

시소러스와 온톨로지의 상호 호환성에 관한 연구

A Study on the Interchangeability between a Thesaurus and an Ontology

조 현 양(Hyun-Yang Cho)*

남 영 준(Young-Joon Nam)**

초 록

본 연구에서는 다국어 시소러스에 수록된 디스크립터간의 관계를 온톨로지 언어로 표현하는 실험을 수행하였다. 대등관계는 equivalentClass와 equivalentProperty, sameAS 등으로, 그리고 연관관계는 ObjectProperty 을 비롯하여 DatatypeProperty, inverseOf 등으로 표현할 수 있었다. 이러한 언어를 기반으로 실제 AAT에 배정된 디스크립터와 ICCD에서 구축한 다국어 시소러스의 디스크립터를 대상으로 디스크립터의 한글화 작업이 수행되었으며, 다국어간 개념일치를 위해 패싯개념이 시소러스 구조에 이용되었다. 본 연구를 통한 실험의 결과 다국어 시소러스를 온톨로지로 표현하기 위해서는 속성관련 온톨로지 언어를 사용하는 것이 가장 효과적임을 확인할 수 있었다.

ABSTRACTS

In this study, the experiment was made to transform the relationship among terms in a thesaurus to ontology language as search tools for multilingual text. As a result, the equivalent relationship in the thesaurus can be expressed by different ways in the ontology, such as equivalentClass, equivalentProperty, sameAS, and so on. On the other hand, the associative relationship can be represented by ObjectProperty, DatatypeProperty, and inverseOf. For this test, first of all, the descriptors assigned by AAT and the descriptors from bilingual thesaurus by ICCD were translated into Korean. Then, the facet was used for conceptual equivalence among terms from different languages. The result of the study showed that using rdf:Property in ontology was the most effective way of transforming multilingual thesaurus into ontology.

키워드: 시소러스, 용어 관계설정, 온톨로지, OWL, 패싯 시스템
thesaurus, term relationships, ontology, OWL, facet system

* 경기대학교 문헌정보학과 교수(hychol180@yahoo.co.kr)

** 중앙대학교 문헌정보학과 교수(namyj@cau.ac.kr)

■ 논문접수일자 : 2004년 10월 15일

■ 게재확정일자 : 2004년 11월 26일

1. 서론

도서관은 전통적으로 모든 종류의 정보들을 수집·관리하는 역할을 수행해 왔으며, 책자형태와 같은 물리적 매체위주의 수집과 이를 기반으로 하는 정보서비스가 이루어졌다. 그러나 인터넷과 첨단 정보기술을 이용한 새로운 정보제공기관들의 등장으로 과거와 같은 물리적 매체 위주의 자료를 수집, 관리, 제공하는 수준에 머문다면 도서관은 최고의 정보서비스 기관으로서의 입지를 위협받게 될 지도 모른다. 이러한 위기감과 더불어 다양한 형태의 매체와 방대한 양의 정보를 효율적으로 관리하기 위해 모든 도서관은 새로운 정보기술을 도입하고 주제 영역이나 기관별로 최적의 정보서비스 모델을 개발해 왔다.

또한 도서관은 상호협력 체제를 통한 정보의 상호교환 및 공동 활용을 위하여 일찍부터 정보의 조직화 및 데이터베이스화와 관련하여 최소한의 표준이 이루어질 수 있도록 MARC와 같은 표준 메타데이터 등을 채택하여 사용하고 있다.

이러한 표준 메타데이터는 대부분 자관 소장 자료의 편목과 서지데이터의 교환을 위한 도구로 사용하고 있으며, 실제적인 검색 서비스를 제공함에 있어 검색의 효율성을 증대시킬 수 있는 시소러스와 같은 보조 도구는 제한적으로 이용되고 있는 실정이다. 그나마 검색 보조도구로 기존에 사용되고 있는 시소러스는 모든 주제 분야에 적용되어 검색지식으로 활용될 수 있는 매크로 형태보다 특정 주제 분야에서 사용될 수 있는 마이크로 형태로 개발·운용되고 있다. 이러한 검색도구는 기본적으로 자

체 제작한 데이터베이스의 검색에서 최적의 효율을 얻을 수 있도록 개발되고 있는 것이 현실이다. 그러나 인터넷과 같은 네트워크가 활성화된 실시간 정보교환시대에서는 다양한 형태의 데이터베이스를 종합적으로 검색하기 위한 요구는 증가할 수밖에 없다. 따라서 네트워크로 연결되어 있는 모든 데이터베이스를 통합적으로 수용할 수 있는 매크로 데이터베이스의 검색에 적합한 시소러스나 혹은 새로운 검색도구의 개발이 필요하다.

본 연구의 목적은 정보검색에 있어서의 효율성을 향상시키기 위하여 사용하는 시소러스와 같은 검색용 도구들이 독립성을 유지하면서 최소한의 통합검색이 가능하도록 온톨로지 개념을 이용하는 방안을 모색하는 것이다. 예를 들면, A라는 데이터베이스에서 main entry에 대한 개념을 '기본저록'이란 용어를 사용하고, B라는 데이터베이스에서는 이를 '기본기입'이라고 사용할 경우에, '기본저록'과 '기본기입'이 동의어 관계에 있다는 용어간의 관계를 이용하여 복수의 데이터베이스를 검색할 수 있는 알고리즘을 개발하는 것이다. 이러한 추론은 다국어 전자자원을 대상으로 온톨로지 기반의 시소러스 검색에 유용하게 활용될 수 있을 것이다.

2. 시소러스의 의미구조

2.1 시소러스의 역할

KDC나 DDC와 같은 분류체계는 수록된 표제항의 개념들을 이차원의 계층구조로 전개하

고 있다. 또한 전통적으로 주제명 표목표에서 표목은 색인어가 될 수 있는 주요어를 가나다 순으로 전개하고, 서로 유사성이 높은 표목을 대등관계로 표시할 것을 권고하고 있다. 상기에 언급된 이 두 개의 도구들은 원래 축적된 정보에 대한 검색보다는 색인작업을 위한 용도로 개발되었다. 따라서 주제명 표목표나 분류 체계가 갖는 이차원적인 개념구조로는 이용자의 검색에 대한 높은 기대 수준을 제한적으로 밖에 충족시킬 수 없다.

시소러스는 기본적으로 도서관 분류체계나 주제명 표목표로써 표현하기 어려운 검색개념을 다차원적인 용어 구조를 이용하여 효과적으로 검색할 수 있도록 개선한 통제어휘집의 일종이다. 따라서 시소러스는 키워드간의 관계를 다차원적인 형태로 구조화한 새로운 검색도구로 활용하였다. 주제명 표목과 분류표가 주제어의 나열과 함께 특정 영역을 구분 짓는 기준으로서의 역할이 강조된다면, 시소러스는 주제어를 의미적 관점에서 구조화한 것으로 의미에 기반한 개념 검색에 초점을 맞출 수 있다. 검색을 특정 표목집에 등재된 검색어(주제명 표목)에 기반하여 검색하던 전통적 방법에서, 시소러스를 이용한 검색은 검색된 결과에 대한 이용자의 만족도에 따라 검색의 관점이나 범위 등을 고려하여 개념의 폭을 조정하는 것이 가능하다.

