

# 시소러스를 활용한 온톨로지 구축방안 연구\*

- 시소러스의 SKOS 변환을 중심으로 -

## A Study of Ontology Construction Using Thesaurus: Transformation of Thesaurus into SKOS

한 성 국(Sung-Kook Han)\*\*

이 현 실(Hyun-Sil Lee)\*\*\*

### 초 록

본 연구에서는 시소러스를 SKOS 기반 온톨로지 구축하는 단계적 변환 방법을 제시하였다. 변환과정을 3단계로 형식화하였고, 각 단계별 수행작업, 산출물 및 가이드 라인을 개발하였다. 각 단계는 (1) 시소러스의 어휘 용어 체계와 의미 관계를 이해하기 위해 시소러스의 기초자료를 수집하고 분석한다. (2) 대상온톨로지의 분석결과를 활용하여 변환방식을 정의하고 SKOS 온톨로지를 생성한다. (3) 대상시소러스와 생성된 SKOS 온톨로지 사이의 형태와 의미 구조 보존 여부를 검사하여 보정하고 문서화한다. 본 연구에서는 제시하고 있는 시소러스 기반형 온톨로지 구축 방법은 복잡한 어휘 관계를 갖는 시소러스에도 적용이 가능하다. 향후, 본 연구에서 제시한 단계별 변환 알고리즘을 프로그램으로 구현하여 시소러스의 자동 SKOS 온톨로지 변환기를 구현하여야 할 것이다.

### ABSTRACT

This study suggests the method of converting thesauri to SKOS step by step and it is formalized in three stages of the conversion process. The study develops output and guidelines for each stage. The converting stages are: (1) Collecting and analyzing thesauri for understanding about structure of terms and semantics of relation, (2) Defining the conversion method and creating ontology of the thesauri, (3) Examining the preservation of forms and various semantic relations between the thesauri and then creating SKOS ontology. This method can be applied to the thesauruses with complicated relations in concepts. In the future, it is needed to have an embodiment of conversion after making the algorithm of conversion by stage with the method suggested in this research.

키워드: 시소러스, SKOS 온톨로지, 의미관계, 도메인  
thesaurus, SKOS, ontology, semantic relation, domain

\* 본 논문은 2005학년도 원광대학교의 교비 지원에 의해서 수행됨.

\*\* 원광대학교 컴퓨터학과 교수 (skhan@wku.ac.kr)

\*\*\* 원광대학교 도서관 사서 (hyunsil@wku.ac.kr)

논문접수일자 : 2006년 5월 22일 논문심사일자 : 2006년 5월 28일 게재확정일자 : 2006년 6월 1일

## 1. 서론

최근 구현되고 있는 정보시스템은 대부분이 지식처리 기능을 목표로 하고 있다. 도메인의 개념 모델 방법을 제공하는 온톨로지에 대한 관심이 고조되고 있을 뿐 만 아니라, 도메인 온톨로지 구축이 당면 과제로 대두되고 있다. 도메인 온톨로지는 도메인에 내재한 개념요소의 명시적 정의나 개념간의 의미 관계 및 도메인의 공리체계를 정의하여 지식처리 시스템의 기반을 제공한다.

각 분야에서 도메인 온톨로지 구축이 활발하게 추진되고 있지만, 도메인 온톨로지 구축은 많은 인력과 자원이 소요되는 장기간의 개발 과제이고, 구축된 온톨로지의 질적 수준에 문제가 제기되기도 한다. 대표적인 의학 온톨로지인 GALEN(Open Galen 2005)은 1980년 후반에 개발이 착수되었지만 현재까지도 지속적으로 수정 보완되고 있는 것을 볼 때, 질적 수준이 보장되는 온톨로지 개발의 어려움을 알 수 있다. 때문에 기존의 정보자원 모델링 기술과 통제 어휘집, 시소러스 등과 같이 축적된 자원을 활용한 온톨로지 구축이 하나의 대안(Matthews, Brian, 2003)이 되고 있다.

그동안 색인 용어의 통제와 정보 검색의 효율성 등을 위하여 많은 시소러스가 개발되어 왔는데, 시소러스는 도메인을 구성하는 용어와 이들 사이의 의미 관계를 구조화 한 도메인 개념 모델링의 일종이다. 시소러스 구축을 위해서 도메인 전문가들은 장기간에 걸쳐 도메인 어휘를 수집하고 의미 관계를 분석하여 왔다. 시소러스의 유용성이 확대되면서 이러한 작업을 효과적으로 수행할 수 있도록 표준적인 개

발지침이 필요하게 되었고 ISO 2788:1966과 ANSI/NISO Z39.19 등이 제정되었다. 시소러스가 도메인의 개념 모델링의 일종이지만, BT, NT, RT 등 단순 의미 관계만을 제공하고 있기 때문에 상세한 도메인 개념 구조의 표현에는 한계가 있다. 이러한 한계를 극복하고자 시소러스의 용어 관계 확장에 대한 연구가 많이 이루어졌다(한상길 1999; 고영만 2006).

한편, 시소러스가 도메인의 핵심 어휘와 이들 간의 의미 관계를 표현할 수 있는 도메인 모델의 성격을 갖는다는 관점에서 시소러스를 활용하여 도메인 온톨로지 구축하기 위한 연구도 활발히 진행되어 왔다(Matthews 2003; Assem et al. 2004; Wielinga 2006). 이러한 연구의 대부분은 시소러스를 RDF/S 또는 OWL 등 온톨로지 언어로 변환하는 방법을 제시하고 있다. 하지만 이러한 방법은 특정 시소러스를 대상으로 하고 있기 때문에, 제시된 방법을 일반화하여 다른 시소러스에 적용하는데 어려움이 있다. 또한, 온톨로지로 변환시에 시소러스 고유의 형태적 특성 유지에도 문제가 있다. W3C에서는 시소러스를 활용한 온톨로지 구축의 중요성이 부각됨에 따라, 시소러스의 형태적 상호운용성을 보장하고, 온톨로지의 다양한 의미관계를 활용할 수 있도록, 통제어, 분류, 시소러스 등 기존의 용어 체계를 포괄할 수 있는 RDF 기반의 표준 프레임워크인 SKOS(Simple Knowledge Organization System)를 개발하였다. 이로써 시소러스를 활용한 경량 온톨로지(light-weight ontology) 구축에 대한 연구가 활성화되는 계기가 마련되었다(Assem 2006).

SKOS는 기존의 시소러스가 당면한 문제를 해결하고 온톨로지화 함으로써, 효과적으로 도

메인 온톨로지를 구축하여 지식 기반 시스템을 실현할 수 있는 기반을 제공하고 있지만 시소러스는 고유한 도메인 개념 모델과 표현 방식을 갖고 있기 때문에 SKOS를 변환하는 것이 용이하지 않다. SKOS는 범용적인 프레임워크를 제시하고 있기 때문에 개별 시소러스의 특성을 정확하게 분석하여 SKOS로 대응시켜야 한다.

