

대단위 우리말 온톨로지 구축을 위한 시소러스의 개발

Constructing a Grand-Scale Korean Thesaurus as a Base for Ontology Development

최석두¹⁾, 이우범²⁾, 김이겸³⁾, 최상기⁴⁾, 한상길⁵⁾, 이정연⁶⁾

Suk-Doo Choi, Woo-Bum Lee, Lee-Kyum Kim, Sang-Ki Choi,
Sang-Kil Han, Jung-Yeoun Lee

ABSTRACT

This paper reports an effort to construct a grand-scale Korean thesaurus that can be used for enhancing retrieval performance in various fields. This thesaurus is currently being used for indexing and retrieving purpose and new terms are being added to it. As the new demands on retrieval performance increase in Korea, developing a grand-scale ontology appears to be necessary so a project is undertaken to transfer the current thesaurus into an ontology system. The paper describes how the thesaurus is constructed and prepared to be the base for an ontology system.

초 록

한국어 정보처리를 위하여 전 분야를 포괄하는 대단위 시소러스를 계획하고 이를 개발하고 있다. 이 시소러스는 현재 현장에서 정보의 색인과 검색용으로 사용되고 있으며 새로운 용어가 추가되면서 계속하여 증식하고 있다. 그런데 정보처리에 대한 새로운 요구가 증가함에 따라 대단위 온톨로지의 개발이 필요하게 되었으며, 이에 시소러스를 온톨로지로 변환할 준비를 하고 있다. 본고에서는 온톨로지를 만들기 위하여 지금까지 구축한 시소러스의 내용에 대하여

-
- 1) 한성대학교 지식정보학부 교수(sdchoi@hansung.ac.kr)
 - 2) 한성대학교 지식정보학부 교수(moon0610@hansung.ac.kr)
 - 3) 광주대학교 문헌정보학과 교수(leekyum@kwangju.ac.kr)
 - 4) 전북대학교 문헌정보학과 교수(choisk@moak.chonbuk.ac.kr)
 - 5) 대림대학 문헌정보과 교수(skhan@daelim.ac.kr)
 - 6) 한국학술진흥재단 지식정보센터 박사후 연구원(shampoo@krf.or.kr)

논한다.

핵심어: 시소러스, 한국어 시소러스, 온톨로지, 온톨로지 개발, 개념관계, 용어 풀, thesaurus, Korean thesaurus, ontology, ontology development, conceptual relationships, pool of terms

1 서론

우리나라에서는 1961년에 출판된 『주제명표목표』(이재철 1961)가 처음이었으며 이 주제명표목표에는 약 5,000어의 주제명이 수록되었다. 불행하게도 이 주제명표목표는 지금까지 한 번도 개정되지 않아 쓰는 곳이 없게 되었다. 1980년대에 들어와서 외국의 시소러스를 번역하거나 특정 기관용 혹은 전문 분야별 시소러스가 만들어지기 시작하였으며 지금까지 그 수는 10종 정도를 헤아릴 수 있다. 그러나 현재 정보처리시스템에 정상적으로 적용되고 있는 시소러스는 드물다.

시소러스의 개념관계는 군더더기 없이 간결하게 추상화 되어 있다. 그런 데 이에 대한 한계는 근본적으로 개념관계의 단순화에서 오는 것과 외부적인 요인으로 인한 것으로 나눌 수 있다. 외부적인 요인은 다음과 같은 것을 생각할 수 있다. 첫째, 이용방법을 잘 모르는 데 있다. 시소러스는 응용시스템이 아니라 데이터이다. 따라서 시소러스를 아는 만큼 응용하게 되며 정보처리시스템의 목적과 그 사용능력에 따라 시소러스의 효율성은 크게 달라진다. 둘째, 1회성으로 구축하는 데 있다. 시소러스의 대상은 용어이다. 용어와 용어관계는 계속적으로 추가, 확장, 분할, 통합되고 용어관계가 변화한다. 따라서 기 구축된 시소러스를 계속적으로 갱신하지 않는다면 멈춘 순간부터 낡기 시작하며 어느 정도의 시간이 지나면 전혀 쓸모없는 것이 되고 만다. 셋째, 시소러스를 발전시키지 않는 데 있다. 시소러스 관련 표준은 고착되어 있는 강제형식이 아니라 개념이며 이론이다. 정보처리기술이 진화하면 시소러스도 진화하여야 한다. 넷째, 완전성이 결여되어 있다. 하나의 용어에 대하여 모든 측면을 망라하여 수집하고 관계를 정의하는 것은 어려운 일임에 틀림없다. 그러나 용어의 수, 관계 등의 측면에서 완전할 필요가 있다. 다섯째, 일반적으로 전거에서 주제명전거를 제외하며 인명전거, 지명전거, 작품전거, 상품명전거, 기관/단체명전거 등과는 별도로라는 생각을 바탕으로 데이터와 시스템을 구축하는 데 있다. 이들을 분리해서 생각하는 것은 무리이다. 여섯째, 용어 및 기술규칙의 표준화

활동이 미비한 데 있다. 심지어는 우리말의 표현에서 가장 기본적인 띄어쓰기나 한글자모의 배열순서도 모호하다. 일곱째, 범용 시소러스의 국가적인 개발 계획이 없는 데 있다. 전 분야에 걸친 일반 혹은 종합시소러스, 그리고 소위 기타 전거데이터는 특정 개인이나 기관이 할 일도 아니며 할 수 있는 일도 아니다. 그것은 시간, 인력, 비용이 엄청나게 필요한 일이기 때문이다. 그것은 국가가 중심이 되어 만들고 필요한 기관에게 무료로 제공하여야 한다. 그리고 개념관계의 단순화에서 오는 한계를 극복하기 위해서는 개념관계를 확장하는 수밖에 없다.

그런데 최근의 차세대 인터넷에 대한 필요성과 이를 위한 시멘틱 웹(Berners-Lee et al. 2001)에 대한 구상으로 많은 관심이 온톨로지로 집중되고 있다. 한편으로는 시소러스도 온톨로지임에도 불구하고 시소러스의 한계를 불필요할 정도로 시소러스를 온톨로지와 차별화하고 그 차이를 강조하고 있는 경향도 있다.

이에 1995년부터 모든 분야를 망라하여 색인과 검색에 적용할 수 있는 우리말 시소러스의 구축을 시작하여 현재까지 약 52만 용어를 등록하고 있으며 개념관계의 단순화에서 오는 한계를 극복하기 위해서 온톨로지로 변환할 준비를 하고 있다. 그리고 앞으로 계속되는 온톨로지의 확장도 시소러스를 중간단계로 거치는 방법을 고수하려고 한다. 시소러스형식 자체를 유지하여 시소러스 고유기능을 활용할 필요가 있으며, 또한 시소러스에서는 모든 관계를 직접 술어형식으로 표현할 수 있을 뿐만 아니라 이를 포괄적으로 조망할 수 있어서 개념관계를 설정하고 수정하는 데 매우 편리하기 때문이다.

시소러스를 이용하여 온톨로지를 자동 혹은 반자동으로 구축하는 연구는 다수 있다(Clark et al. 2000, Wielinga et al. 2001, Soergel et al. 2004, Kawtrakul 2005). 이들 연구는 특정한 시소러스를 대상으로 일정한 조건을 만족하는 용어간의 관계를 자동 혹은 반자동으로 생성하는 연구들이며 범용성은 거의 없으나 자동생성의 가능성을 보여주고 있다. 다만, 본 시소러스의 개발에 즈음해서는 정밀한 개념관계를 설정하기 위하여 완전히 수동으로 구축하고 있다.