에치슨 등(Aitchison etc. 2000)은 시소러스의 역할을 첫째, 문헌자료에 대한 정보검색 도구로, 둘째, 특정분야를 이해하기 위한 일련의 의미지도(semantic maps)로써 개념간의 상호관계를 표시하고, 용어간의 정의를 이해할 수 있도록 지원하는 것으로 정의하였다.

라시(Lassi 2002)는 시소러스의 역할에 대해 다음과 같이 설명하였다.

- 용어가 갖고 있는 혹은 사물이 갖고 있는 개념간의 관계구조화 사전
- 의미적으로 구조화하여 검색행위시의 검색어 선정 역할과 검색에 필요한 개념을 확장할 때 지식으로 활용될 수 있는 의미사전

ANSI는 시소러스 개발 원칙(2003)에서 시소러스의 역할을 다음 4가지로 정의하고 있다.

- 저자나 색인이거나 혹은 이용자들이 가지고 있는 자연어로 표현되는 개념을 정보검색을 위해 최적화된 통제어로 변환하는 역할
- 부여된 색인어의 일관성을 유지할 수 있도록 지원
- 용어간의 의미적 관계를 지시하기 위한 역할
- 문헌검색을 위해 탐색도구로서의 역할

또한 ISO는 시소러스 구축지침(1986)을 통하여 시소러스의 기능은 문헌의 주제를 결정짓기 위한 기능, 주제를 요약할 수 있는 용어를 선정하는 기능, 그리고 이들 용어로 표현된 개념간의 관계를 지시하는 기능 등의 3가지로 제시하고 있다.

이상에서 제시된 시소러스의 기능 및 역할을 종합해 보면 시소러스는 정보검색의 보조 도구로서 개념별로 검색과 색인 작업시 대표어휘의 선정을 위한 사전의 역할과 의미에 기반한 용어간의 관계나 구조를 보여주는 사전으로서의 역할을 수행하는 것으로 볼 수 있다.

2. 2 주요 권고안의 관계설정규칙

시소러스에서 용어간 의미 구조화와 다차원적인 검색을 위해 디스크립터들은 기본적으로 계층관계를 비롯하여 연관관계, 대등관계 등으로 구조화하고 있다. 이러한 기본적인 구조화 원칙은 국제 및 국가 표준으로 제정된 권고안의 형태로 제안되어 있으며, 실제 개발된 시소러스는 이러한 기준을 근거로 많은 부분을 개발 주체기관에서 필요한 특성에 따라 자의적으로 반영하고 있다.

국내의 주요 권고안에서 제안하는 관계설정을 위한 구조화의 틀을 살펴보면 다음과 같다.

2. 2. 1 한국데이터베이스진흥센터 개발지침

시소러스 구축에 대한 우리나라 국가 표준으로는 한국데이터베이스진흥센터(Korea Database Promotion Center)에서 시소러스 개발지침(한국데이터베이스진흥센터 2000)을 제시하였다. 이 지침에서는 ISO 2788에서 제안하는 관계설정의 유형과 특성을 고려하여 시소러스 관계를 다음과 같이 등가관계와 계층관계, 관련관계 등 세 가지 요소로 구분하여 설명하고 있다.

1) 대등관계: 대등관계는 등가관계로도 표현되며, 이 관계는 색인작업시 복수의 용어가 동일개념을 나타낸다고 인정되는 경우에 우선어 및 비우선어간의 관계이다. 따라서 우선어와 비우선어는 등가어의 집합을 형성한다. 책자형 시소러스에서의 자모순 배열은 보통은 우선어와 비우선어를 인쇄 서체로 구별한다. 이에 속하는 것으로는 유사어와 유사동의어, 업워드포스팅으로 구분한다. 이 가운데 동의어는

광범위한 문맥에서 의미가 동일하다고 인정되기 때문에 실제로 교환 가능한 용어로 정의된다. 동의어는 일상적인 언어에는 그다지 출현하지 않으나, 과학기술 분야를 비롯한 주제 분야 전체에 걸쳐서 자연어보다 통제언어 쪽에 동의어가 많다. 이것은 통제언어에서도 의도적으로 용어의 의미를 제한하기 때문이다. 유사동의어는 근사동의어라고도 하며, 일반적으로 의미가 상이라고 인정되나, 색인작업에서는 동의어로 취급하는 용어라고 정의된다. 유사동의어에는 의미가 현저하게 중복되는 용어를 포함한다. 업워드 포스팅은 하위개념어를 상위개념어의 종(種)의 관계보다 등가 관계로 취급하는 방법이다. 그 목적은 어휘의 수를 감소시키고 동시에 하위개념인 특정용어에서 하위개념을 나타내기 위하여 이용하는 상위개념어에 액세스하기 위한 방법을 확보하는데 있다.

2) 계층관계: 계층관계는 시소러스 디스크립터 용어를 상·하위 개념으로 연결시켜주는 관계를 말한다. 상위개념은 클래스 또는 전체를 나타내며, 하위개념은 한 요소 또는 일부분을 나타내는 것이다. 계층관계는 상위 및 하위개념을 논리적으로 전개하는 순서를 나타낸다. 이에 속하는 관계는 “속(屬)관계”를 비롯하여 “계층적인 전체-부분관계”, “사례관계” 등 세 가지를 제시하고 있다.

3) 연관관계: 연관관계란 계층적이 아니지만 개념적으로 밀접하게 관련되어 있으나, 등가집합에는 포함되지 않는 용어간의 관계이다. 관련관계에 있는 용어 즉, 관련어는 색인작성과 탐색에 이용될 가능성이 있는 대체용어를 제시한다는 관점에서 볼 때, 용어간의 연결을 시소러스 중에서 명시하는 것이 보다 좋을 것이라는

심리적인 연상에 따라 만들어진 것이다. 이 관계는 상호보완적인 관계로 설명하고 있다. 이 관계에는 속하는 것으로는 “동일 범주에 속하는 용어”를 비롯하여 “상이한 범주에 속하는 용어”로 두 개의 경우를 열거하고 있다.

2. 2. 2 ISO의 가이드라인

시소러스를 구조화하기 위한 대표적인 국제 가이드라인은 ISO에서 제정한 ISO 2788(ISO 1986)과 ISO 5964(ISO 1985)가 있다. 전자는 단일어로 구성되는 시소러스를 구성하는데, 그리고 후자는 다국어 시소러스를 구성하는데 원용된 표준이다.

1) 유사관계: 이 관계는 동의어와 유사동의어로 구분할 수 있다. 이 관계는 색인어 선정에 있어 어느 것을 선정해도 무리가 없는 관계의 용어들을 균집화할 때 필요한 용어간 관계로 정의한다. 다국어 시소러스의 경우, 디스크립터로 선정되는 대표어와 이에 대응하는 외국어를 비디스크립터어로 선정할 경우가 이 관계에 해당한다.