이에 본 연구에서는 기존의 시소러스를 SKOS로 변환하는 단계적 방법을 제안하여 시소러스 기반 온톨로지를 구축할 수 있는 방안을 마련하고자 한다. 시소러스 모델링을 분석하여 명세화하는 방법을 제시하고, 기본 관계인 BT, NT, RT 뿐 만 아니라 용어관계의 확장에 의해서 생성될 수 있는 다양한 의미 관계(한상길 1999; 고영만 2006)를 포괄할 수 있는 시소러스 기반 온톨로지 구축 방법을 제시한다. 본 논문은 한국역사 용어 시소러스를 대상으로 서술하였지만, 확장된 시소러스 관계를 갖은 시소러스에도 적용이 가능하다.

## 2. 이론적 배경

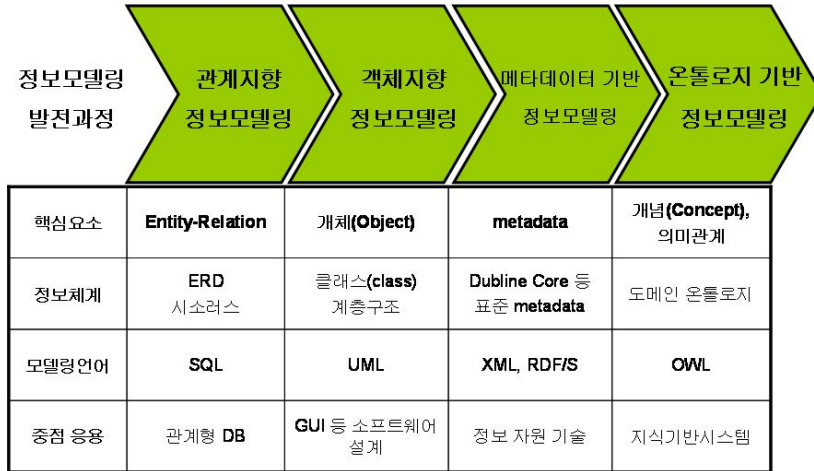
### 2.1 정보모델링의 변천

통제어와 같은 용어 체계는 키워드 검색이나 색인 등에 활용될 수 있지만, 용어간의 관계 또는 의미를 기반으로 하는 다양한 형태 정보 검색 등에 활용할 수가 없다. 효과적인 정보의 관리와 검색을 위해서는 용어 또는 개념 요소간의 의미 관계가 표현되는 정보 모델링이 이루어져야 한다.

정보 모델링은 현실세계의 실체를 정적 특성(사건)과 동적 특성(행동)을 중심으로 사고하고 표현하는데 필요한 수학적으로 잘 정의된 개념들의 집합이다. 정적 특성은 스키마(schema)에 정의되고 동적 특성은 기능 명세(Specification)로 정의되고 있으며, 정보 모델은 기본적으로 객체(object), 연산(operation), 제한(constraints)의 개념을 갖는다(Brodie, Mylopoulos, and Schmidt 1984).

정보 모델링은 정보의 뜻을 파악하고 유용하게 조직하는 컴퓨터 기반의 구조적 표현으로 도메인을 구성하는 정보 개체를 식별하고 이들의 상호 관계를 컴퓨터 처리가 가능한 형태로 구조화한다. 컴퓨터 시스템 구축의 토대를 제공하는 정보 모델링의 중요성으로 인하여, 그동안 다양한 정보 모델링 기법 개발에 많은 연구가 이루어져 왔고, 이러한 정보 모델링 기법을 요약하면 그림 1과 같다.

그림 1에서 보면, 정보 모델링은 정보 요소와 요소간의 단순 관계를 표현하는 관계 지향 모델링에서 정보 요소의 개념 구조와 의미 관계를 표현하는 온톨로지 기반 모델링으로 발전하고 있으며, 정보 자원의 의미구조와 도메인 지식을 처리할 수 있는 시스템 구현을 목표로 하고 있다. 최근 이러한 목표를 달성하는데 필요한 표준 언어와 개발 도구들이 다양하게 지원되고 있어 효과적으로 지식 기반 시스템을 구현할 수 있는 개발 환경이 마련되어 있다. 따라서 도서관 정보시스템도 발전된 정보 기술을 능동적으로 활용하여야 하며, 정보시스템을 지식 기반 시스템으로 구현하여 처리 능력과 수준을 향상시켜야 한다.



(그림 1) 정보 모델링의 발전 과정

## 2.2 시소러스와 온톨로지

시소러스는 특정 도메인의 어휘 정보 모델링으로써, 정보의 색인과 검색 과정에서 디스크립터와 검색어를 선정하기 위한 도구로 널리 사용되어 왔으며, 시소러스를 활용함으로써 도메인의 어휘 용어를 통제하고 정보검색의 정확도를 높일 수 있다. 방대한 도메인 어휘체계를 구축하고 있는 시소러스는 BT, NT, RT, UF 와 같은 어휘간의 중요한 관계를 정의하고 있다.

정보 검색 시스템에서 색인 용어를 통제하고 검색 효율성을 높여주는 도구으로써 시소러스의 중요성이 크게 증대됨에 따라, 분야별로 많은 시소러스들이 개발되었다. 시소러스는 단순히 색인과 검색에서 통제 역할만을 하는 것이 아니라, 개념의 확장과 축소를 통해서 도메인의 개념 체계 표현으로도 활용된다. 시소러스에 대한 표준 기술 체계가 요구됨에 따라 관련 표준안도 함께 마련되어, 용어 중심의 시소러스 표준 기술 체계인 ISO 2788:1986이 개발되었

고, 이를 확장하여 개념을 표시하는 어휘 레이블(lexical label)로써의 용어를 정의한 ANSI/NISO Z39.19가 개발되었다.

하지만 현재 정보시스템에 탑재되어 도메인 정보검색에 사용되는 시소러스는 그 구조가 유연하지 않아 언어적 변화를 수용하는데 있어서 어려움이 있다(Schmitz-Esser 1991). 또한 시소러스는 중요성과 유용성에 비하여 의미 관계 표현의 단순성과 한계가 있다. 시소러스는 용어간에 의미 구조의 불명확성, 개념에 대한 속성 파악의 어려움, 의미 검색 및 추론 기능의 부재 등으로 인하여 정보시스템에 지식처리의 기능을 제공하고 있지 못하는 것이다(고영만 2005).

반면에 인공지능과 시맨틱웹에서 지식 표현을 위하여 널리 활용하고 있는 온톨로지는 도메인 지식에 대한 공유된 개념화의 형식적 명세 체계로써, 도메인에 내재한 개념 요소의 의미 특성과 개념간의 다양한 의미 관계를 표현하며, 공유 및 재사용과 의미 추론을 통하여 지식 처리 기능을 제공한다. 온톨로지는 시맨틱

웹의 확산과 더불어 도메인 지식 체계를 명세화하고 의미적 상호운용성을 실현하는 도구로 알려지고 있다. 온톨로지는 지식 기반 시스템의 지식 베이스를 구축하는 기본 정보 모델로 활용되고 있으며, 지식 추론 등의 다양한 지식 처리 기능을 소프트웨어 에이전트에게 제공한다. 때문에, 온톨로지는 시맨틱웹 뿐만 아니라, 지식 기반 시스템의 핵심요소로 인식되고 있다.

