 요구를 수용하는 유연한 지식 체계라는 것이었다. 이용자의 검색 내비게이션 방식을 잘 이해하고 이를 반영한 정보 체계를 구축할 수 있는 기반 기술을 제공하였다. 세 번째는 온톨로지 모형의 구성요소를 통해 특정 주제 영역의 지식 구조를 완성하는 것이었다. 네 번째는 지식시스템과의 연계가 가능하다는 것이었다. ISO 25964에서는

시소러스를 일종의 지식조직시스템(Knowledge Organization System, KOS)으로 간주하고, 시맨틱 웹 환경에 적용 가능하도록 설계하도록 하였다. 온톨로지 모형이 이를 가능하게 하는 강력한 장치로 활용될 수 있음을 본 연구를 통해 확인하였다.

본 연구에서는 매핑 유형을 고려한 시소러스와 온톨로지 간의 상호운용성 모형을 설계하고, 이에 대한 온톨로지 클래스와 속성을 선언하였다. 온톨로지는 특정 주제 영역을 중심으로 전개되는 정보표현 방식이므로, 본 연구에서도 특정 주제 영역을 선정하였다. 본 연구에서 온톨

로지 변환 모형의 효용성을 확인하기 위해 활용한 데이터는 행정부처 시소러스였다. 행정부처의 명칭을 관리하기 위한 클래스, 속성을 선언하고, 인스턴스 입력을 통해 모형의 구체성 및 상세성을 확인하였다.

ISO 25964-2를 구체적으로 적용한 국내 연구가 없는 상황에서 본 연구는 전자적인 환경에서 시소러스의 범위를 확장하고, 재설계하는데 도움이 되는 토대 연구로 활용될 수 있을 것이다. 앞으로의 지속적인 연구를 통해 좀 더 다양한 어휘집과의 연결을 시도가 필요가 있겠다.

참 고 문 헌

- 국가기록원 홈페이지. www.archives.go.kr
- 김성원, 김정우 (2011). 통제어휘 표준의 변화 및 시사점에 대한 연구. *한국문헌정보학회지*, 45(1), 211-232.
- 김태수 (2001). 용어 정의를 도입한 시소러스 개발 연구. *정보관리학회지*, 18(2), 231-254.
- 박옥남 (2011). ISO 25964를 통해 살펴본 시소러스 구축 표준에 관한 연구. *정보관리연구*, 42(2), 43-65.
- 박지영, 윤소영, 이혜원 (2017). 기록시소러스 구축지침 개정에 관한 연구. *한국기록관리학회지*, 17(1), 117-141.
- Brownson, H. (1957). Current United States research on topics related to the conference. In *Proceedings of the International Study Conference on Classification for Information Retrieval 13-17 May 1957, Dorking, England.* (pp. 99-100). London: ASLIB.
- Cardillo, E., Folino, A., Trunfio, R., & Guarasci, R. (2014). Towards the reuse of standardized thesauri into ontologies. Retrieved July 2, 2017, from http://ontologydesignpatterns.org/wiki/images/0/03/Paper_8.pdf
- Clarke, Stella G. Dextre (2011, July). ISO 25964: a standard in support of KOS interoperability. (Presentation slides) In: Gilchrist, Alan and Vernau, Judi, Editors. *Facets of Knowledge Organization*, London. Retrieved July 1, 2017, from <http://www.iskouk.org/conf2011/papers/dextreclarke.pdf>

- ISO (2011). ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval.
- ISO (2013). ISO 25964-2:2013 Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies.
- Järvelin, K., Kristensen, J., Niemi, T., Sormunen, E., & Keskustalo, H. (1996). A deductive data model for query expansion. In Proceedings of the 19th Annual International ACM - SIGIR Conference on Research and Development in Information Retrieval (pp. 235-243). New York, NY: ACM Press. Retrieved July 3, 2017, from <http://www.webcitation.org/5m2kxJw2E>
- Mader, C., Haslhofer, B., & Isaac, A. (2012). Finding Quality Issues in SKOS Vocabularies. International Conference on Theory and Practice of Digital Libraries 2012, 222-233.
- Pastor-Sanchez, J.-A. (2015, september). Elaboration of Controlled Vocabularies Using SKOS. Paper presented at the DCMI Global Meetings & Conferences, São Paulo. Retrieved July 5, 2017, from <http://dcevents.dublincore.org/IntConf/dc-2015/paper/view/404>
- Pastor-Sanchez, J.-A., Mendez, F. J. M., & Rodríguez-Muñoz, J. V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. Information Research, 14(4). Retrieved July 8, 2017, from <http://www.informationr.net/ir/14-4/paper422.html>
- Roche, C., Damas, L., & Roche, J. (2014, september). Multilingual Thesaurus: The Ontoterminology Approach. Paper presented at the CIDOC, Dresden. Retrieved July 8, 2017, from <https://hal.archives-ouvertes.fr/hal-01356631/document>
- W3C (2009). SKOS Simple Knowledge Organization System. Retrieved July 2, 2017, from <https://www.w3.org/TR/skos-reference/>
- Will, L. (2012). The ISO 25964 data model for the structure of an information retrieval thesaurus. Bulletin of Association for Information Science and Technology, 38(4), 48-51. Retrieved July 3, 2017, from https://www.asis.org/Bulletin/Apr-12/AprMay12_Will.pdf

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Kim, Sungwon, & Kim, Jeongwoo (2011). A Study on the Changes in Standards Related to Controlled Vocabulary and Their Implications. Korean Society for Library and Information Science, 45(1), 211-232.

Kim, Tae-Soo (2001). A Study on the Changes in Standards Related to Controlled Vocabulary and Their Implications. Korea Society for Information Management, 18(2), 231-254.

National Archive of Korea Homepage. www.archives.go.kr

Park, Ok-Nam (2011). The Study on the Standards of Thesaurus Construction Based on ISO 25964. Journal of information management, 42(2), 43-65.

Park, Zi-young, Yoon, SoYoung, & Lee, Hyewon (2017). A Study on the Revision Archival Thesaurus Construction. Korean Society of Archives & Records Management, 17(1), 117-141.