2) 계층관계: 이에 속하는 관계로는 ISO 2788과 ISO 5964에서 제시하는 두 가지 이론에 속하는 경우에만 활용한다. 하나는 속관계이며, 하나는 계층개념으로 전체와 부분관계로써 용어의 동일한 범주에 속하는 것이다. 예를 들면, 지리적인 공간개념과 신체와 신체 일부와의 관계 등이 이에 속한다.

3) 연관관계: ISO의 가이드라인에서도 연관 관계 설정이 다른 관계의 개념적 설정이 상대적으로 어렵다고 기술하고 있다. 따라서 연관관계에 대한 분명하고 주관적인 기준을 제시하기보다는 유사관계나 계층관계에 속하지 않

는 용어간 관계들은 대부분 연관관계에 속한다고 정의하고 있다.

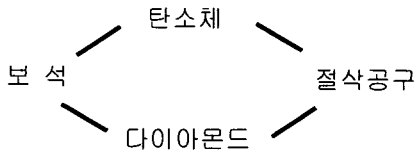
2. 2. 3 ANSI의 가이드라인-2003

미국 국립표준원(American National Standard)에서 단일어 시소러스 구축을 위해 제안한 가이드라인으로써 1980년에 UNESCO 가이드라인의 기초가 된 ANSI/NISO Z39.19을 기반으로 작성한 것이다. 이는 2003년에 디지털 정보자료 검색을 위해 많은 부분을 수정하여 개정판(NISO 2003)에서 새롭게 제안되었다. 이 가이드라인에서의 가장 큰 변화는 전체-부분관계를 연관관계보다 계층관계적 시각으로 처리하도록 권고한 점이다. 기본적인 관계로는 다른 가이드라인과 같이 대등관계를 비롯하여 연관관계, 계층관계로 구분하고 있으나 해당 관계에 속하는 사례부분에서 의미적 연관성을 중시하여 다른 개념어와 연관 관계를 갖지 않는 고립어를 인정하지 않고 있다.

1) 대등관계: 동일한 개념을 두 개 이상의 단어로 표현하는 용어간의 구조화에 사용되는 관계이다. 이 관계에 속하는 사례로써는 동의어관계와 어휘파생어(Lexical Variants)로써 이형동의어의 범주를 포함한 구체적인 관계, 유사어, 복합명사 가운데 상호참조가 가능한 용어관계로 확대하여 제안하고 있다.

2) 계층관계: 이 관계에 속하는 것은 용어간 소속이나 포함범주에 의해 분명하게 상위개념어와 하위개념어로 구분되는 관계이다. 이 관계에는 속관계를 비롯하여 전체와 부분관계, 사례관계, 다계층관계로 제안하고 있다. 특히 다계층관계는 하나의 개념이 시소러스 내에서 복수의 계층관계를 갖는 경우이다.

〈그림 1〉은 다이아몬드가 기계시소러스내에서 보석(gem)과 절삭공구(cutters)라는 복수의 상위어를 가지며 보석과 절삭공구는 탄소체(carbons)라는 단일 상위어를 갖는 예이다.



〈그림 1〉 기계 시소러스 예

3) 연관관계: 이 관계는 대등관계와 계층관계에 속하지 않는 관계로써 개념의 크기와 용어 수준이 대칭이 되는 관계에 속하는 용어간 관계가 이에 해당한다. 이 관계로 구조화하기 위한 우선 조건으로 다음과 같은 경우를 제안하고 있다.

- 동일한 범주에 속해있는 용어간에 설정할 수 있다.
- 다른 범주에 속해있으면서 동시에 출현하여 색인과 검색이 도움이 되는 용어간에 설정할 수 있다.
- 관련어의 노드레이블(node labels)로 검색이나 색인에 연상이 되는 용어간에 설정할 수 있다.
- 기계가독형 색인어집에서 특정 의미를 갖는 특정기호로 처리될 수 있는 용어들을 연관관계로 설정할 수 있다. 예를 들면, 국가코드가 붙는 용어는 해당 국가코드를 기준으로 서로 연관관계가 설정될 수 있다.

2. 2. 4 UNESCO 지침

UNESCO 시소러스 구축 지침은 SC/76/

WS/555의 구축 원칙을 확대 개선한 것이다. 이 지침은 오스틴(D. Austin)과 데일(Peter Dale)이 ISO 2788을 개발하기 이전에 영국도서관에서 시소러스를 우선 구축할 수 있는 기준으로 활용하였다. 이 가이드라인에서는 디스크립터의 형태에 관한 것을 주로 제안하고 있으며, 해당 디스크립터 관계설정은 크게 다음과 같은 네 가지 관계를 제안하고 있다 (UNESCO 1981).

1) 대등관계: 이 관계는 우선어와 비우선어간의 관계로써 비우선어는 복수로도 설정할 수 있다. 이 관계의 유형으로는 동의어를 비롯하여, 유사동의어, 업워드포스팅이 있다. 이 가운데 업워드포스팅 관계는 색인어를 고려한 관계로써 검색상황에 따라 계층관계의 용어도 이에 속할 수 있다.

2) 계층관계: 이 관계는 비구조화된 용어군 가운데 가장 분명하게 계층관계를 설정할 수 있는 용어들의 관계이다. 이 관계는 최상위어를 선정하고 해당 최상위어에 속해있는 용어들을 파생하위어(subordinate terms)로 선정하는 방법을 택한다. 이에 속하는 관계로는 “속성관계”를 비롯하여 “계층적인 전체-부분 관계”, “사례관계(instance relationships)”를 제안하고 있다.

3) 연관관계: 이 관계는 대등관계나 계층관계에 속하지 않는 용어 가운데 서로 군집화하여 색인이나 검색에 도움이 되는 관계를 의미한다. 이 관계는 다른 용어를 의미적으로나 계층적으로 대체하지 못하는 관계이다. 이 관계에 속하는 것으로는 “동일한 카테고리에 속해 있는 용어”와 “다른 카테고리에 속해 있는 용어”로써 서로 관련성이 있는 경우를 제안하고 있다.

이 관계를 일련의 제안된 형태로 표현하는 것은 어려움이 있기 때문에 이 가이드라인에서는 9가지의 구체적인 예를 들어 설명하고 있다.

4) 상호범주 관계(Inter-category relationships): 이 관계는 일련의 용어들이 속하게 되는 범주가 복수에 해당하는 용어들이 이에 해당한다. 이러한 관계가 설정될 수 있는 것으로는 다음 두 가지를 설정하고 있다.