본고에서는 개념관계의 단순화와 외부적인 요인으로 인한 한계는 범용 온톨로지를 구축함으로써 극복할 수 있다고 보며, 이를 위한 작업의 일환으로 지금까지 구축하고 있는 시소러스의 내용에 대하여 논한다. 다만, 개념관계의 확장에 초점을 맞추어 논하고, 형용사 시소러스에 대한 구축내용, 그리고 시소러스 구축과 실험을 위한 시스템, 환경 및 실험결과는 다른 기회로 미룬다.

2 시소러스

2.1 용어의 대상

한 나라의 국어로 표현할 수 있는 용어는 무수히 많다. 이들 용어를 모두 수집하는 것은 거의 불가능에 가까운 일임에 틀림없다. 그러나 정보처리의 세계에서는 이들 용어의 망라적인 집합을 필요로 한다. 본 시소러스에서는 대상이 되는 용어의 범위를 다음과 같이 정하고 있다. 첫째, 기본적으로 한글용어를 대상으로 한다. 명사, 동사, 형용사, 수사를 포함시키며, 주제나 개념을 표현한다면 구, 문도 포함시킨다. 즉, 한글로 표기된 국내외의 모든 事象이 대상이 되므로 외국어의 음차어를 포함한다. 예를 들면, 일반 주제어를 필두로 우리말로 표기된 국내외의 인명(모든 이명 및 칭호를 포함한다), 법률명, 동식물명, 기관/단체명, 상품명(모형을 포함한다), 지명(일반 목록규칙에서 정의하고 있는 지명을 말한다), 기념물명, 국보, 작품명(모든 분야의 작품을 포함한다), 회의명이 대상이 되며, 문자열의 길이에 관계없이 하나의 용어로 취급한다. 둘째, 로마자 두문자어를 포함한다. 셋째, 특정 저작물 혹은 제목이 우리말로 번역되지 않은 경우에는 외국어 제목을 포함한다. 넷째, 이상의 범주에 속하고, 사전에 포함하여 문헌상으로 한 번 이상 사용된 용어라면 방언, 속어, 고어, 잘못 사용되는 용어를 불문하고 대상이 된다. 다만, 식별기호, 분류기호, 한자, 대응외국어 등은 필요한 경우에만 참조데이터로 이용한다.

2.2 용어의 표준화

2.2.1 표준용어

분야를 불문하고 하나의 개념을 표현하는 데 두 가지 이상의 용어가 사용되는 예는 허다하다. 완전한 동의어인 경우도 있으나 잘못 알고 사용하는 경우도 많다. 심지어는 같은 개념에 대하여 전문분야에 따라 다른 용어를 사용하며 하나의 용어로 통일하는 것을 반대하기도 한다. 예컨대, 국어연구원(전 국립국어연구원)이 펴낸 패션 디자인 용어에 대한 『국어순화자료집』(1999)을 보면 ‘순화한 용어’가 다음과 같이 복수로 권장되고 있음을 볼 수 있다(다음 예에서, 왼쪽 난은 순화 대상 용어, 중간 난은 순화한 용어이며, 오른쪽 난의

기호 X는 순화한 용어만 씀, 기호 →는 될 수 있으면 순화한 용어를 씀, 기호 ○는 순화 대상 용어와 순화한 용어를 모두 씀을 의미한다).

고 (고데) (鋏 こて 일)	→ 인두, 머리인두, 머리인두질	X
캐주얼 웨어 (casual wear 영)	→ 평상복, 캐주얼복	→
콤비네종 (combinasion 프)	→ 속치마, 항공복	→
노하우 (know-how 영)	→ 기술, 비결, 비법	○
컬러 (color 영)	→ 색, 색채, 색상, 빛깔	○

또한 Dublin Core의 기본요소인 *Creator*에 대한 한국어의 역어는 열 가지 이상이 가능하며 어떤 의미에서는 모두가 타당한 것이다. 이와 같은 일이 생기는 것은 개인의 기호, 문맥, 공기(共起), 언어습관, 조어규칙, 복합명사 생성 시 조합되는 단위개념 등에 따라 달리 선택될 소지가 있으며 근본적으로 그것을 허용할 필요가 있다. 물론 가능한 한 여러 가지 기준을 만들어 통일하려는, 즉 표준화하려는 노력을 열심히 하고는 있으나 궁극적으로 모든 용어에 대하여 하나의 개념에 하나의 용어(문자열)를 대응시키는 완전한 표준화는 원칙적으로 불가능하다. 따라서 어휘를 통제하는 입장에서는 대표어(전거)를 정할 수가 없게 되므로 대표어를 고정시킬 수가 없다. 따라서 어휘를 통제하는 입장에서는 대표어(대표표현; 전거; 표준용어; authority)를 정할 수가 없게 되므로 대표어를 고정시킬 수가 없다. 이를 논리적으로 해결하기 위해서는 주제명의 세계에서 대표어를 인정하지 않는 방법밖에 없다. 본 시소러스에서는 대표어(가변가능)와 UF의 관계를 서로 호환이 가능하도록 설계하고 있다(최석두 1993). 즉, 어떤 이용자가 대표어를 참조했을 때는 아무런 변함이 없으나, UF에 해당하는 용어로 참조했을 때는 UF 용어를 대표어의 자리에 놓고 대표어를 UF로 옮긴다. 원래의 대표어에는 그것이 대표어라는 부가정보는 가지고 있으나 이용자의 입장에서 보면 이용자가 참조한 것이 대표어의 위치에 있게 된다.

2.2.2 용어의 풀

모든 분야 혹은 특정 분야의 용어를 모아 사용빈도와 중요도 등을 고려하여 이차원 평면에 정리한다면 대략 삼각형과 유사한 모양이 될 것이다. 즉, 사용빈도와 중요도가 높고, 해당 분야의 주제를 정확하게 표현하는 용어일수록 그

수가 줄어들 것이고, 인접분야와 공유하는 용어일수록 그 수가 늘어날 것이다. 그러나 수집된 용어를 모두 해당분야의 표준 집합으로 삼을 수는 없다. 왜냐하면 실제로 ‘해당 분야의 용어’의 집합은 해당 분야 개념체계와의 거리를 어디까지로 보느냐에 따라 달라질 뿐만 아니라, 그 분야의 용어를 완전히 수집하는 것은 불가능하기 때문이다. 예를 들면, 농업분야의 용어라고 했을 때, 동식물이 관계 되므로 ‘식물학’, ‘동물학’과 ‘임업’ 등, 비료나 농약과 관계되므로 ‘화학’, 기상과 관계되므로 ‘기상학’, 농기계가 관계되므로 ‘기계학’, 토양과 관계되므로 ‘토양학’과 ‘지질학’, 문화와 관계되므로 ‘문화인류학’ 등과 같이 관련 분야를 이어가면 결국 ‘역사, 문화, 교육’까지 확장되며, 그 중간 어디에서나 임의로 잘라 그곳까지가 농업분야라 할 수 있다. 어떤 분야에서도 이와 같은 현상이 일어난다. 결국 수집한 용어에서 적절한 수준에서 농업분야 용어를 선정할 필요가 있으며, 용어집합의 표준화를 위해서는 용어의 삼각형을 밀변과 평행인 선분으로 분할하는 것이 가장 이상적일 것이다. 그러나 실제로는 그 삼각형을 만들기도 밀변과 평행인 평행선을 찾는 것도 어렵다. 이와 같은 상황에서는 새로운 해결방법을 도입할 필요가 있다. 그것은 용어의 풀을 만드는 방법이다.