도메인의 개념구조를 명세화하는 과정에서, 시소러스와 온톨로지는 유사성이 있고, 시소러스가 경량 온톨로지로 간주되기도 한다. 시소러스와 온톨로지는 도메인의 개념 요소에 대한 공유된 이해 체계를 제공하는 개념 기반 표현 방식이다. 시소러스가 정보 검색 및 색인에 필요한 용어 체계에서 출발한 반면에, 온톨로지는 도메인 지식 체계의 명시적 표현을 목적으로 하고 있는 점에서 차이가 있다. 시소러스와 온톨로지의 특성을 비교하면 표 1과 같다.

표 1에서와 같이, 시소러스는 어휘 중심의 단순 의미 관계 형성을 지원하고 있고, 온톨로지는 논리 형식 체계를 기반으로 한 지식 표현 방식임을 알 수 있다. 시소러스가 정보 검색 및

색인에 유용하지만, 질의 및 논리 추론을 위해서는 온톨로지가 필요하다. 즉 정보시스템이 지식베이스로 발전하려면 온톨로지 구축이 선행되어야 할 것이고, 시맨틱웹 등의 기술을 활용하기 위해서는 시소러스 온톨로지로 변환할 필요가 있다.

### 3. 시소러스의 온톨로지 변환

#### 3.1 시소러스 변환

시소러스는 특정 도메인 어휘 정보를 모델링하고 있고 표준 형식으로 개발되어 있어 이를 온톨로지로 변환해 도메인 온톨로지를 용이하게 구축할 수 있다. W3C에서는 시소러스의 온톨로지 변환을 지원하기 위해, 시소러스의 온톨로지 모델링인 SKOS를 제안하고 있다 (Wielinga et al. 2001). 때문에 시소러스의 표준 기술 체계를 SKOS로 변환할 수 있는 효과적인 절차와 알고리즘을 개발하여 단 기간에 도메인 온톨로지를 구축하는 것이 가능하다.

(표 1) 시소러스와 온톨로지의 특성 비교

구분	시소러스	온톨로지
공통점	<ul style="list-style-type: none"> <li>▪ 도메인의 개념 요소에 대한 공유된 이해 체계</li> <li>▪ 도메인 지식의 개념 기반 표현 방식</li> <li>▪ 어휘 간의 계층 구조 및 의미 관계의 네트워크 형 표현</li> <li>▪ 개념 기술을 위한 어휘 체계 이용</li> </ul>	
어휘요소	통제어	개념 어휘(통제어 불필요)
어휘의미표현	계약·한정된 표현(RT 등)	다양한 의미 표현
형식체계	형식체계 없음	FOL, DL 등 논리 형식체계 기반
지식표현	제한적 또는 어려움	명시적이고 형식적인 지식표현
처리방법	질의 처리	질의 및 논리 추론
추론기능	없음	지식 추론
활용	정보검색 및 색인	지식기반 시스템 구축

시소러스는 BT, NT, RT 등 용어간의 단순 관계만을 표시하고 있어 도메인의 개념 구조와 다양하고 심층적인 개념의미 관계를 표현하는 도메인의 중량 온톨로지 구축에는 한계가 있다. 이를 해결하기 위해서는 시소러스 용어간의 관계를 세분화할 필요가 있으며, 이를 표현할 수 있는 표준 기술 체계가 필요하다. 또한 시소러스를 온톨로지로 변환할 경우 의미 관계의 충돌 또는 불필요한 관계의 첨가 등이 발생하므로 변환된 온톨로지를 사후 보정해야 한다. 시소러스 변환으로 생성된 온톨로지는 시소러스의 구문·의미 구조를 보존하여야 하며, 생성된 온톨로지 간의 호환성을 유지해야 하기 때문에 이러한 목적을 달성하려면 SKOS와 같은 시소러스 표준 메타 모델로 시소러스를 변환할 필요가 있다.

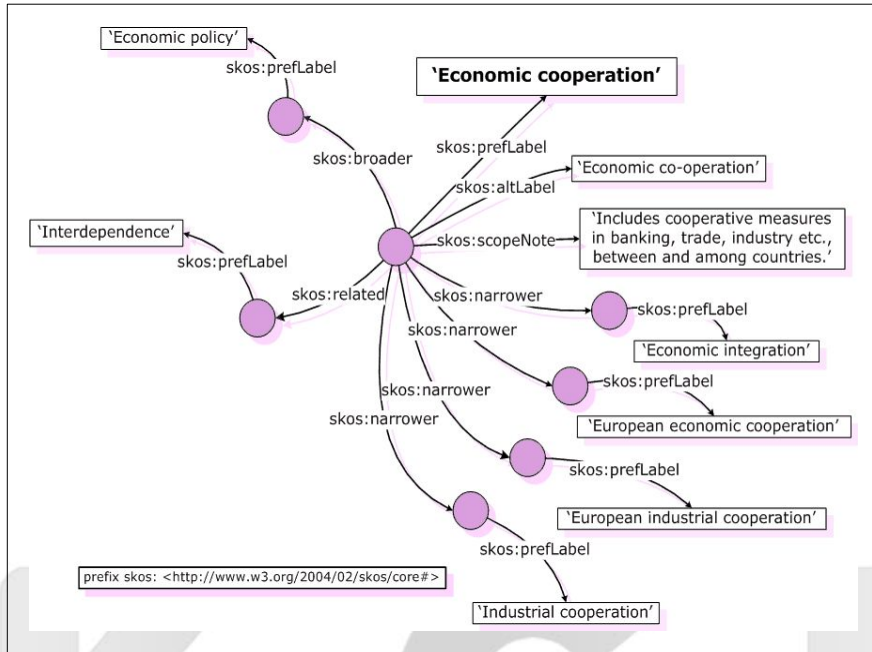
### 3.2 SKOS 개요

SKOS(Simple Knowledge Organization System)는 W3C에서 시맨틱웹 기반의 지식 구조화 프레임워크 표준으로 개발하여 2005년 발표하였다. SKOS는 RDF 기반의 프레임워크로서 분산 정보 처리 환경에서 유연하고 확장이 가능하며, 공유와 재사용할 수 있는 다양한 지식 구조화 방법을 제공한다. SKOS는 통제어, 용어, 시소러스, 텍소노미, 주제표목 등을 모델화 하는데 적합한 기능을 갖고 있다.

SKOS Core Vocabulary는 RDF 클래스와 프로퍼티로 구성되며, RDF 그래프를 활용해서 개념과 개념 구조를 정의 할 수 있다. 예를 들어, UK Archival Thesaurus(UKAT 2004)의 일부인 그림 2의 시소러스 용어는 그림 3과 같은 RDF 그래프로 표현이 가능하고, SKOS를 활용해서 그림 4와 같이 기술할 수 있다

Term: Economic cooperation
Used For: Economic co-operation
Broader terms: Economic policy
Narrower terms: Economic integration European economic cooperation European industrial cooperation Industrial cooperation
Related terms: Interdependence
Scope Note: Includes cooperative measures in banking, trade, industry etc., between and among countries.