- 특정 주제 분야나 기초이론들 사이에서 기본이 되는 대상
- 패싯체계에서 기본이 되는 대상

이와 같은 시소러스 구조화 틀은 외형적으로는 계층관계와 연관관계, 대등관계 기호를 사용하고 있으며, 각각의 관계 내에서는 해당 영역에 속하는 유형들에 대해 자세한 속성을 설명하고 있다. ANSI(2003)를 비롯한 대부분 가이드라인은 계층관계에 속하는 것으로써 속성 관계를 비롯하여 전체와 부분의 관계, 사례관계 등과 같이 패싯개념을 도입하고 있다. 즉, 디스크립터의 계층에 대한 구조화 기호 가운데 계층의 일반적 관계이외에 포괄 속성(generic)을 나타내는 것과 사례적 속성(instance), 전체-부분(partitive)을 사용하고 있다. 또한 최상위어를 의미하는 TT(top term)기호를 사용하도록 하여, 특정개념에 대한 의미의 전개 점을 확인할 수 있도록 하였다.

3. 온톨로지의 의미구조

온톨로지는 특정 분야의 지식을 표현하기 위해 클래스를 비롯하여 관계, 함수(Functions),

공리(Axioms), 인스턴스의 다섯 가지 요소를 사용한다.

특정 개념을 정의하는 과정은 해당 개념이 갖고 있는 작업영역과 기능, 행동, 전략, 추론 절차 등을 기술하는 행위를 의미한다. 또한 해당 개념에 대한 공식적이고 명시적인 설명을 구조화하는 것이 클래스이다. 따라서 해당 영역의 개념을 설명하는 클래스는 온톨로지에서 가장 중요한 요소로 꼽힌다. 기본적으로 클래스는 계층적으로 상위클래스와 하위클래스로 구성된다.

클래스의 특징을 나타내는 것을 속성이라고 하며, 이를 슬롯이나 역할 기호로 표현할 수 있다. 즉, 속성(슬롯)은 클래스와 인스턴스의 특성을 설명하는 역할을 수행한다. 속성에는 제한사항이 있으며, 이것을 패싯이라고 한다. 이를 기반으로 개념은 관계, 함수, 인스턴스, 공리로 분류되어 조직화된다.

관계는 특정 개념 사이의 상호작용 형태를 나타내고 함수는 관계 중 특별한 경우를 뜻하는 것이다. 공리는 항상 참인 명제를 모델화하는데 사용되고, 인스턴스는 요소를 설명하기 위해 사용된다.

3.1 온톨로지의 인지적 역할

온톨로지는 용어의 의미나 개념간의 관계를 인간이나 컴퓨터 모두 인식할 수 있도록 하여 용어간의 관계를 특정한 딜리미터로 표현하고 있다. 이 기호는 컴퓨터를 이용한 기계처리가 가능하도록 선언하고, 기호를 실생활 용어로 사용함으로써 코딩된 구조화 자료를 육안으로도 관찰과 이해가 가능하도록 하였다.

온톨로지를 인지적인 측면에서 역할을 살펴 보면 다음과 같다.

1) 온톨로지는 특정 주제영역에서 사용되는 일반적인 용어를 제공한다. 주제 영역에 있어 공유된 개념을 상호 공유하여 사용하는 것은 온톨로지의 기본적인 역할이다. 또한 온톨로지는 데이터베이스 구축과정에서 개념적 스키마 역할을 수행한다. 즉 온톨로지는 정보의 설명이나 교환을 위한 적정 데이터 구조를 제공하는 지식이다. 따라서 온톨로지는 문제해결에 필요한 함축적인 전제조건과 가정뿐만 아니라 그런 가정들을 반영하는 주요 객체의 개념화를 설명할 수 있다.

2) 온톨로지는 지식의 체계화를 지원한다. 이것은 인간이 인지하고 기술·설명하는 현상, 이론들을 개념, 용어로 개념을 구조화하는 도구이다. 여기서 온톨로지는 지식 체계화의 주요 근간으로 활용될 수 있다. 따라서 시맨틱 웹에서 사용된 메타데이터는 각 태그 의미와 가치 값을 부분적으로 정의하는 온톨로지를 기반으로 개발된다. 따라서 메타데이터의 해석과 전환은 온톨로지를 통해서 이루어 질 수 있다. 이렇게 온톨로지는 메타데이터 간의 의미적 상호운용성을 보장하는 중요한 요소로 작용한다.

3) 온톨로지는 특정 영역이나 개념에 대해 함축적인 지식을 표현할 수 있다. 모든 인간 활동은 함축적인 가정을 갖고 있기 때문에 정형화된 포맷들은 함축적인 가정 속에서 일반적인 기본 단어를 비롯하여, 관계 및 제한사항에 대한 정의를 찾아내서 온톨로지에 포함시킨다. 지식기반이라 함은 개발자가 갖고 있는 개념을 기반으로 만들어지기 때문에, 인간 개념은 함축적일 수밖에 없다. 이런 개념과 가정을 명시

적으로 표현하는 것은 단순한 설명의 차원을 넘어선다. 따라서 온톨로지를 활용하면 지식의 공유와 재사용을 어렵게 하는 중요한 요인이었던 함축성을 설명할 수 있다.

3. 2 웹 온톨로지의 언어

대표적인 웹 온톨로지의 언어 가운데 DAML+OIL의 기술적 지식을 바탕으로 클래스와 속성 간의 관계 정의 기능을 강화한 것이 OWL (Web Ontology Language)이다. OWL은 DAML+OIL를 확장한 것으로, 기존의 온톨로지 언어에 비해 속성과 클래스를 좀 더 상세하게 기술할 수 있도록 상대적으로 풍부한 어휘를 제공하고 있다. 하위언어로는 OWL Full과 Lite, DL로 구분한다. 최근에는 OWL Lite와 OWL DL의 장점만을 조합하여 Semantic Web Rule Language(SWRL)을 개발하여 Protege에 적용하고 있다. 다음은 OWL로 표현할 수 있는 기본 특성이다.

3. 2. 1 RDF 스키마 특성

1) Class

클래스는 특정 속성을 공통적으로 갖는 개체(individuals)를 군집화하여 이를 정의한 것이다. 표현은 'subClassOf'를 사용하여 클래스의 계층구조를 형성할 수 있다. 현재 정의된 클래스 가운데 가장 일반적인 클래스는 'Thing'이다. 'Thing'은 모든 개체를 포함하는 클래스이며, 모든 OWL 클래스들 중의 최상위 클래스가 된다. 본래 정의되어 있는 클래스 중 최하위 클래스는 'Nothing'이다. 이에 대한 예를 들면 다음과 같다.

예) 와인 영역의 온톨로지를 구축과 루트 클래스¹⁾의 생성.

```
<owl:Class rdf:ID="Winery"/>
<owl:Class rdf:ID="Region"/>
<owl:Class rdf:ID="ConsumableThing"/>
```

2) rdf:subClassOf

하위클래스(rdf:subClassOf)는 어떤 클래스의 하위 개념을 나타낼 때 사용된다. 가령 'Person'은 'Mammal'의 하위클래스가 된다. 이 클래스 표현에 따르면 사람이 포유동물임을 유추할 수 있다.

```
예) 사람과 포유동물간의 하위생성클래스 예
<owl:Class rdf:ID="Person">
<rdfs:subClassOf rdf:resource="#Mammal"/>
...
</owl:Class>
```

3) rdf:Property

속성(rdf:Property)은 개체 사이의 관계나 개체의 데이터 값을 부여할 때 사용된다. 예를 들면, 'hasChild'를 비롯하여 'hasRelative', 'hasSibling', 'hasAge'라는 속성이 있을때 'hasChild', 'hasRelative', 'hasSibling'은 'Person' 클래스의 인스턴스간 연관성을 나타내는데 사용된다. 개체 간의 관계를 나타내는 속성은 'ObjectProperty'로 표기된다. 'hasAge'는 'Person' 클래스의 인스턴스에 정수 값을 부여하는데 사용된다. 개체의 값을 나타내는 속성은 'DatatypeProperty'로 표기된다. 'ObjectProperty'와 'DatatypeProperty'은

모두 'rdf:Property'의 하위클래스이다.