용어의 풀(용어저장소)은 많은 집합으로 구성되어 있으며, 모든 집합은 크게 네 가지 특성을 갖는 그룹으로 나눌 수 있다. 첫 번째 그룹은 용어 특성으로 나눈 그룹이며, 전술한 바와 같이 품사, 인명, 법률명, 동식물명, 기관/단체명, 상품명, 지명, 기념물명, 작품명, 회의명 등의 기본집합을 갖게 된다. 두 번째 그룹은 분류체계로 나눈 그룹이며, 각종 학문분류체계 및 동식물 분류체계를 비롯하여, 상품분류, 직업분류 등의 기본집합을 갖게 된다. 세 번째 그룹은 외국어 그룹이며, 한국어에 대응되는 모든 외국어(현재 34개 국어)는 각각의 기본집합을 갖게 된다. 그리고 네 번째 그룹은 UF 용어를 포함하는 사용기관별 그룹이며, 본 시소러스를 사용하는 각 기관은 그 기관이 선택하여 사용하고 있는 용어의 집합을 갖게 된다. 이 용어 풀은 모든 기본집합 단위로, 임의로 선택된 집합 단위로, 그리고 전체적으로 관리되고 브라우징할 수 있다. 용어 풀의 전체적인 개념은 그림 1과 같다.

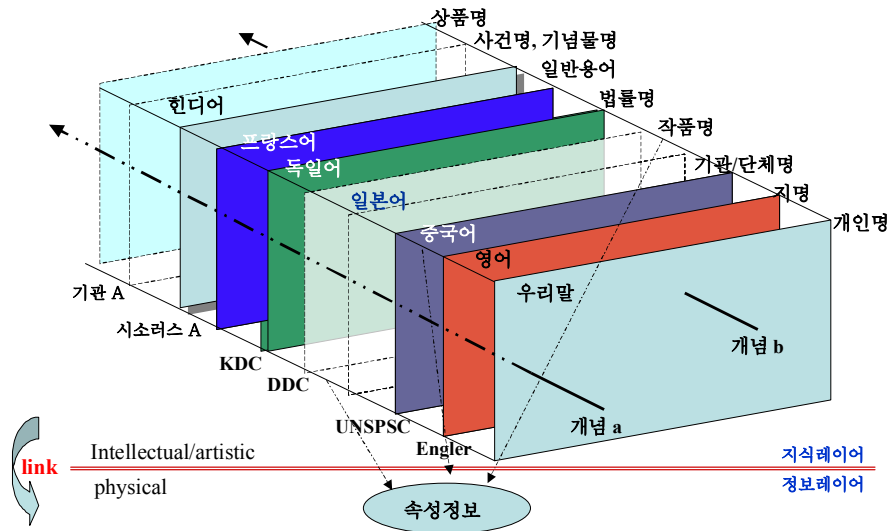


그림 1 용어 풀

모든 용어는 일차적으로 전체집합의 관점에서 추가하여 용어관계 풀을 구성하게 되며, 이차적으로 기본집합을 구성하게 된다. 특히 사용기관별 기본집합은 용어 풀의 사용기관이 풀 안에 있는 어떤 용어를 선택하여 사용하면 그 용어가 그 기관의 기본집합에 포함되는 것이다. 이 때, UF에 해당하는 용어의 선택도 허용하게 된다. 따라서 용어관계 풀에는 어떤 용어라도 계속적으로 추가될 수 있으며 용어의 수가 늘어나면 늘어날수록 기본집합의 형성이 점점 쉬워지고 또한 충실해지게 될 것이다. 하나의 객관적이고 고정적인 용어집합의 표준이란 존재할 수 없으며 보는 측면에 따라 기관에 따라 달라지는 동적인 표준이 있을 뿐이다.

3 시소러스와 온톨로지

온톨로지는 주지하고 있는 바와 같이 원래 철학용어이며 ‘존재론’ 혹은 ‘존재에 관한 체계적인 이론’이라는 의미이다. 그런데 이제는 오랫동안 철학자들이 ‘존재’라는 불가해한 개념을 이해하고 설명하기 위해 만든 온톨로지를 한마디로 ‘상위온톨로지’라고 구분하여 별칭하고 있다. 온톨로지에 대한 정의는 매우 다양하여 사람, 목표, 대상 등에 따라 다르다. 그중 범용적인 정의로는 다음과 같은 정의가 있다. 즉, “대상 세계의 정보처리 모형을 구축하는 사람이 그 세계를 어떻게 ‘보았는지’, 다시 말하면 그 세계에는 ‘무엇이 존재하고 있다’고 보고 모형을 구축했는지를 (공유를 목적으로) 명시적으로 표현한 것이며, 그 결과로 얻어진 기본개념이나 개념간의 관계를 토대로 모형을 기술할

수 있는 것”(溝口 2005, 9). 여러 가지 정의에서 언제나 포함되고 있는 구성적인 내용은 매우 단순하여 온톨로지는 개념의 집합이며, 개념의 계층관계, 그 외에 필요한 관계, 그리고 개념과 관계의 정의(의미제약의 공리화)이다. 이 기준으로 보면 주위에 있는 대부분의 분류표와 기존의 시소러스는 모두 온톨로지에 속한다.

그러나 대부분의 분류표와 시소러스는 소위 ‘가벼운 온톨로지’(light-weight ontology)에 가깝다. 지식베이스의 틀을 만드는 입장에서 개념간의 관계에 의미를 엄밀하게 개입시켜, 개념과 관계의 형식적 정의 혹은 의미제약을 기술함으로써 재이용성, 공유성, 의미의 명확성을 높이고, 개념과 관계의 인스턴스를 작성함으로써 그 대상의 모형을 작성, 구축, 조작할 수 있는 ‘무거운 온톨로지’(heavy-weight ontology)와는 어느 정도 구별된다.

필자는 가벼운 온톨로지가 가지고 있는 개념관계를 확장함으로써 무거운 온톨로지를 만들 수 있다고 보고 있으며 이를 만들기 위하여 확장하고 있는 여러 가지 확장관계에 대하여 논한다.

4 관계의 확장

모든 이론과 그 이론을 바탕으로 하는 응용은 상황에 따라 진화한다. 전통적인 시소러스의 용어관계가 단순하면 이를 확장하면 되는 것이며, 정보처리시스템의 목적에 따라 적절히 수정하여 최적의 데이터를 구축할 필요가 있다.

가까운 예로는 *LCSH*(Library of Congress 1992)를 들 수 있다. 전조합 색인언어인 주제명표목에서 후조합 색인언어인 시소러스형으로 모습을 바꾸었다. *ROOT thesaurus*(BSI 1988)에서는 관계지시기호를 기호로 표현하고 있으며, 중국분류주세사표(중국도서관도서분류법편위원회 1994)에서는 D, Z, C 등의 문자를 사용하여 관계를 표현하고 있다. 독일이나 프랑스에서도 자체의 관계지시기호를 사용한다. 또한, ANSI/NISO Z39.19-1991에서는 이미 부록으로 다음과 같이 우리나라 현대그룹이 자회사 정보를 시소러스형식으로 사용하고 있는 응용례를 들고 있다. 여기에서 사용한 관계명 JV와 SB는 JV=Joint venture with (a type of related term), SB=Subsidiary of (a type of narrower term)를 나타내고 있다. 표준은 그 틀을 정의하고 있는 것이므로 그 틀 내에서는 자유롭게 대응시킬 수 있으며, 틀을 깨지 않는 한 필요에 따라 다른 정보들을 추가할 수도 있는 것이다.

08487624 APAC

LANGUAGE: English

COMPANY: Hyundai Group [South Korea]

JV Hyundai Motors Co. [South Korea] 85% with Mitsubishi Group
[Japan]

SB Hyundai Auto Canada Inc. [Canada]

SB Hyundai Construction & Engineering Co. [South Korea]

SB Hyundai Electrical Engineering Co. [South Korea]. [Further
information available on this company; Use MC-]

SB Hyundai Electronics America Inc. [USA]. [Company address(s)
available; Use CO- and DIRECTORIES/DE.]