(그림 2) UKAT 시소러스의 예



(그림 3) UKAT 시소러스 용어의 RDF 그래프 표현

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.ukat.org.uk/thesaurus/concept/1750">
    <skos:prefLabel>Economic cooperation</skos:prefLabel>
    <skos:altLabel>Economic co-operation</skos:altLabel>
    <skos:scopeNote>Includes cooperative measures in banking, trade, industry etc.,
      between and among countries.
    </skos:scopeNote>
    <skos:inScheme rdf:resource="http://www.ukat.org.uk/thesaurus"/>
    <skos:broader rdf:resource="http://www.ukat.org.uk/thesaurus/concept/4382"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/2108"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/9505"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/15053"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/18987"/>
    <skos:related rdf:resource="http://www.ukat.org.uk/thesaurus/concept/3250"/>
  </skos:Concept>
</rdf:RDF>
  
```

(그림 4) UKAT 시소러스 용어의 SKOS 표현

시소러스 정보 표현의 유용한 기능을 갖고 있는 SKOS의 중요한 특성은 다음과 같다.

① 레이블링 (labelling) 특성

레이블은 주로 사용자의 편의를 제공하기 위한 것으로 정보 자원을 지시하기 위해 부여된 명칭 또는 기호와 같은 것이다. SKOS에는 skos:prefLabel, skos:altLabel, skos:hiddenLabel, skos:prefSymbol, skos:altSymbol 과 같은 다양한 레이블링 방법을 제시하고 있다. 예를 들어, 정보 자원 ex:animals에 대하여 그림 5와 같이 레이블링 할 수 있다.

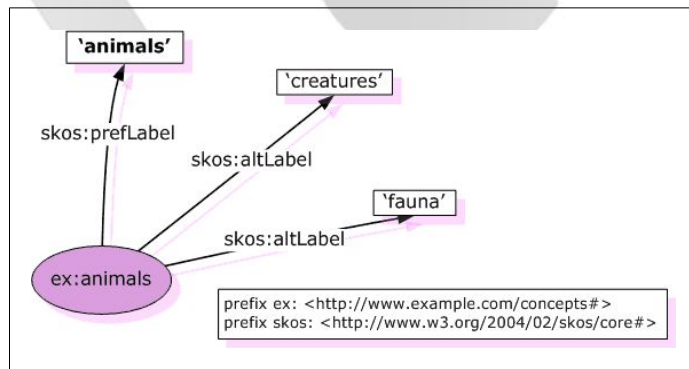
또한 SKOS에서는 다국어 레이블링, 기호에 의한 레이블링 기능도 제공하고 있는데 다국어 레이블링 기능을 활용하여, 국내 개발 시소러스들이 포함하고 있는 한국어, 영어, 중국어 등

의 용어를 표현할 수 있다.

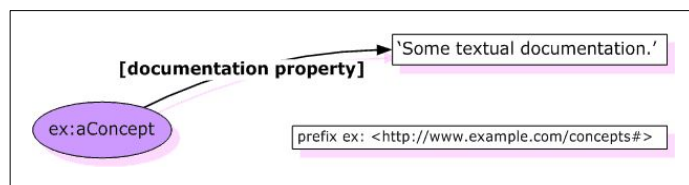
② 문서화 특성

SKOS에서는 작성된 문서를 효과적으로 관리할 수 있도록 문서화 기능을 제공하고 있다. 대표적인 것으로 skos:definition, skos:scopeNote가 있으며 이들의 상위 속성으로 skos:note가 정의 되어 있다. 이러한 문서화 속성은 SKOS 문서 관리의 편리성을 제공하는 한편, 다양한 형태의 문서화 방법이 가능하도록 하고 있다 (Mikhailenko 2005).

SKOS에서 제공하는 문서화의 종류를 요약하면 첫째, RDF 리터럴(literal)로서의 문서, 둘째, 관련 정보 자원 기술로서의 문서, 셋째, 문서 참조로서의 문서로 구분하여 정의 할 수 있다. 그림 6은 리터럴로서의 문서의 예를 보인 것이다.



(그림 5) 레이블링 예



(그림 6) 리터럴로서의 문서의 예

③ 의미 관계 표현

SKOS는 시소러스의 기본 의미 관계인 BT, NT, RT에 대한 표현 방식을 제공하고 있다. 이를 위해 skos:broader, skos:narrower, skos:related를 정의하고 있으며, 이들 속성은 skos:semanticRelation의 하위 속성이다. 그림 7에 의미 관계 표현과 관련된 SKOS 기술을 보였다.

skos:broader와 skos:narrower 사이에는 역관계(inverse relation)가 있으며, 각자는 추이율 관계(transitive relation)가 정의되어 간단한 추론을 수행할 수 있다.

④ 주제 색인 특성

주제 색인 기능을 제공하기 위해서는 skos:subject, skos:isSubjectOf, skos:PrimarySubject, skos:isPrimarySubjectOf가 제공되고 있다. skos:subject와 skos:isSubjectOf는 서로 역관계가 있으며, skos:subject와 skos:broader 사이에는 주제 일반성 규칙(subject generality rule)이 적용되고 있다. 주제 색인의 예는 그림

8과 같다.

⑤ 개념 그룹 특성

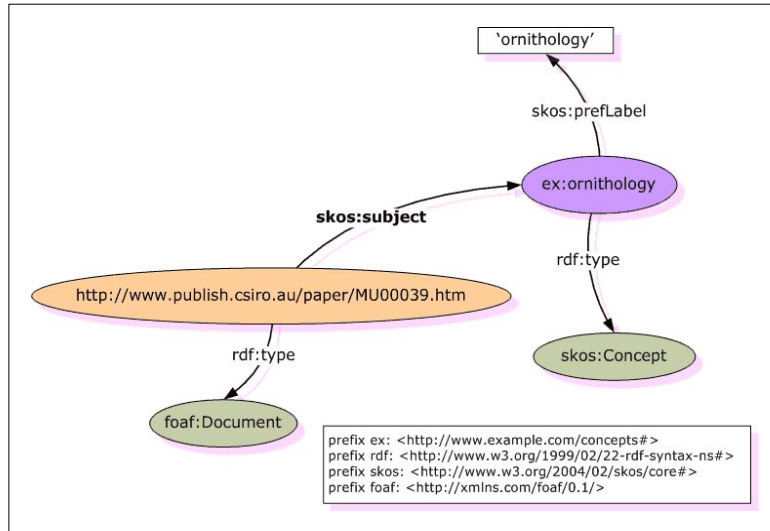
SKOS에는 개념을 그룹화 할 수 있는 기능을 제공하고 있다. 예를 들어, buffalo milk, cow milk, goat milk, sheep milk와 같은 개념들을 묶어서 <milk by source animal>로 정의 할 수 있다. 이와 같은 기능은 SKOS의 기반인 RDF의 rdf:Bag, rdf:Seq 등을 유사한 기능이다. 그림 9에서 이 같은 내포된 그룹화(nested collection)의 예를 보였고, 이외에도 여러 그룹화 기능을 제공한다.

## 4. 시소러스의 온톨로지 구현

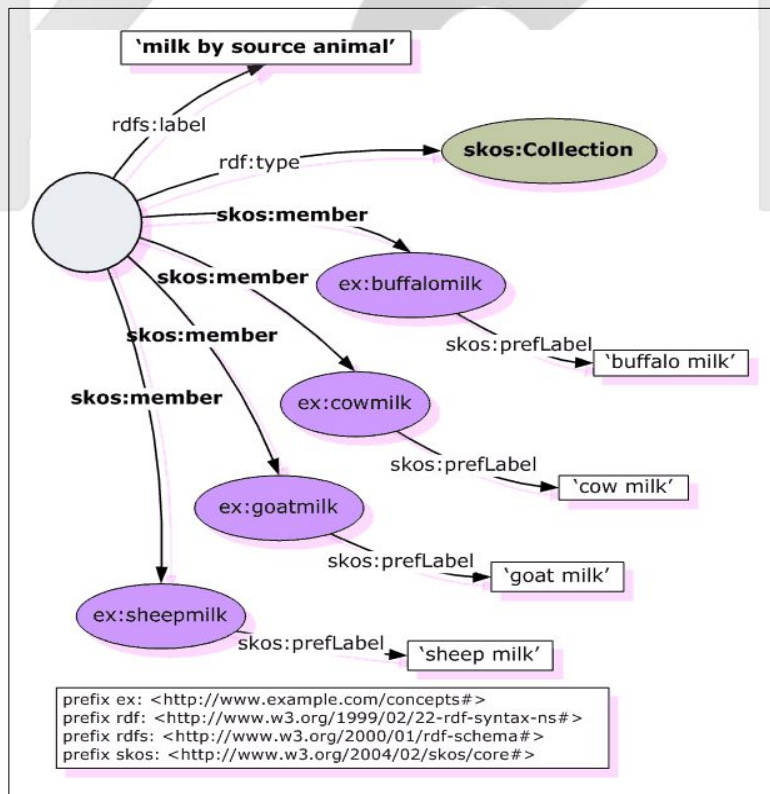
본 연구에서는 그림 10과 같이 시소러스의 표준 SKOS 변환의 단계별 전략을 수립하였다. 각 단계별로 독립된 수행 작업을 정의하고 산출물과 가이드라인을 정하여 체계적인 개발이

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.example.com/concepts#mammals">
    <skos:prefLabel>mammals</skos:prefLabel>
    <skos:broader rdf:resource="http://www.example.com/concepts#animals"/>
  </skos:Concept>
  <skos:Concept rdf:about="http://www.example.com/concepts#animals">
    <skos:prefLabel>animals</skos:prefLabel>
    <skos:narrower rdf:resource="http://www.example.com/concepts#mammals"/>
  </skos:Concept>
</rdf:RDF>
```