예) 'Sister'와 'Woman' 클래스의 인스턴스들이 'hasSibling'이라는 속성으로 연관성을 갖는 예.

```
<owl:ObjectProperty rdf:ID="hasSibling">
<rdfs:domain rdf:resource="#Sister"/>
<rdfs:range rdf:resource="#Woman"/>
</owl:ObjectProperty>
```

4) rdfs:subPropertyOf

하위속성(rdfs:subPropertyOf)은 어떤 속성의 하위 속성을 나타낼 때 사용된다. 예를 들면, 'hasSibling'은 'hasRelative'의 하위속성이다. 이를 통해 'hasSibling' 속성을 갖는 개체가 'hasRelative' 속성에도 관련됨을 유추할 수 있다. 속성도 클래스처럼 계층구조를 갖는다.

예1) 'Sibling'을 'Relative'의 하위클래스로 표현한 예.

```
<owl:Class rdf:ID="Relative"/>
<owl:Class rdf:ID="Sibling">
<rdfs:subClassOf rdf:resource="#Relative"/>
...
</owl:Class>
```

예2) 'Person' 클래스와 'Relative' 클래스 개체에서 'hasRelative'라는 속성을 갖는 예.

```
<owl:ObjectProperty rdf:ID="hasRelative">
<rdfs:domain rdf:resource="#Person"/>
<rdfs:range rdf:resource="#Relative"/>
</owl:ObjectProperty>
```

1) 'rdf:ID='구문으로 루트 클래스를 지정한다.

예3) 'hasSibling'은 'hasRelative'의 하위 속성이다. 'Person' 클래스가 'hasSibling' 속성을 갖는 예

```
<owl:ObjectProperty rdf:ID="hasSibling">
<rdfs:suPropertyOf rdf:resource="#hasRelative"/>
<rdfs:range rdf:resource="#Sibling"/>
</owl:ObjectProperty>
```

5) rdfs:domain

영역(rdfs:domain)은 속성이 적용될 개체를 제한한다. 속성이 개체와 관련된다면 그 영역으로 하나의 클래스를 갖으며, 개체들은 반드시 그 클래스에 속한다. 가령 'hasChild' 속성은 '포유동물' 영역에서 언급된다. A개체가 B개체라는 자식을 갖게 되면, A개체는 분명한 포유동물이다.

예) 포유동물이 'hasChild'라는 속성을 가지면, 해당 영역에 포함된 개체는 포유동물임을 표현한 예.

```
<owl:ObjectProperty rdf:ID="hasChild">
<rdfs:domain rdf:resource="#Mammal"/>
.....
</owl:ObjectProperty>
```

6) rdfs:range

범위(rdfs:range)는 속성을 가치 값으로 가질 수 있는 개체를 제한한다. 즉, 개체들이 어떤 속성으로 연관되면 속성의 범위로 클래스를 갖는다. 그러면 개체들은 반드시 범위 클래스에 속하게 된다. 가령 'hasChild' 속성은 'Person'의 범위를 갖는다. A개체와 B개체가 'hasChild'라는 속성으로 연관될 경우에(B개체가 A개체의

자식이면), B개체는 포유동물이다.

예) 'hasChild'라는 속성을 갖고 동일한 범주를 갖는 예

```
<owl:ObjectProperty rdf:ID="hasChild">
<rdfs:domain rdf:resource="#Mammal"/>
<rdfs:range rdf:resource="#Person"/>
</owl:ObjectProperty>
```

7) individual

개체는 클래스의 인스턴스이다. 속성이란 한 개체와 다른 개체를 연관짓기 위해 사용된다. 가령 A라는 이름의 개체는 'Person' 클래스의 인스턴스가 될 수 있다. 'hasEmployer'라는 속성은 A개체와 B개체를 연관짓는데 사용될 수 있다.

3. 2. 2 (비)동격 속성

1) equivalentClass

동격클래스는 두 클래스가 동등함을 나타낸다. 동격클래스들은 반드시 똑같은 인스턴스를 갖는다. 동격은 유사한 클래스를 생성하는데 사용된다. 가령 'Car'와 'Automobile'은 동격클래스가 될 수 있다. 그러면 'Car'의 어떤 인스턴스가 'Automobile'의 인스턴스가 될 수 있다는 것을 유추할 수 있다. 그 반대의 경우도 성립한다.

예) Car 온톨로지에서 Automobile이 동격임을 나타낸 예.

```
<owl:Class rdf:ID="Car">
<owl:equivalentClass rdf:resource="#Automobile;Car"/>
</owl:Class>
```

2) equivalentProperty

동격속성은 두 속성이 동등함을 나타낸다. 동격은 유사한 속성을 생성하는데 사용된다. 가령 'hasLeader'는 'hasHead'와 동격속성이 될 수 있다. 어떤 두 개체가 'hasLeader' 속성을 갖는다면 'hasHead'의 속성도 갖는다. 그 반대의 경우도 성립한다. 또한 'hasLeader'가 'hadHead'의 하위속성이 될 수 있고 그 반대의 경우도 성립한다.

```
<owl:ObjectProperty rdf:ID="hasLeader">
  <owl:equivalentProperty>
    <owl:ObjectProperty rdf:ID="has
      Head"/>
  </owl:equivalentProperty>
</owl:ObjectProperty>
```

3) sameAS

sameAS는 두 개체(individual)가 같음을 나타낸다. 가령 두 개체 William Jefferson Clinton과 Bill Clinton은 같다고 할 수 있다.

```
<rdf:Description rdf:about="#William__
  Jefferson__Clinton">
  <owl:sameAs rdf:resource="#BillClinton"/>
</rdf:Description>
```

4) differentFrom

differentFrom은 한 개체가 다른 개체와 다름을 나타낸다. 가령 개체 Frank는 다른 개체 Deborah나 Jim과 다르다고 할 수 있다. 개체들이 다르다는 점을 명시하는 것은 개체들이 하나의 이름을 갖는다는 것을 가정하지 않은 OWL과 같은 언어를 사용할 때 중요하다. 추가적인 설명이 없으면 Frank와 Deborah가

다른 개체라는 것을 유추하지 못한다.