SB Hyundai Electronics Industries Co. Inc. [South Korea]

SB Hyundai Heavy Industries Co. [South Korea]

SB Hyundai Merchant Marine Co. [South Korea]

SB Hyundai Mipo Dockyards Co. [South Korea]

SB Hyundai Motor America Inc. [South Korea]

SB Hyundai Motor Service Co. [South Korea]

SB Hyundai Precision Industrial Co. [South Korea]

SB Hyundai Wood Products Co. [South Korea]. [Further information
available on this company; Use MC-]

SB Suneel Alaska Corp. [USA] 50%

용어 혹은 개념의 관계에 대한 연구는 무수히 많다. 본 연구에서는 모든 관계를 1차 개념관계, 2차 개념관계 및 속성관계로 크게 그룹화한다(Choi 2006). 1차 개념관계는 일반적으로 표준에서 정하고 있는 관계가 중심이 되는 기본관계이며, 2차 개념관계와 속성관계는 본 시소러스에서 확장하고 있는 관계이다. 2차 개념관계는 1차 개념관계가 정하고 있는 것 외에 계층관계, 동등관계, 관련관계를 각각 더 확장하고 있으며 계층관계의 확장은 다시 의미역관계와 패킷관계로 나눌 수 있다. 속성관계는 지금까지 고유명사류의 전거에서 일부 적용하고 있었으나 주제명 전거나 시소러스에는 볼 수 없었던 새로운 관계로 무거운 온톨로지로를 위하여 중요한 역할을 한다.

확장관계명은 모두 시소러스에 등록되어 있는 우리말 용어로 표현하며 동일

한 한 쌍의 용어에 둘 이상의 관계명을 가지는 것을 허용하고 있다.

4.1 관계의 방향성

관계를 설정하기 위해서는 관계명과 관계의 방향성을 정할 필요가 있다. 관계의 방향성은 네 가지 유형을 생각할 수 있다. 첫째, 양방향관계(대칭관계: \leftrightarrow)이다. 예를 들면, <남편 [부부] 아내>, <아내 [부부] 남편>의 관계가 그것이다. RT관계도 여기에 속한다. 둘째, 대응관계(\rightleftharpoons)이다. 예를 들면, <아버지 [아들/아버지] 아들>, <아들 [아버지/아들] 아버지>의 관계가 그것이다. 어느 쪽에서 보느냐에 따라 그 관계가 달라진다. NT/BT 및 USE/UF 관계도 여기에 속한다. 셋째, 일방향관계(\rightarrow)이다. 예를 들면, <모차르트 [출생년/] 1756>의 관계가 그것이다. 그리고 넷째, 이들 중 두 가지 이상의 유형을 갖는 복합유형이 있다.

4.2 1차 개념관계

기본적인 관계의 종류로는 ISO 2788:1986(E)에서 정의하고 있으며 계층관계(hyponymy: NT/BT), 동등관계(synonymy: USE/UF) 및 관련관계(association: RT)가 그 중심이 된다. 특히 ISO에서는 세 관계 중에서 계층관계만을 더 확장하고 있는데 전체부분관계, 사례관계 및 속관계가 그것이다. 전체부분관계(NTp/BTp)는 사물 전체와 그 전체를 구성하는 부분을 중심으로 본 관계이다. 예컨대, ‘손 NTp/BTp 손등, 손가락; 자동차 NTp/BTp 자동차 엔진, 자동차 타이어; 총무부 NTp/BTp 총무과, 서무과, 회계과’ 등이 이에 해당한다. 사례관계(NTi/BTi)는 하위개념어가 특정한 사례인가 아닌가를 중심으로 본 관계이다. 예컨대, ‘은행 NTi/BTi 한국은행, 국민은행; 꽃 NTi/BTi 진달래, 장미; 소설가 NTi/BTi 이광수, 김동리’ 등이 이에 해당한다. 속관계(NTg/BTg)는 각종 분류체계가 중심이 되나 본 시소러스에서는 동식물의 속종관계에만 적용한다. 상기 세 관계에 포함되지 않는 관계는 NT/BT만을 부여한다.

계층관계는 다른 관계보다는 비교적 정의하기가 쉽지만 그 유형은 상당히 복잡하다. 계층관계 간의 단위개념관계의 유형을 보면 그림 2와 같다. 그림 2의 대문자 한 자는 단위개념이며 세 개념까지를 조사하여 전개하고 있다. 복합어에 포함되어 있는 단위개념을 더 늘이면 더욱 복잡하게 될 것이다.

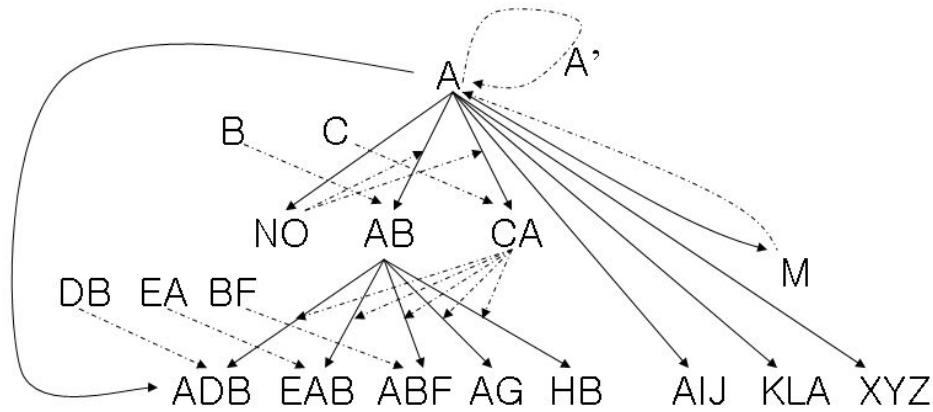


그림 2 계층관계의 단위개념관계 유형

계층관계를 판별할 때 조어적인 기준과 의미적인 기준을 적용할 수 있다. 조어적인 기준은 ‘어떤 문자열의 단위개념을 나타내는 부분문자열은 그 문자열의 상위개념어가 될 확률이 높다’는 기준이다. 예를 들면, ‘판매 경로 → 경로; 디지털 SLR 카메라 → 디지털 카메라’인 경우 우변이 좌변의 상위개념어가 될 확률이 높다. 그러나 조어적 기준도 용어의 개념을 파악하지 못하면 그 관계의 판단이 잘못되는 경우도 많다. 예컨대, ‘콩팥(臟器)→콩(豆)/팥(小豆); 간담(肝膽)→담(牆)’ 등은 우변이 상위개념어가 아니다. 조어적인 경우는 그림 2에서 ‘A→AB; A→CA; AB→EAB; AB→ABF’ 등이 해당된다. 의미적 기준은 문자열이 아니라 의미로 판단할 수밖에 없는 경우이다. 예컨대, ‘눈동자 → 눈 → 얼굴 → 머리’인 경우의 계층관계는 의미로서만 판단할 수 있다. 의미적인 경우는 그림 2에서 ‘A→M; A→NO; CA→ADB’ 등이 해당된다. 이들 기준은 계층관계 설정 시 참조기준일 뿐 그 정보를 기술하지는 않는다.