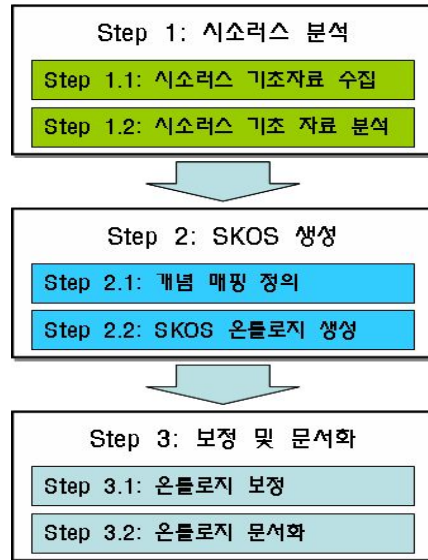
(그림 7) 의미 관계 표현의 SKOS의 표현



(그림 8) 주제 색인의 예



(그림 9) 그룹화 기능의 예



(그림 10) 시소러스의 SKOS 변환 절차

이루어지도록 하였다.

#### 4.1 시소러스의 분석

시소러스 분석단계(Step 1)는 온톨로지 변환 준비 단계로 대상 시소러스의 어휘 용어 체계와 의미 관계를 이해하고 모델화하는 단계이다. 더불어 구축되어질 온톨로지에 대한 요구 사항을 정리한다. 시소러스 분석 단계는 수행 작업에 따라 다음과 같이 두 단계로 나누어진다.

##### 1) 시소러스 기초 자료 수집(Step 1.1)

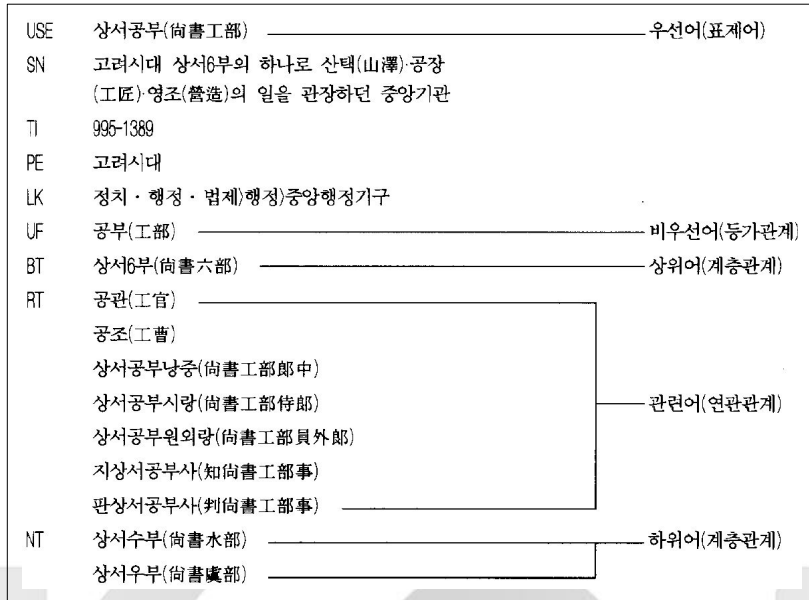
시소러스의 구조와 특성을 전문적으로 이해하기 위해, 시소러스와 관련된 자료를 수집한다. 수집할 자료는 (1) 시소러스 개발자와 면담, (2) 시소러스 설계 및 개념 모델, (3) 시소

러스 구현 형태, (4) 시소러스 응용 사례 및 활용 도구 등으로 시소러스의 어휘 모델을 이해하는데 필요한 자료이다.

그림 11에 예시한 한국역사용어 시소러스(국사편찬위원회 2005)에서는 TI, PE, LK 등 색인 또는 분류 목적의 어휘 요소가 추가되어 있지만, 국내에서 개발된 대부분의 시소러스는 BT, NT, RT 관계를 중심으로 구축되어 있기 때문에 시소러스 어휘 모델이 단순하다. LIMBER<sup>1)</sup> 프로젝트 등 외국에서 개발된 많은 시소러스는 복잡한 어휘 모델을 기반으로 하고 있는 경우가 많이 있다. 시소러스의 어휘 모델은 온톨로지 변환에 중요한 역할을 함으로 명시적인 모델 설정이 필요하다.

수집된 시소러스 자료의 검토와 함께 구축할 온톨로지의 목적, 용도 등을 고려하여 온톨로

1) LIMBER는 메타데이터와 다언어 시소러스를 이용하여 웹상에 분산되어 있는 데이터에 다언어 접근을 지원하기 위한 개발 도구이다. LIMBER에서는 RDF언어를 이용하여 메타데이터와 다언어 시소러스가 정의된다



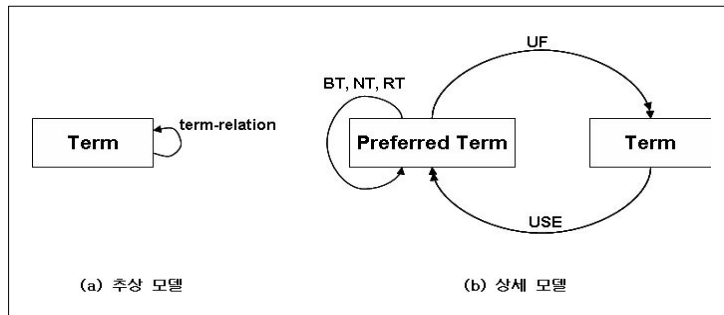
(그림 11) 한국역사용어 시소러스의 기본 구조의 예( 류주희 2004)

지 구축 요구 사항을 작성한다. 이러한 과정을 통해 시소러스 개요 및 특징 기술서, 시소러스 어휘 용어 모델 설명서, 시소러스 온톨로지의 요구 사항 기술서 등의 산출물을 작성한다.