예1) 개체 'Deborah'는 개체 'Frank'와 다름을 표현한 예

```
<PersonName rdf:ID="Frank" />
<PersonName rdf:ID="Deborah">
  <owl:differentFrom rdf:resource="#Frank"/>
</PersonName>
```

예2) 개체 'Jim'은 개체 'Frank', 'Deborah'와 다름을 표현한 예

```
<PersonName rdf:ID="Jim">
  <owl:differentFrom rdf:resource="#Frank"/>
  <owl:differentFrom rdf:resource="#Deborah"/>
</PersonName>
```

5) AllDifferent

AllDifferent는 사용하여 모든 개체가 서로 다름을 나타낼 수 있다. 가령 Frank, Deborah, Jim은 AllDifferent를 사용하여 서로 다르다고 언급할 수 있다. AllDifferent는 distinctMembers와 함께 사용된다.

예) PersonName 개체인 'Frank', 'Deborah', 'Jim'은 서로 다름을 표현한 예

```
<owl:AllDifferent>
  <owl:distinctMembers rdf:parseType="
    Collection">
    <PersonName rdf:about="#Frank"/>
    <PersonName rdf:about="#Deborah"/>
    <PersonName rdf:about="#Jim"/>
  </owl:distinctMembers>
</owl:AllDifferent>
```

4. 시소러스와 온톨로지의 의미 병합

OWL Lite는 앞서 조사한 바와 같이 시소러스 구조와 용어개념 구조화와 같은 용어간 개념 구조 사전의 변형의 이론적 배경을 살펴 보면 상호간 호환성이 매우 높은 것으로 나타나고 있다. 이런 유사성을 시소러스의 기본 구조로 비교하여 상호간 언어적 호환의 가능성을 조사한다. 본 연구에서 참고한 관계의 기준은 국내외 구축 지침 가운데 최신성이 높은 자료인 ANSI(2003)에 근거하여 분석한다.

4. 1 계층관계

현재 ANSI에서 제안하는 계층관계로 구성할 수 있는 관계는 크게 다섯 가지로 구분한다. 이를 구체적으로 살펴보면 다음과 같다. 참고로 이에 해당하는 온톨로지 어휘는 subClassOf와 subPropertyOf 등으로 대별된다.

4. 1. 1 속관계(Generic Relationship)

이 관계는 하나의 클래스에 속하는 구성요소나 혹은 종(species)과의 관계를 나타낸 것으로 하나의 개념과 총칭(대표어)로 표현되는 관계이다. 기본적인 것은 영어의 문형으로 표현할 경우에 술부 IS를 기준으로 주어와 술어의 관계로 표현할 수 있다. 이때 계층은 <협의어 IS '광의어'>로 표현된다. 또한 이론적으로 전부와 일부(all-and-some)의 관계도 이에 속한다. 예를 들면, 선인장은 다즙 식물 가운데 일부이다. 한편 이 관계에 속하기 위해서는 광의어와 협의어간에 완전한 내삽이 필요하다. 부분적인 내삽이 이루어지는 경우에는 해당

용어들은 반드시 다른 계층에 배정되어야 한다. 예를 들면, 선인장과 사막식물은 완전하게 포함되지 않는다. 왜냐하면 선인장은 사막식물의 일부일 수도 있으며, 다른 유에 속하는 식물일 수 있기 때문이다. 다음은 이 관계를 OWL Lite로 구조화한 예이다.

```
<owl:Class rdf:ID="선인장"/>
  <rdfs:subClassOf rdf:resource="#다즙식물"/>
<owl:Class/>
```

4. 1. 2 전체-부분 관계(Whole-Part Relationship)

이 관계는 하나의 개념이 다른 용어로의 상속적 정보가 앞의 속관계보다 약한 경우이다. 부분이 모여 상위 개념으로써 전체를 구성하는 것이다. 예를 들면, 자동차의 엔진과 타이어 등과 같은 하위 구성요소가 열거되어 상위어으로써 자동차를 결정하는 관계이다. 이 관계로 표현될 수 있는 것은 다음 4가지이다.

- 시스템 및 신체기관: 예 중추신경계와 두뇌, 脊髓(척수)
- 지리요소: 캐나다와 온타리오
- 학문분야 혹은 주제의 원리: 과학과 생물학
- 일반적 계층조직: 군대와 사단, 연대

다음은 이들 관계중 일부 관계를 OWL Lite로 구조화한 예이다.

```
<owl:Class rdf:ID="온타리오"/>
<owl:Class rdf:ID="알버타"/>
<owl:Class rdf:ID="캐나다"/>
<owl:Class rdf:ID="온타리오">
  <rdfs:subClassOf rdf:resource="#캐나다"/>
```

...

</owl:Class>

4. 1. 3 사례 관계(Instance Relationship)

이 관계는 하나의 일반 범주를 일반명사로 표현될 수 있는 개념어와 이에 배정된 고유명사와의 관계이다. 예를 들면, '산악지역'이라는 일반 명사로 표현되는 범주에 알프스와 히말라야와 같은 고유명사는 계층관계로 표현될 수 있다. 이 관계는 subClassOf의 관례로 표시할 수 있다.

```
<owl:Class rdf:ID="알프스">
<owl:Class rdf:ID="히말라야">
<rdfs:subClassOf rdf:resource="#산악지역"/>
...
</owl:Class>
```

4. 1. 4 다계층 관계(Polyhierarchical Relationship)

일정한 개념은 태생적으로 복수의 상위어로 구조화될 수 있다. 이러한 관계를 다계층 관계라 한다. 예를 들면, 피아노는 내부적으로는 현악기이며, 건반을 이용하는 타악기, 좀 더 세부적으로 건반악기 하에 하위개념어가 될 수 있다. 즉, 피아노의 관점에서는 세 개의 상위어를 갖게 된다. 역으로 하위 다계층 관계도 가능하다. 예를 들면, 생화학은 생물학과 화학이라는 하위어를 가질 수 있기 때문에 하위 다계층 관계라 할 수 있다.

```
<owl:Class rdf:ID="피아노">
<rdfs:subClassOf rdf:resource="#현악기"/>
<rdfs:subClassOf rdf:resource="#타악기"/>
<rdfs:subClassOf rdf:resource="#건반악기"/>
...
```

</owl:Class>

```
<owl:Class rdf:ID="생물학">
<owl:Class rdf:ID="화학">
<rdfs:subClassOf rdf:resource="#생화학"/>
...
</owl:Class>
```

4. 1. 5 계층내 노드 레이블(Instance Relationship)

이 관계는 기존의 지침서에서 표현되지 않는 관계이다. 다만 블리스 분류표(Bliss Classification)에서 시작된 패킷의 개념을 보다 구체적으로 표현한 관계이다. 예를 들면, 다음과 같다.

```
cars
  by motive power
    diesel cars
    electric cars
  by purpose
    racing cars
    sports cars
```