그러나 이와 같은 표준용어관계만으로는 관계가 너무 추상화되어 있어서 용어관계를 확장할 필요가 있다. 본 시소러스에서는 우리말의 특성을 고려하여 1차 개념관계를 중심으로 다음과 같이 임의로 확장한 용어관계가 있다. 즉, 학문분류기호로 DDC와 KDC의 두 분류체계, 특수 분류체계로 상품분류, 직업분류, 동식물분류 체계 등, 우리나라의 독특한 상황 때문이지만 북한어와 우리말, 학명과 동식물명(특히, 외국어 동식물명과 우리말 동식물명을 대응시킬 때 매우 유용하다), 이전용어(prior term: PT)와 나중용어(later term: LT)의 관계, 저자/작품/연도의 관계, 품사(현재는 명사와 형용사가 주임), 출전, 용어의 정의(일반용어의 정의 이외에, 국보, 보물, 유적, 사적, 문화재 등의 지정호수 및 소재지; 지명, 기관/단체명의 소재지; 작품의 서지사향; 사건의 시기; 인명의 이력; 법률명의 호수 및 제정/폐기 일자; 경축일의 일자 등의 정보도 포함

된다), 용어의 수준(1단계밖에 없으며 초중등학교의 교과서에 출현하는 용어인지 아닌지를 구별하고 있을 뿐이다), 용어의 범주(모든 등록용어는 다음의 중에서 하나 이상의 범주를 가져야 한다. 즉, 주제어, 법률명, 동물명, 식물명, 상품명, 국보, 기념물명, 한국인명, 외국인명, 외국의 왕이름, 국내의 왕이름, 국내기관/단체명, 국제기관/단체명, 지명, 국명, 행정구역, 서명, 총서명, 회의명), 그리고 각종 외국어 관계(외국어 코드는 ISO 639-2:1998에서 규정하고 있는 Alpha-3 코드를 사용하며 두 가지 이상의 표현이 있는 경우에는 이를 반복한다)가 있다. 외국어는 단순히 대응되는 용어일 뿐이고 우리말 개념체계에 대응되는 한 가지 외국어의 관계가 그 외국어의 개념체계를 나타내는 것은 아니다. 즉, 대응되는 외국어가 여럿 있다고 해서 다국어 시소러스가 되는 것은 아니다. 이외에 명시적인 관계는 아니지만 용어를 보완하는 정보들이 있다. 즉, 노드레이블, 띄어쓰기, 용어의 입력/수정/점검과 관련된 작업자 및 각종 일차정보, 한자, 동형이의어 구별을 위한 식별어, 이미지, 고립어의 관리 등이 있다. 또한 의미가 다르면 이를 한정어로 구분한다. 예컨대, ‘수박차’가 ‘수박을 싣고 팔고 다니는 차’, ‘수박 모양의 차’, ‘수박으로 만든 차’라는 세 가지 의미를 갖는다면 각각 ‘용도, 형상, 재료’를 의미하는 한정어를 부가하여 이를 구분한다. 이에 해당하는 모든 관계(확장관계를 포함하여)를 1차 개념관계(The first conceptual relationships: FCR)라 부르기로 한다.

4.3 2차 개념관계

전술한 용어관계도 개념관계의 일종이다. 개념이 중심이 되지 않으면 용어관계를 만들 수 없기 때문이다. ISO 2788-1986(E), ISO 5964-1985(E), ANSI/NISO Z39.19-1993 및 이를 기준으로 만들어진 기타 국가별 표준은 용어의 개념관계를 잘 표현하고 있다. 그러나 이들 기준에 의한 용어의 개념관계로는 실제의 색인이나 검색에 응용하는 데는 어느 정도 부족한 면이 있을 뿐만 아니라, 이와 같은 내용의 시소러스를 중간언어로 삼아 대단위 다목적 온톨로지를 만드는 데는 무리가 있다. 시소러스는 계층관계, 동등관계 및 관련관계를 중심으로 개념관계를 너무 추상화시켜 단순하게 표현하고 있으므로 실제의 정보처리상황에서는 모호성과 다의성이 너무 높기 때문이다. 따라서 시소러스 표준에서 정의하고 있는 이와 같은 개념관계를 1차 개념관계라고 보면 이를 더 깊게 세분하는 개념관계가 필요하며 이 확장관계를 2차 개념관계(The second conceptual relationships: SCR)라 하자. 즉, 1, 2차 개념관계를

함께 가지고 있을 필요가 있다.

본 시소러스에서 적용하고 있는 2차 개념관계를 계층관계, 동등관계, 관련관계로 나누어 순서대로 설명하기로 한다.

4.3.1 계층관계의 확장

용어간의 계층관계는 명시하지 않아도 자연스럽게 스스로 존재하고 있다. 그러나 시소러스의 세계에서는 두 용어관계가 계층관계라고 명시해야만 계층관계가 된다. 즉, 절대적인 의미에서 계층관계인지 아닌지 모호한 경우에도 특정 시소러스에서 계층관계라고 명시하면 그 시소러스에서는 계층관계가 되는 것이다. 이렇게 이미 정의된 계층관계를 대상으로 그 관계를 확장하는 것이다. 다시 말하면, 상위개념어를 세분한 하위개념어들이 어떤 측면에서 세분되었는가를 판단할 수 있으면 그 계층관계를 확장할 수 있다고 전제한다. 계층관계의 2차 개념관계는 의미역관계와 패킷관계로 확장할 수 있다.

앞에서 언급한 바와 같이 계층관계의 유형을 상위개념어/하위개념어 순서로 보면, 단일어/단일어, 단일어/복합어, 복합어/복합어의 세 가지를 생각할 수 있다. 결국 2차 개념관계를 생각할 때는 복합어 또는 합성명사의 구성 성분 사이의 의미관계를 파악하는 것이 일차적인 관건이다. 이에 대하여 김인균(2005, 104-129)은 합성명사의 의미관계를 다음과 같이 유형화 하고 있다. 즉, 그는 선·후행 명사관계를 ‘용도, 형상, 재료, 수단/방법, 유형/대상, 시간, 처소, 기원/소유, 용도’의 9가지로 나누고 각각의 경우를 구성성분의 의미유형을 분류하였다. 예컨대, <유형/대상>인 경우는 ‘사물+사물, 사건+사물, 사람+사람, 사물+사람, 위치+추상물’, <형상>인 경우는 ‘사물+사물, 사람+사물, 상태+사물, 사물+사람, 사물+사건, 상태+사건, 추상물+사건, 사물+추상물, 사람+추상물’, <재료>인 경우는 ‘사물+사물’, <수단/방법>인 경우에는 ‘사물+사물, 사물+사건, 사건+사물’, <시간>인 경우는 ‘위치+사물, 위치+사람, 위치+추상물, 위치+사건, 위치+위치’, <처소>인 경우는 ‘위치+사물, 위치+사람, 사물+사건, 사물+사물’, <기원/소유>인 경우는 ‘사건+사물, 사물+사물, 사물+추상물, 사물+사건, 추상물+추상물, 추상물/사물’, <용도>인 경우에는 ‘사물+사물, 사건+사물, 추상물+사물, 추상물+사건’의 결합양상을 띠고 있다고 보고 있다. 사물이나 개념에 반드시 있어야 하는 일차적인 의미속성인 ‘모양, 빛깔, 크기, 성질, 방식, 기능, 재료, 수단, 방법’ 등을 이들과 결합한다면 계층관계의 확장인 의미역관계와 패킷관계를 정의하는 데 크게 도움이 될 것이다.

4.3.1.1 의미역관계

명사는 보통명사, 고유명사, 의존명사, 대명사, 서술명사 등으로 분류할 수 있으며 이중 서술명사(verbale noun: 서술형 명사)는 논항(격)을 취한다(김종복 2004, 44). 예컨대 서술명사로는 ‘공부, 수출, 발견, 검사, 처리, 연구’ 등을 들 수 있다. 이러한 서술명사가 다른 용어와 결합하여 복합어 또는 합성명사(복합명사)를 만들고 원래 서술명사의 하위개념어가 되었을 때 상하위개념어간의 관계는 상위개념어를 세분한 측면, 즉 일방향관계의 논항 측면을 중심으로 보는 것이 합당하다고 본다.

Fillmore(1967)가 격문법을 제창한 이래 사람에 따라 여러 가지 격이 사용되었다. 용어관계에서는 표층격이 아니라 심층격을 대상으로 하며, ‘근원(원천), 대상, 도구, 시간, 목표, 목적, 장소(처소), 행위자(동작주), 이유, 경로, 원인, 방법, 방식, 방향, 수혜자’ 등 많은 격을 의미역 관계명으로 생각할 수 있다. 본 시소러스에서는 그 종류를 가장 판별이 쉬운 ‘근원, 대상, 도구, 시간, 목표, 장소, 행위자, 방식’의 8가지로 좁히고 용어관계의 예를 따라 확장할 방침이다. 의미역관계명의 예를 들면 다음과 같다.