2) 시소러스 기초 자료 분석(Step 1.2)  
시소러스의 개념 모델을 분석하는 작업을 수

행하는 단계이다. step 1.1의 내용과 산출물을 기반으로 시소러스의 개념 모델을 정립한다. 일반적으로 시소러스의 개념 모델에는 용어 중심 모델과 개념 중심 모델의 두 형식이 있다.

용어 중심 모델은 그림 12처럼 용어 간의 관계를 기반으로 시소러스가 구축된 형태이다. 용어 중심 모델은 용어 및 복합 용어(compound



(그림 12) 용어 중심 시소러스 모델

term), BT, NT, RT, USE, UF(Use For) 등 용어간의 관계 등 시소러스 문서화와 관련된 요소로 구성된다(Assem et al. 2004).

개념 중심 모델은 개념 및 복합 개념, 우선 레이블, 비우선 레이블 등 개념 표현 요소와 BT, NT, RT 등 개념 관계와 문서화 요소로 구성된다. 개념 모델에서는 노드 레이블과 패킷이 중요한 역할을 한다. 개념 중심 모델의 구조와 예를 그림 13에 보였다

시소러스의 개념 모델 정립을 위해 대상 시소러스의 용어(또는 클래스)와 용어 관계를 분석하여 시소러스 모델 분석서를 작성한다. 본 연구에서는 한국역사 용어 시소러스에 대한 시소러스 모델 분석 예를 그림 14에 보였다

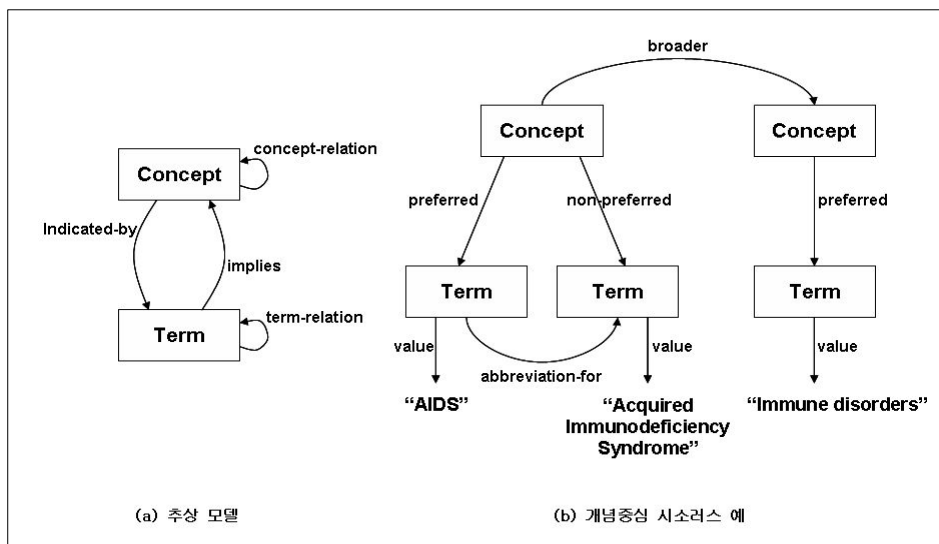
그림 14의 시소러스 모델 분석에서는 TI, PE, LK에 사용되는 Time(년도), Period(시대), Link(분류체계)를 독립적인 요소로 설정하였다. 이외에도 시소러스 용어 관계의 확장에 의

해서 생성되는 비표준적 요소를 용어간의 관계와 의미 특성에 따라 독립 요소로 모델화 할 수 있다. 이러한 비표준적인 요소를 SKOS에서 사용하기 위해서는 이들을 사전에 정의하여야 하는데, 먼저 용어 요소인 Time, Period, Link를 정의한 후, 의미 관계를 정의하여야 한다. 그림 14 시소러스 모델을 표현할 수 있도록 SKOS를 확장하면 그림 15와 같다.

그림 15에서는 LK 분류체계의 계층성을 고려하여, SubjectHierarchy의 상위 관계와 하위 관계인 broaderLink와 narrowerLink를 별도로 정의 하였다.

#### 4.2 시소러스의 SKOS 생성

SKOS 생성 단계(Step 2)에서는 Step 1의 분석 결과를 활용하여, 변환 방식을 정의하고 실제 SKOS 온톨로지를 생성한다. SKOS 온톨



(그림 13) 개념 중심 시소러스 모델

일반 사항	시소러스 명칭	한국역사용어 시소러스	
	저자	홍길동	
	프로젝트	역사정보 시스템 구축 프로젝트	
	기관	한국역사연구소	
	날짜	2006년 4월 5일	
클래스 (class)	관계 (relation / property)	치역 (range)	비고
Term			용어
	USE	Term	
	SN	rdf:literal	
	TI	Time	년도주기
	PE	Period	시대주기
	LK	Link	분류
	UF	Term	비우선등가관계
	BT	Term	
	RT	Term	
	NT	Term	
Time			년도
Period			시대
Link			분류체계
	broadLink	SubjectHierarchy	
	narrowerLink	SubjectHierarchy	
SubjectHierarchy			분류주제어

(그림 14) 시소러스 모델 분석서 형식과 예

```

<rdfs:class rdf:ID="Time">
  <rdfs:subClassOf rdf:resource="&skos;Concept"/>
  <rdfs:label>time</rdfs:label>
</rdfs:class>
<rdfs:class rdf:ID="Period">
  <rdfs:subClassOf rdf:resource="&skos;Concept"/>
  <rdfs:label>period</rdfs:label>
</rdfs:class>
<rdfs:class rdf:ID="SubjectHierarchy">
  <rdfs:subClassOf rdf:resource="&skos;Concept"/>
  <rdfs:label>subject hierarchy</rdfs:label>
</rdfs:class>
<rdf:Property rdf:ID="broaderLink">
  <rdfs:subPropertyOf rdf:resource="skos:broader"/>
  <rdfs:range rdf:resource="#SubjectHierarchy"/>
  <rdfs:label>broader subject link</rdfs:label>
</rdf:Property>
<rdf:Property rdf:ID="narrowerLink">
  <rdfs:subPropertyOf rdf:resource="skos:narrower"/>
  <rdfs:range rdf:resource="#SubjectHierarchy"/>
  <rdfs:label>narrower subject link</rdfs:label>
</rdf:Property>
<rdf:Property rdf:ID="Link">
  <rdfs:subPropertyOf rdf:resource="skos:semanticRelation"/>
  <rdfs:range rdf:resource="#Subject Hierarchy"/>
  <rdfs:label>subject link</rdfs:label>
</rdf:Property>

```

(그림 15) SKOS의 확장 방법

로지 생성은 시소러스 개념 모델을 SKOS 개념 모델로의 매핑 방식 정의와 실제 SKOS 온톨로지 생성의 두 단계로 수행된다.