자동차의 하위어로 디젤자동차나 경주용 자동차는 논리적으로 정상적인 배정방식이다. 단 디젤자동차와 경주용 자동차는 연관성이 상대적으로 떨어진다. 이 경우에 이를 동력(motive power)과 자동차 사용목적(purpose)에 따라 구분하여 배열할 수 있다. 즉, 패킷 개념을 도입하여 단순배열을 의미적으로 개선할 수 있는 관계이다. 다음은 이들 패킷을 OWL Lite로 구조화한 예이다.

```
<owl:ObjectProperty rdf:ID="hasPurpose">
<rdfs:subPropertyOf rdf:resource="#has
```

```
CarsDescriptor"/>
<rdfs:range rdf:resource="#CarsPurpose"/>
...
</owl:ObjectProperty>
```

4. 2 연관관계

연관관계는 디스크립터의 관계를 계층이나 대등관계로 처리할 수는 없으나 서로 연관성이 높은 개념을 구조화한 관계이다. 다만 이 두개의 관계는 의미적이나 개념적으로 그 연관성이 계층에 비해 느슨한 경우이다. 따라서 용어간 대칭적이며 독립적인 용어간 구조화할 수 있는 관계이다. 이 관계로 지정되는 경우는 크게 동일한 계층에 지정된 용어를 구조화하는 것과 상이한 계층에 지정된 용어를 구조화한 것으로 구분할 수 있다. 참고로 이에 해당하는 온톨로지 어휘는 ObjectProperty를 비롯하여 Data-typeProperty, inverseOf 등으로 대별된다.

4. 2. 1 동일범주내 중복 자매어(Overlapping sibling terms)

이 관계에 속하는 것은 선박과 보트와 같이 의미도 중복되며 동일한 상위어(운송수단)를 갖는 자매어이다. 각 용어가 갖는 개념의 중요도나 혹은 범위에 차이가 있다. 간혹 이 관계는 해당 시소러스의 특성에 따라 서로간의 대체도 가능한 수준으로써 하나의 용어가 다른 용어를 연상할 수 있는 수준의 용어간 관계를 의미한다.

```
<owl:Class rdf:ID="선박">
<owl:equivalentClass rdf:resource="#보트"/>
</owl:Class>
```

4. 2. 2 동일범주내 상호 독립적 자매어

(Mutually Exclusive sibling terms)

이 관계는 자매어간에 대치가 이루어질 수 없는 관계를 말한다. 예를 들면, 장미와 수선화는 꽃이라는 상위어를 갖지만 이 두개의 자매어는 앞의 경우와 달리 대치가 될 수 없는 매우 독립적 상호 관계이다.

```
<꽃 rdf:ID="장미"/>
<꽃 rdf:ID="수선화">
<owl:differentFrom rdf:resource="#장미"/>
</꽃>
```

4. 2. 3 동일범주내 파생 관계(Derivational sibling terms)

서로 친숙하거나 파생적 관계로 연결된 개념어들은 연관관계로 처리된다. 예를 들면, 말료에 하위어로서 나귀와 노새, 말 등이 있다. 이때 노새는 나귀와 말의 잡종으로써 '나귀와 노새'와 '노새와 말'은 연관관계로 구조화될 수 있다. 이 때 말과 나귀는 서로 파생관계가 없기 때문에 연관관계로 구조화하지 않는다. 특히 말과 나귀가 다르다는 것을 선언하기 위해서는 세개의 단계로 선언을 한다. 우선 말과 말은 말이 포함되어 있다는 것을 선언하고, 말과 노새는 sibling관계라는 것을 선언한다. 최종적으로는 말과 나귀는 서로 다른 속성을 갖는다고 선언한다. 다음은 초기 선언과정이다.

```
<owl:ObjectProperty rdf:ID="has말과">
<rdfs:domain rdf:resource="#말"/>
<rdfs:range rdf:resource="#말과">
</owl:ObjectProperty>
```

4. 2. 4 다른 범주의 용어간 관계

서로 다른 범주에 속하면서 연관관계로 연결될 수 있다. 이 관계는 하나의 용어가 다른 용어를 보완하는 경우에 주로 해당한다. 동일한 파생원을 갖고 있지 않으며, 동일한 범주에 속하지도 않는 경우에 속하나, 의미적으로 보완되는 관계이다. 예를 들면, 음악과 음악가는 서로 다른 범주에 속하는 용어이지만 이 두개의 개념어는 연결되어 있어 서로 보완하는 역할을 수행할 수 있다. 따라서 이 두개의 용어가 같이 보완적 용어들은 연관관계를 설정할 수 있다. 이 때 포함하는 사례는 다음과 같다.

- 학문영역이나 주제영역과 그 분석대상이 되는 주제간의 관계: 식물과 식물학
- 특정 행위와 이를 측정할 수 있는 도구: 온도계와 온도조절
- 작용제와 그 대상물: 풀과 제초제
- 행위와 행위 결과: 재봉질과 의복
- 행위와 목적: 수확과 곡식
- 특정물과 그의 특성: 액체와 표면장력
- 일반적으로 서로 연결된 개념: 죽음과 매장
- 개념과 이를 측정할 수 있는 장치: 전압과 전류
- 실제 상위어가 아닌 명사구. 예를 들면, 고무오리와 오리간의 관계로 설명된다.

```
<owl:ObjectProperty rdf:ID="온도를측정한다">
<rdfs:domain rdf:resource="#온도계"/>
<rdfs:range rdf:resource="#온도조절"/>
</owl:ObjectProperty>
```

‘온도계’ 클래스와 ‘온도조절’ 클래스의 개체들은 ‘온도를측정한다’라는 속성을 갖는다.

4. 2. 5 관련어용 노드 레이블

연관관계를 설정할 때 해당 관계어의 주제에 따라서 주관적 관점에 따라 의미적으로 구조화한 관계이다. 예를 들면, 다음과 같다.

BOOK

```
<operation>
binding
printing
<topic>
fiction
biography
```

책(book)과 관련된 하위어를 패킷개념으로 구분하고 해당 패킷에 배정된 용어는 서로 연관관계를 설정할 수 있다. 즉, 패킷 개념을 도입하여 단순배열을 의미적으로 개선할 수 있는 관계이다. 다음은 이들 패킷을 OWL Lite로 구조화한 예이다.

```
<owl:ObjectProperty rdf:ID="hasTopic">
<rdfs:DatatypePropertyOf rdf:resource=
"#hasBookDescriptor"/>
<rdfs:range rdf:resource="#BookPurpose"/>
...
</owl:ObjectProperty>
```

4. 3 대등관계

대등관계는 하나의 개념이 두개 이상의 복수 용어로 사용될 때 용어간의 관계를 맺어주기 위한 것이다. 이 관계는 크게 일반적 대등관계와 동의어관계, 변형어휘관계, 유사동의어관계로 구분된다. 참고로 이에 해당하는 온톨로지 어휘는 `equivalentClass`를 비롯하여 `equi-`

valentProperty, sameAS 등으로 대별된다.

4. 3. 1 일반적 대등관계

하나의 개념을 두개 내지 세 개 이상의 언어로 표현되는 관계로서 우선어와 비우선어로 표현된다. 이 관계는 복수개의 용어가 갖고 있는 개념 및 범주의 크기가 일치하는 경우가 이에 해당한다. 또한 동의어와 유사동의어, 변형어휘 관계 등을 제외한 모든 대등관계의 용어가 이에 해당한다.