[예] 검사 NT [대상/] LCD 검사
검사 NT [도구/] 내시경 검사
검사 NT [방식/] 비파괴 검사

4.3.1.2 패킷관계

이 패킷관계는 시소러스의 세계에서 오래 전부터 사용해 왔다. 특히 *ROOT Thesaurus*(BSI 1988)나 *Unesco Thesaurus*(Unesco 1977) 등의 분류시소러스에서는 많은 패킷지시어(패킷관계를 표현하는 관계어)를 사용하고 있다. 본 시소러스에서의 패킷관계는 어떤 두 개념 간에 상하위개념관계가 성립할 때 그 상위개념어를 하나 이상의 하위개념어로 세분하는 기준이 되었던 의미속성(내포; 意味素性; 의미자질)을 두 개념간의 패킷관계명으로 사용하고자 한 것이다. 예컨대, ‘온실’이 ‘가정 온실, 과수 온실, 목조 온실, 반원형식 온실, 반철조 온실, 부등식 온실, 실험용 온실, 알루미늄 온실, 양쪽 지붕형 온실, 연동식 온실, 영리 생산 온실, 외쪽 지붕형 온실, 일반 작물 온실, 채소 온실, 철조 온실, 표본 식물

온실, 화훼 온실’ 등을 가진다고 할 때, 그 패싯관계와 패싯관계명은 다음과 같다.

[예] 온실 [재료/] 목조 온실, 반철조 온실, 알루미늄 온실, 철조 온실
온실 [재배식물/] 과수 온실, 일반 작물 온실, 채소 온실, 화훼 온실
온실 [목적/] 가정 온실, 실험용 온실, 영리 생산 온실, 표본 식물 온
실
온실 [모양/] 반원형식 온실, 부등식 온실, 양쪽 지붕형 온실, 연동식 온
실, 외쪽 지붕형 온실

상기 예에서 ‘온실’과 ‘목조 온실’의 관계는 ‘목조’라는 명사의 의미속성 ‘재료’와 일치한다고 보는 것이다. 이와 같이 시험한 의미속성의 수는 1,200종을 헤아리며 대부분 서로 관계가 있는 특정 용어군에서만 사용되는 것이 보통이었다. 이를 시험적으로 추상화 시킨 결과가 ‘공간(부위, 위치, 지명, 지형, 차원, 천체), 물질(기기, 재료, 제품), 부분(부품, 분야, 성분, 요소), 생물(동물, 식물, 속종, 친족, 조직), 속성(감각(미각, 시각, 청각, 촉각, 후각), 길이(거리, 높이, 두께), 나이, 넓이, 부피, 비용, 상태, 속도, 압력, 온도, 전압, 질량, 크기, 현상), 유형(내용, 단체, 방식, 사례, 성별, 언어, 에너지, 조직, 종목, 질병, 형태), 작용(결과, 경로, 기능, 방법, 사건, 행위, 현상)’ 등이었다.

그러나 이를 적용하여 패싯관계를 정의한 결과, 다음과 같은 문제점을 발견하였다. 첫째, 데이터가 늘어나게 되면 패싯관계명이 상당히 늘어나게 되리라 생각되지만 적정 수가 몇 개인지 혹은 얼마나 상세히 세분해야 하는지에 대한 기준을 정할 수가 없었다. 둘째, 각 패싯관계는 배타적으로 적용할 수 있는 전형적인 경우를 가지고 있으나 두 가지 이상의 패싯관계가 하나의 용어관계에 적용될 수 있는 경우가 많았다. 예컨대, ‘방식, 유형’, ‘공간, 위치’, ‘물질, 성분, 재료’, ‘기기, 부품, 제품’, ‘상태, 속성, 현상’ 등이 이에 해당한다. 셋째, 어의상으로는 의미가 구분되나 관계패싯명으로 사용했을 때는 그 의미의 구분이 어려운 경우가 많았다. 예컨대, ‘목적’과 ‘용도’는 전혀 다른 의미를 가지고 있으나 관계패싯명으로 사용하게 되면 그 의미가 비슷하게 되어 구분하기가 어려웠다. 넷째, 패싯관계명 간의 계층관계가 서로 교차되는 경우가 있었다. 예컨대, ‘물질(기기, 재료, 제품)’과 ‘부분(부품, 성분, 요소)’은 서로 교환할 수 있으며 경우에 따라서는 계층관계가 바뀌기도 하였다.

그래서 현재는 그 구분이 어느 정도 배타적이고 객관성이 있는 것들을 통합하여 다음과 같이 16종의 관계패싯명을 적용하고 있다. ‘분야·이론·방법, 감각·

감정, 공간·위치, 기기·장치·부속, 물질·재료, 조직, 언어, 생물, 상태·성질, 시간, 단체, 질병·증상, 내용, 행위, 현상·사건, 사례'가 그것이다(황순희 외 2006). 16종의 관계패킷명으로서 지시하는 의미속성이 넓고 포괄적이며 추상적이라는 것을 알 수 있다. 실증적인 근거를 바탕으로 계속 추가, 세분, 수정해나갈 계획이다. 이 범주에 속하지 않는 용어관계는 정의하지 않고 비워두고 있는 것도 그 때문이다. 적용의 예를 보면 다음과 같다.

[예] 그래픽 카드 NT [기기·장치·부속/] PC 그래픽 카드
 유전자 NT [생물/] 동식물 유전자
 탐색 엔진 NT [공간·위치/] 인터넷 탐색 엔진

4.3.2 동등관계의 확장

동일한 개념을 나타낼 때 경우에 따라서 다른 용어를 사용하는 우리말만의 독특한 동의어들이 있다. 이들은 기본적으로 USE/UF로 표현하지만 이들을 다시 새로운 관계명으로 세분하고 있다. 그 관계명의 종류는 ‘동의어, 거센말, 고어, 궁중어, 낮춤말, 높임말, 대체어, 방언(각 지방), 변한말, 본딤말, 북한어, 비표준어, 속어, 승인어, 여린말, 여성어, 우선어, 유아어, 유의어, 은어, 작은말, 줄인말(준말), 참조어, 신조어, 친근체, 큰말, 폐기어, 표준어, 차음어, 틀린말, 비유’ 등이 있다.

두 가지 이상의 관계에 해당되는 경우에는 모두 부여한다. 예컨대, ‘비유’는 부차적 의미를 나타내는데 ‘속어, 은어’에도 속하는 경우에는 복수의 관계를 부여하고 있다.

[예] 국제연합 UF [차음어/] 유엔(UN)
 디지털 카메라 UF [줄인말/본딤말] 디카
 아버지 UF [유아어/] 아빠
 무능력자 UF [비유/] 바지저고리

4.3.3 관련관계의 확장

관련관계에 대한 정의를 보면 상당히 주관적이고 분야에 종속적이다. 그것은 관련관계를 다음과 같이 정의하고 있기 때문이다(ISO 2788-1986(E)).

첫째, 개념적으로 밀접하게 관련되어 있으나 계층관계 및 동등관계 집합에는 포함되지 않는 용어간의 관계, 둘째, 색인작성과 검색에 이용될지도 모르는 대체용어를 제시한다는 관점에서 용어간의 연결을 시소러스 중에서 명시하는 것이 좋을 것이라고 심리적으로 연상되는 것, 셋째, 이용자의 입장에서 볼 때, 한쪽의 용어가 색인어로서 채택되면 항상 다른 쪽의 용어가 강력하게 암시되는 것, 넷째, 한쪽의 용어가 다른 쪽의 용어를 정의하거나 설명하는 데 필요한 구성요소인 경우.