1) 개념 매핑(Step 2.1)

시소러스의 개념모델은 앞서 기술한 바와 같이, 용어 지향 시소러스와 개념 지향 시소러스의 두 종류가 있다. 또한 ANSI/NISO 39.19를 기반으로 표준형으로 기술된 시소러스와 비 표준형식으로 기술된 시소러스가 있다. 용어 지향 시소러스를 개념 지향 모델로 변환하는 것에 대한 문제가 제기되기도 하지만, 용어 지향 시소러스는 개념 지향 시소러스로 변환이 불가능한 것이 아니므로 용어 지향 시소러스를 SKOS로 변환할 수 있다. 비 표준형식으로 기술된 시소러스는 개념 모델의 형태와 의미 구조를 참조하여 SKOS와의 상호 매핑 관계를 정의 할 수 있다. 이 같은 맥락에서 시소러스와 SKOS 온톨로지 사이에는 표 3과 같은 매핑 관계가 존재함을 분석할 수 있다.

표 3과 같은 기본적인 매핑 관계를 참조하여 시소러스 개념 모델의 형태와 의미 구조를 보존할 수 있는 다양한 매핑 관계를 정의할 수 있다. 매핑 관계를 정의할 때는 먼저 식별자(identifier)를 확인할 수 있어야 한다. 즉, skos:Concept의

rdf:ID로 사용할 용어를 선정하여야 한다. 식별자를 선정할 수 없는 경우에는 다음과 같은 두 가지 방법을 이용할 수 있다.

- ① 기존 용어나 개념과는 무관한 새로운 식별자 생성
- ② 우선어를 식별자로 사용

①의 경우, 새로운 식별자와 원래 용어 사이의 관계를 별도로 관리하여야 하는 단점이 있고, ②의 경우에는 개념이 용어에 의존하게 된다. 이러한 특징을 고려하여 구축하고자 하는 온톨로지의 특징을 파악하여 식별자를 선정하여야 한다. 시소러스가 비표준 형식이거나 독특한 구조를 갖고 있는 경우에는 앞서 서술한 바와 같이 계층구조와 의미 특성을 고려하여 이들을 skos:Concept 또는 skos:TopConcept의 하위 클래스로 생성한다.

이와 같은 방법으로 그림 14는 한국 역사용어 시소러스에 대하여 표 4의 개념 대응 변환표를 작성하였다. SKOS가 시소러스의 온톨로지 표현을 위해서 작성된 표준이므로 1:1 대응관계를 무리 없이 만들 수 있다. 비표준 시소러스의 경우에는 시소러스 개념 모델로부터 1:1 대응관계의 유도가 가능하다(W3C 2005b).

(표 3) 시소러스 모델과 SKOS 온톨로지 매핑관계

시소러스	SKOS 온톨로지
우선어 (preferred term)	개념의 우선레이블 (preferred label) skos:prefLabel
비우선어 (non-preferred term)	개념의 대체레이블 (alternative label) skos:altLabel
BT	skos:broader
NT	skos:narrower
RT	skos:related
SN(Scope Note)	skos:scopeNote
definition	skos:definition

(표 4) 개념 대응 변환표

일반 사항	시소러스	한국역사용어 시소러스
	온톨로지	한국역사용어 온톨로지
	작성자	홍길동
	작성기관	한국역사연구소
	날짜	2006년 4월 10일
시소러스 구성요소	특징 및 기능	SKOS 온톨로지 요소
USE A	우선어(표제어)	skos:Concept rdf:ID="A" skos:prefLabel="A"
SN B	Scope Note	skos:scopeNote="B"
TI C	년도 주기	khis:time="C"
PE D	시대 주기	khis:period="D"
LK E	분류	khis:link="E"
UF F	비우선어(등가관계)	skos:altLabel ="F"
BT G	상위어	skos:broader ="G"
RT H	관련어	skos:related ="H"
NT K	하위어	skos:narrower ="K"

2) SKOS 온톨로지 생성(Step 2.2)

이 단계에서는 개념 대응표를 활용해서 대상 시소러스를 SKOS 온톨로지로 변환한다. 여기서는 알고리즘적인 변환을 수행하며, 별도의 의미 해석은 배제한다. 변환 절차가 알고리즘 방법으로 수행되므로, 수작업으로 변환하는 것보다는 별도의 변환 프로그램을 개발하여 자동 변환하는 것이 바람직하다. 별도의 변환 프로그램은 RDF, API 등을 활용하면 용이하게 개발할 수 있다. 그림 16은 개념대응 변환표에 의해 생성된 한국역사용어 온톨로지를 보인 것이다.

4.3 보정 및 문서화

보정 및 문서화 단계(Step 3)는 시소러스를 SKOS 온톨로지로 변환 후의 처리 단계로서, 대상 시소러스와 생성된 SKOS 온톨로지 사이의 형태와 의미 구조 보존 여부를 검사하고, 생성된 SKOS에 대한 문서화 작업을 수행한다.

1) 보정(Step 3.1)

보정단계에서는 시소러스의 형태 및 의미 구조가 생성된 SKOS 온톨로지에 보존되어 있는지를 확인하고 추가 보완 작업을 하는 것이다. SKOS는 ISO 2788:1986에 적합한 구조를 갖고 있기 때문에 비표준 특성을 갖고 있거나 모호성(ambiguity)이 있는 시소러스의 경우에는 형태와 의미 구조를 재현하는데 어려움이 있다. 보정 단계에서는 step 1에서 분석한 시소러스의 개념 구조를 근거로 하여 이러한 사항을 점검한다.

SKOS에서는 관계간의 특성이 존재한다. 예를 들어, skos:broader와 skos:narrower는 역관계이고, 각각에는 추이율이 있다. skos:related 사이에는 대칭률(symmerticity)이 존재하기도 한다. 또한 다음과 같은 주제 일반성 규칙(subject generality rule)이 성립하고 있다.

[(?x skos:subject ?y)(?y skos:broader ?z) 이면(?x skos:subject ?z) 이다]

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  xmlns:khis="http://www.koreahistory.org/thesaurus#">

  <skos:Concept rdf:ID="상서공부">
    <skos:prefLabel>상서공부(尙書工部)</skos:prefLabel>
    <skos:scopeNote>고려시대 상서6부의 하나로 산택·공장·영조의 일을 관장하던 중앙기관</skos:scopeNote>
    <skhis:time>995-1389</skhis:time>
    <skhis:period>고려시대</skhis:period>
    <skhis:link>중앙행정기구</skhis:link>
    <skos:altLabel>공부</skos:altLabel>
    <skos:broader rdf:resource="#상서6부"/>
    <skos:related rdf:resource="#공관"/>
    <skos:related rdf:resource="#공조"/>
    <skos:related rdf:resource="#상서공부낭중"/>
    <skos:related rdf:resource="#상서공부시랑"/>
    <skos:related rdf:resource="#상서공부원외랑"/>
    <skos:related rdf:resource="#지상서공부사"/>
    <skos:related rdf:resource="#판상서공부사"/>
    <skos:narrower rdf:resource="상서수부"/>
    <skos:narrower rdf:resource="상서우부"/>
  </skos:Concept>
</rdf:RDF>
    