그래서 관련관계가 되는 전형적인 경우를 일반적으로 관련표준에서 나열하고 있다. 그러나 실제의 상황에서는 관련관계의 설정여부에 모호한 경우가 많다. 첫째, 고유명사와 관련된 용어이다. 예컨대, 인명과 관련하여 출생지, 소속단체, 학과, 학설, 수상 등, 지명과 관련하여 유적, 사건, 시설, 법률 등을 관련지워야 하는지 판단이 어렵다. 둘째, 분야에 따라 다른 개념으로 혹은 동일한 개념으로 사용하기도 한다. 예컨대, ‘강간/강간죄’, ‘부모폭력/부모구타’를 법률분야에서는 동의어로 사용한다. 그러나 개념 자체를 보면 동의어로 보기가 어렵다. 셋째, 상이하거나 관련 있는 개념이 특정 핵심개념과 결합하면 의미가 같거나 유사하게 되는 경우가 많다. 예컨대, ‘법, 법률, 법령’이 ‘개정’과 복합어가 되는 경우, ‘국제, 대외, 세계, 외국, 해외’가 ‘무역’과 복합어가 되는 경우, ‘부패, 부정, 불법, 위법, 탈법’이 ‘선거’와 복합어가 되는 경우에는 동의어로 설정하는 경우가 많다. 넷째, 상하관계로도 볼 수 있는 밀접한 관계가 있다. 예컨대, ‘국적/무국적, 대사/명예대사’가 그것이다. 일반적으로 수식어가 특정 개념 앞에 오면 계층관계, 뒤에 오면 관련관계인 경우가 많다, 따라서 ‘무국적’은 ‘국적’의 하위개념어로, ‘명예대사’는 ‘대사’의 하위개념어로 볼 수도 있으나 서로 관련관계로 보는 것이 타당할 것이다.

시소러스에서 1차 개념관계를 어떤 관점에서 설정했는가를 명시하고 있지 않기 때문에 이를 파악하기가 매우 어렵다. 예컨대, ‘남편’과 ‘아내’의 관계는 일반적으로 관련관계(RT)로 설정하고 있으나 측면을 추측해보면 ‘부부, 남편/아내, 남자/여자(남녀), 애인, 배우자, 대응(반의)’ 등 여러 가지 관계를 생각할 수 있다.

관련관계의 확장을 위한 관계지시어를 광범위하게 분석하여 이를 범주화하고 표준화할 필요가 있다. 본 시소러스에서는 그에 대한 시험을 계속하고 있으며 현재 “대응, 구성요소, 기능, 도구, 부속, 수단, 용도, 유형, 처리” 아홉 가지를 정의하고 있다. 특히 대응관계가 세 개 이상의 용어 간에 형성되는 경우가 있다. 즉, 셋인 경우는 ‘상, 중, 하’, ‘전, 중, 후’ 등이 있으며, 넷인 경우는

‘동, 서, 남, 북’ 등이 있다. 본 시소러스에서는 셋까지 대응관계로 설정하고, 넷인 경우에는 이들을 분리하여 ‘동/서, 남/북’을 각각 대응관계로 설정하고 있다. 경우에 따라 이 관계는 같은 개념의 하위개념어 간(형제용어)에는 원칙적으로 참조관계를 만들지 않는다는 관례에 위배될 수도 있으므로 주의를 요한다.

[예] 개방 사회 RT [대응] 폐쇄 사회

차변 RT [대응] 대변

평등 RT [대응] 불평등

제1 차액 지대 RT [대응] 제2 차액 지대

제1 차액 지대 RT [대응] 제3 차액 지대

시어터 시스템 RT [구성요소] AV 리시버

컨버터 RT [기능] 신호 변환

네트워크 RT [도구] 프로토콜

상거래 RT [수단] XML 문서

컴퓨터 RT [부속] 메모리

4.4 속성관계

속성관계는 주로 인명, 지명, 기관/단체명, 작품(문학, 음악, 예술), 상품과 같은 고유명사가 대상이 되지만 시소러스에 등록된 모든 용어는 속성을 가질 수 있다.

그림 3은 용어와 속성 간의 관계를 보인 것이다. 이 관계는 개념적으로는 '(frame_name(slot(facet value)))' 혹은 '(frame_name(slot value))'의 프레임구조와 일치한다. 예컨대, 지명이라면 계층적인 속성을 가질 수 있다. 속성은 다른 용어의 속성이 될 수 있으며, 다른 속성의 속성이 될 수 있다. 뿐만 아니라 그 속성 혹은 속성의 값이 용어로서 시소러스에 포함될 수도 있다. 이와 같은 구조는 시소러스가 있고 시소러스에 등록되어 있는 용어에 대한 각 속성이 메타데이터로서의 역할을 하고 있기 때문이다. 이와 같은 구조를 가질 때 관련되는 지식은 어떤 것이라도 표현할 수 있으며 궁극적인 온톨로지가 될 것이다. 다만, 시소러스에 용어로서 등록되어 다른 용어와의 개념관계를 가져

야 할 속성을 식별하는 것이 선행되어야 한다.

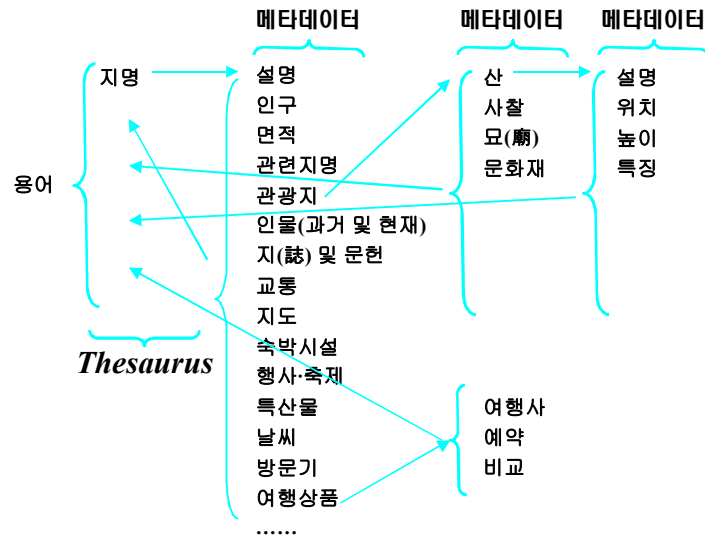


그림 3 용어와 속성의 관계

속성의 종류는 각 대상에 따라 다르며, 현재의 시소러스에서는 다음과 같이 제한적으로 적용하고 있다. <일반용어>인 경우에는 용어의 정의, <인명>인 경우에는 다른 이름(호, 자, 시호, 아명, 본명, 필명, 속명, 법명, 별명), 약력, 생몰년, 출생지, 저작, 수상, <지명>인 경우에는 설명, <저작>인 경우에는 설명, 저자, 저작년, 장르(문학, 음악, 예술분야에 따라 세분), 외국어명, <기관/단체명>인 경우에는 설명, 설립년을 기술하고 있으며 이들을 확장할 계획을 가지고 있다. 다만, <상품>의 속성은 그 종류가 너무나 방대하여(본인이 조사한 바에 의하면 공통속성을 제외하고 10,000종 이상이 된다) 이의 적용을 유보하고 있다.

확장된 시소러스의 예를 그림 4에 보인다.