```

(그림 16) 시소러스 변환 온톨로지의 예

이러한 SKOS 내부 관계와 규칙을 고려하여, 생성된 온톨로지와 시소러스와의 구문 및 의미 관계를 확인한다. step 2.2 SKOS 온톨로지 생성 시 에러가 발생하였다면 그 원인을 파악하고 해결 방안을 모색한다. 이 과정에서 보정 사항 기술서를 작성한다.

2) 문서화(Step 3.2)

생성된 온톨로지에 대한 보정이 완료 된 후에는 문서화 작업을 통하여 관리하여야 한다. 문서화 작업은 skos:historyNote, skos:editorialNote,

skos:changeNote 를 활용하여 생성된 온톨로지 관리에 필요한 사항을 수록한다. 필요시, DCMI 의 메타데이터 등을 이용하면 유용할 것이다. 끝으로 온톨로지 구축 보고서를 작성한다.

5. 결 론

정보를 효과적으로 관리하고 이용하기 위한 방법으로 시소러스가 개발되고 활용되어 왔다. 시소러스는 도메인의 구성 어휘와 어휘 관계를

조직화하여 정보의 색인과 통제, 검색에 효율적인 방법을 제공하지만, 단순 의미관계만을 제공하고 있기 때문에 지적인식 수준의 정보 처리와 관리에 한계가 있다고 지적되고 있다. 이러한 이유로 도메인 개념 모델링의 새로운 방법으로 사용되고 있는 온톨로지의 활용에 대한 연구가 활발하게 진행되고 있다.

온톨로지는 도메인의 개념요소와 다양한 의미관계, 도메인의 공리체계와 규칙 등을 명세화하여 추론과 같은 지식정보 처리 기능을 제공할 수 있어 정보시스템 구축의 새로운 방향을 제시하고 있다. 하지만, 온톨로지 구축은 많은 전문 인력이 요구되는 작업으로 현실적인 많은 문제를 내포하고 있다. 이러한 문제점은 기존의 시소러스를 온톨로지로 변환함으로써 해결이 가능하다. 시소러스를 활용하여 온톨로지를 구축함으로써 기존의 정보자원을 활용할 수 있을 뿐 만 아니라, 실용적인 온톨로지를 효과적으로 구축할 수 있고, 도메인 개념 모델을 보다 명확하게 이해할 수 있는 장점이 있다.

W3C에서는 시소러스와 온톨로지 간의 형태적 상호운용성을 보장하고 시소러스 어휘간의 다양한 관계를 표현할 수 있는 시소러스 기반 온톨로지 구축의 표준 모델로 SKOS를 제안하고 있다. SKOS는 RDF를 기반으로 한 분류, 시소

러스 등을 위한 범용 프레임워크로 어휘 용어와 의미개념의 표현에 효율적인 구조를 갖고 있다. 그러나 시소러스를 SKOS 기반의 온톨로지로 변환하기 위해서는 시소러스의 도메인 모델링과 SKOS의 범용 프레임워크를 일치시켜야 하는 것 등의 문제가 있다. 이러한 점을 고려하여 본 연구에서는 대상 시소러스를 SKOS 온톨로지로 구축하는 단계적 변환 방법을 제시하였다.

본 연구는 시소러스의 SKOS 온톨로지 구축을 시소러스 분석, SKOS 생성, 보정 및 문서화의 3단계로 도출하였고, 각 단계별 수행작업과 온톨로지 변환 방법에서는 온톨로지 구축을 위한 구체적 표준 절차를 제시하였으며, 단계별 작업을 통해서 효율성을 제공할 수 있는 방안을 마련하였다. 이것은 용어 관계의 확장에 의해서 생성될 수 있는 다양한 시소러스 의미관계를 포괄할 수 있는 변환 방법이다.

본 연구에서 제시한 시소러스의 단계별 온톨로지 변환 방법은 일관된 작업 흐름과 정형화된 방법을 제시하고 있으므로, 단계별 변환 알고리즘을 프로그램으로 구현하는 것이 가능하다. 향후 본 연구에서 제시한 방법을 구현한 시소러스의 자동 SKOS 온톨로지 변환기의 개발이 수행되어야 할 것이다.

## 참 고 문 헌

고영만. 2006. 시소러스 기반 온톨로지에 관한 연구. 『정보관리』, 5: 5-22.  
국사편찬위원회. 2005. 한국역사용어시소러스

[cited 2006. 5. 20].

〈<http://thesaurus.history.go.kr/>〉

류주희. 2004. 한국역사용어 시소러스의 구축

- 방법론. 『한국사학회』, 73: 207-226
- 한상길. 1999. 『시소러스 용어관계의 확장에 관한 연구』. 박사학위논문. 중앙대학교 대학원, 문헌정보학과.
- Assem, M. van, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga, 2004. *A Method for Converting Thesauri to RDF/OWL*. Proceedings of ISWC-2004, Hiroshima, Springer-Verlag.
- Assem, Mark, Veronique Malaise, Alistair Miles and Guus Schreiber. 2006. A Method to Convert Thesauri to SKOS. [cited 2006. 5. 20]. <<http://epubs.cclrc.ac.uk/bitstream/956/skospapereswc06.pdf>>
- Brodie, M. L., J. Mylopoulos and J. W. Schmidt. 1984. *On Conceptual Modeling*. New York : Springer-Verlag.
- Doerr, M. 2001. 『Semantic problems of thesaurus mapping』. *Journal of Digital Information*, 1(8). Article No. 52, 2001-03-26. [cited 2006. 6. 17] <<http://jodi.tamu.edu/Articles/v01/i08/Doerr/>>
- LIMBER(Language Independent Metadata Browsing of European Resources). 2000. [cited 2006. 5. 20]. <<http://www.limber.rl.ac.uk/>>
- Matthews, Brian. 2003. Migrating Thesauri to the Semantic Web, W3C Semantic Tour, London Zoo 12/06/2003. [cited 2006. 5. 20]. <<http://www.w3c.rl.ac.uk/pastevents/matthews/semantictour/title.html>>
- Mikhaleiko, Peter. 2005. Introducing SKOS. [cited 2006. 5. 20]. <<http://www.xml.com/pub/a/2005/06/22/skos.html>>
- Open Galen. GALEN. 2005. [cited 2006. 5. 20]. <<http://www.opengalen.org/>>
- Schmitz-Esser, Winfried. 1991. New Approaches in Thesaurus Application. *International Classification*, 18(3) : 143-147
- W3C. 2005a. Quick Guide to Publishing a Thesaurus on the Semantic Web. [cited 2006. 5. 20]. <<http://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050517/>>
- W3C. 2005b. SKOS(Simple Knowledge Organisation System). [cited 2006. 5. 20]. <<http://www.w3.org/2004/02/skos/>>
- Wielinga, B. J., A. Schreiber, J. Wielemaker, J. A. C. Sandberg. 2001. 『From Thesaurus to Ontology』. [cited 2006. 5. 20]. <<http://www.cs.vu.nl/~guus/papers/Wielinga01a.pdf>>
- UKAT(UK Archival Thesaurus). 2004. [cited 2006. 5. 20]. <<http://www.ukat.org/>>

к с і