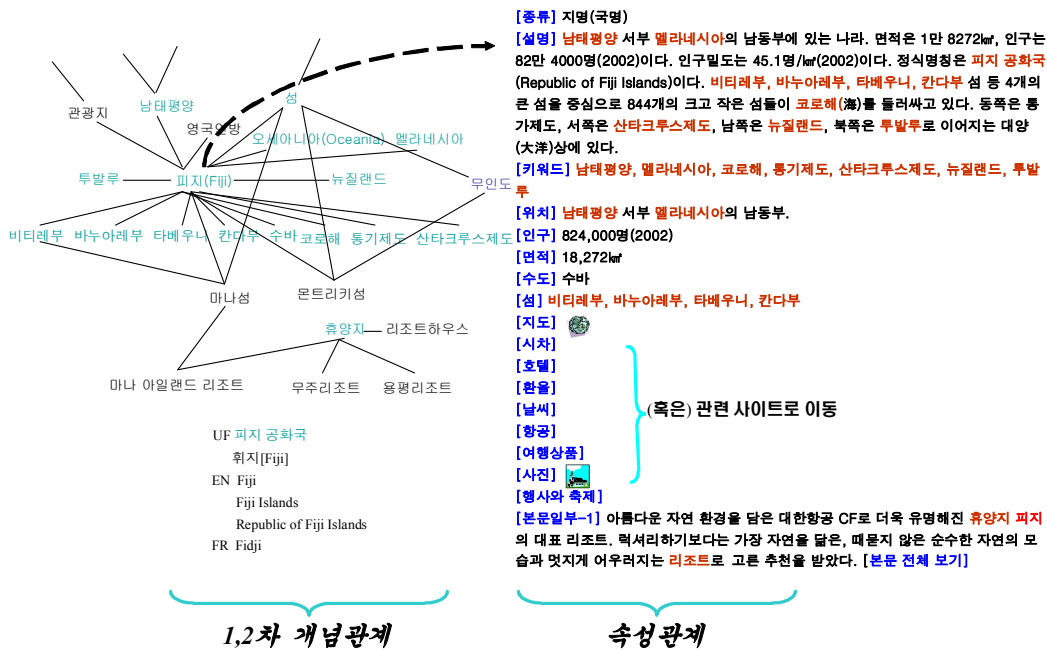


그림 4 확장된 시소러스의 예

5 문제점 및 추후 계획

정보처리에 대한 새로운 요구가 증가함에 따라 기존 시소러스에 새로운 정보를 추가할 수밖에 없다. 결국 시소러스는 소위 온톨로지로 진화하고 있으며 그 중간단계에 서 있다. 본 시소러스는 도서관과 같은 정보처리기관에서 색인과 검색용 주제명표목표로서 충분히 이용될 수 있으며 이용되고 있다. 특히 사용기관별 집합도 개별 또는 통합적으로 관리될 수 있으므로 해당 기관의 상황에 맞추어 용어집합을 조정할 수 있는 이점이 있다. 또한 지금까지 준비해 온 용어와 의미관계정보는 온톨로지로의 변환이 가능하며 보다 효율적이고 광범위하게 이용될 수 있을 것이다. 필요하다면 단위 레이어별 및 그룹별 온톨로지로 분할할 수 있으며, 분류기호를 이용하면 분야별 및 용도별 온톨로지를 만들 수 있어서 기본 온톨로지로서의 역할을 할 수 있을 것이다. 또한 이 모형을 이용하면 분야에 구애되지 않고 대부분의 필요 지식을 조직적으로 표현할 수 있으리라 생각된다.

그러나 지금까지의 처리결과 보완해야 할 다음과 같은 문제점이 노정되고 있다. 첫째, 용어수의 획기적인 확장이 필요하다. 현재의 52만 용어로는 지식의 겨우 일부를 표현할 수 있을 뿐이다. 둘째, 다양하고 표준화된 개념관계명이 필요하다. 특히, 관련관계의 개념관계가 복잡하고 모호하다. 이를 위해서는 용

어의 의미관계에 대한 깊이 있는 연구가 요구된다. 표준화된 개념관계명의 개발은 온톨로지 개발의 가장 기본적이고 필수적인 과제이다. 셋째, 속성관계는 실제적인 지식을 표현하는 데 필수적이다. 그러나 속성의 종류가 너무나 다양하여 그 종류를 파악하고 분류하기가 매우 어렵다.

현시점에서 지금까지 파악된 문제점을 보완하고 있다. 특히, 용어의 확장, 개념관계명 집합의 객관화, 속성관계의 분류 등이 우선되고 있다. 온톨로지 표현언어는 W3C RDF/OWL(Smith et. al. 2004)과 ISO XML/Topic Maps(ISO/IEC 13250:2003) 둘 다 사용하기로 하였다. 용어 데이터는 시소러스 형식을 그대로 유지하여 전통적인 시소러스의 용도로 이용할 수 있도록 하고, 이 시소러스를 중간언어로 삼아 언제라도 RDF/OWL나 XML/Topic Maps 인터페이스를 통하여 온톨로지로 이용할 수 있도록 시스템을 구축하고 있다.

참고문헌

- 溝口理一郎. 2005. 『オントロジー工学』. 東京 : OHM.
- 국립국어연구원. 1999. 『국어 순화 자료집』(1999-1-4).
- 김인균. 2005. 『국어의 명사 문법 I』. 서울 : 도서출판 역락.
- 김종복. 2004. 『한국어 구구조문법』. 서울 : 한국문화사.
- 이재철. 1961. 『주제명표목표』. 한국도서관협회.
- 中國圖書館圖書分類法編委會編. 1994. 『中國分類主題詞表』. 北京 : 華藝出版社.
- 최석두. 1993. 無典據시스템에 관한 研究. 『한국문헌정보학회지』, 25: 233-264.
- 황순희, 정한민, 성원경. 2006. 패싯(Facet)을 이용한 과학기술분야 시소러스 구축과 활용방안. 『정보관리연구』, 37(3): 61-84.
- ANSI/NISO Z39.19-2003 *Guidelines for the construction, format, and management of monolingual thesauri.*
- Berners-Lee, Tim, James Hendler and Ora Lassila. 2001. "The semantic web." *Scientific American*, (May): 35-42.
- BSI. 1988. *ROOT Thesaurus*. Vol. 1-2. BSI.
- Choi, Suk-Doo. 2006. "Ontology development using thesaurus." *The Second International Symposium on Knowledge Processing and Service for China, Japan and Korea: Metadata and Ontology*, October 15-16,

- 2006, Beijing, China. 43–54.
- Clark, P., John Thompson, Heather Holmback, and Lisbeth Duncan. 2000. "Exploiting a thesaurus based semantic net for knowledge-based search." In *Proc 12th Conference on Innovative Applications of AI(AAAI/IAAI'00)*, 988–995.
- Fillmore, C. J. 1967. "The case for case." In E. Bach and R. Harms, eds. *Universals in Linguistic Theory*, 1–88. New York : Holt.
- ISO 639-2:1998 *Codes for the representation of names of languages: Alpha-3*.
- ISO 2788:1986(E) *Documentation – Guidelines for the establishment and development of monolingual thesauri*.
- ISO 5964-1985(E) *Documentation – Guidelines for the establishment and development of multilingual thesauri*.
- ISO/IEC 13250:2003 *Information technology -- SGML applications -- topic maps*.
- Kawtrakul, Asanee, Aurawan Imsombut, Aree Thunyakijjanujit, Dagobert Soergel, Anita Liang, Margherita Sini, Gudrun Johannsen, and Johannes Keizer. 2005. "Automatic term relationship cleaning and refinement for AGROVOC." *2005 EFITA/WCCA 2005*, 1146–1154.
- Library of Congress. 1992. *Library of Congress Subject Headings*. 15th ed.
- Smith, Michael K., Chris Welty, and Deborah L. McGuinness. 2004. *OWL Web Ontology Language Guide*. W3C Recommendation 10 February 2004.
- Soergel, D., B. Lauser, A. Liang, and F. Fisseha. 2004. "Reengineering thesauri for new applications. The AGROVOC example." *Journal of Digital Information*, 4(4). <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Soergel>.
- Unesco. 1977. *Unesco Thesaurus*. Vol. 1–2. Paris : Unesco.
- Wielinga, B.J., A. Th. Schreiber, J. Wielemaker, and J.A.C. Sandberg. 2001. "From thesaurus to ontology." In *Proceedings of the International Conference on Knowledge Capture*. ACM Press, 194–201.