다음은 이 관계를 OWL Lite의 가장 일반적인 예로 구조화한 예이다.

```
<rdf:Description rdf:about="#William_Jefferson_Clinton">
  <owl:sameAs rdf:resource="#BillClinton"/>
</rdf:Description>
```

4. 3. 2 동의어 관계

복수개의 용어가 갖고 있는 개념 및 범주의 크기가 거의 일치하는 관계이다. 실제적으로 완전한 동의어는 존재하지 않기 때문에 이 관계는 주제영역이나 상황에 따라 조금의 차이가 발생한다. 즉, 공식적인 용어, 전문적인 용어 대 일반 용어의 관계를 포함하여, 경멸어(pejorative) 대 평어 대 찬사어(complimentary) 관계 등이 동의어 관계라 할 수 있다. 이 관계에 속하는 사례는 다음과 같다.

- 다른 어원을 갖는 용어: 소금과 나트륨
- 일반 용어와 과학 용어: 아스피린과 아세티살릭산
- 속명과 상품명: 복사와 제록스
- 새로운 개념에 대한 변형 명사: 하바크래프트와 수직이착륙기

- 유행어와 비유행어: 후진국과 개발도상국
- 방언: 엘리베이터와 리프트

다음은 이 관계를 OWL Lite로 구조화한 예이다.

```
<rdf:Description rdf:about="#후진국">
  <owl:sameAs rdf:resource="#개발도상국"/>
</rdf:Description>
```

4. 3. 3 변형어휘 관계

변형 어휘는 동의어관계의 일종이며, 이 관계는 철자법이나 문법적 변형이 이에 속한다.

- 어휘 변형: 직접화법과 간접화법이 경우를 비롯하여 용어간 철자변형 등의 관계가 이에 해당한다: 그리이스와 그리스
- 그러나 한글의 경우는 이러한 관계가 외래어나 외국어에 대한 이형 표기에 의해 달라지는 용어간 관계가 이에 해당한다.

- 완전명과 약어의 관계: 폴리비닐클로라이드와 피브씨

다음은 이 관계를 OWL Lite로 구조화한 예이다.

```
<rdf:Description rdf:about="#Poly_vinyl_Chloride">
  <owl:sameAs rdf:resource="#PVC"/>
</rdf:Description>
```

4. 3. 4 유사동의어 관계

유사동의어의 관계를 결정짓는 것은 용어보다 개념에 의존하여야 한다. 또한 다른 관계에 비해 사례로 표시하기 어려우며, 또한 관점에 따라 다른 관계로 표현될 수 있는 관계도 검색 환경과 같은 외부적 요인에 따라 유사동의어로 표현될 수 있다(남영준 2004). 예를 들면, 임

산관련 시소러스에는 나무의 하위어로서 소나무나 향나무, 밤나무 등이 포함될 수 있으나, 항공분야에서는 이를 크게 구분할 필요없이 나무라는 우선어와 나머지 용어로 이루어진 비우선어로 처리될 수 있는 것이다. 따라서 유사동의어는 주제 영역의 환경에 따라 조정하는 것이 일반적이다.

4. 3. 5 복합명사에서 상호참조 관계

이 관계는 특정 개념을 복합명사로 구성화할 경우, 검색환경에 따라 각각의 단일명사가 상호보완적인 영향을 미칠 경우에 상호참조표시를 하는 관계이다. 예를 들면, 석탄 채굴이라는 행위 개념이 석탄과 채굴이라는 용어의 비우선어로 사용됨으로써 석탄과 채굴이라는 각각의 개념이 서로 관계가 있음을 알 수 있도록 하는 관계이다.

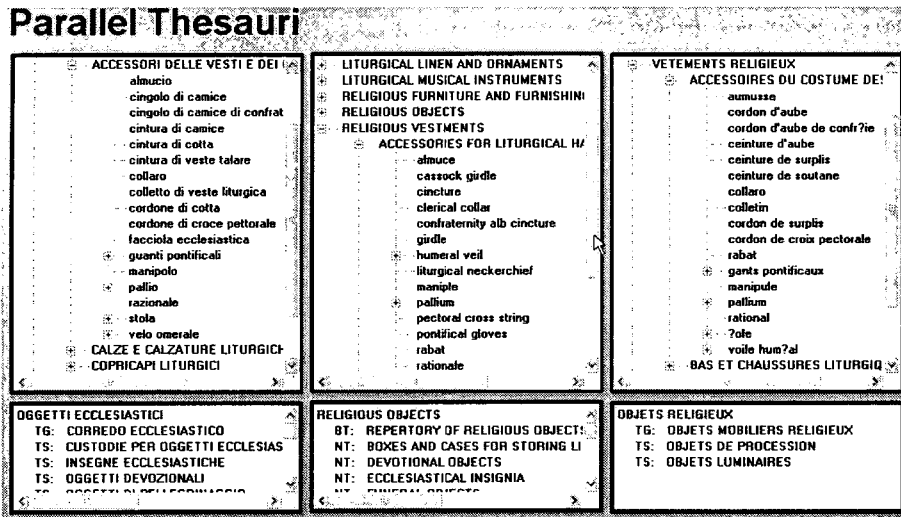
```
<owl:ObjectProperty rdf:ID="석탄을채다">
  <owl:equivalentProperty>
    <owl:ObjectProperty rdf:ID="채굴한다">
      </owl:equivalentProperty>
    </owl:ObjectProperty>
```

4. 4 다국어 수용의 온톨로지 구성

다국어 전자자원을 동시에 검색할 수 있는

시소러스는 이론적으로 우선어로서 한글 디스크립터와 이에 대등되는 외국어 디스크립터를 동시에 시스템에 전송하는 방법이 있다. 앞의 유사동의어 관계에서 지적인 바와 같이 완전한 유사동의어는 존재하지 않기 때문에 다국어간의 대등관계는 오히려 개념의 해석보다 단순히 대상물에 대한 각국의 사용언어를 병렬관계로 나열하는 것이 효과적일 수 있다.

<그림 2>는 미국의 게티연구소의 AAT부서와 캐나다 문화정보네트워크(CHIN), 프랑스의 문화커뮤니케이션부, 이탈리아의 목록도큐멘테이션 센터(ICCD)와 공동으로 개발한 다국어 시소러스이다(ICCD 1999). <그림 2>와 같이 영미계의 종교관련물(Religious Objects)의 개념에 대응되는 용어(개념)로는 'Oggetti Ecclesiastici(대응 이탈리아어)'과 'Objects Religieux(대응 프랑스어)'의 수준은 병렬관계로 표시가 가능하다. 그러나 이의 하위어에 대해서는 이탈리아의 경우는 상위어 1개와 하위어 15개를 갖고 있으며, 영어의 경우도 동일하다. 그러나 프랑스어의 경우는 이에 해당하는 하위어는 2개로 구성되어 큰 차이를 보이고 있다. 이는 각국의 문화적 차이에 따라 발생하는 것으로 단순한 병렬규칙만으로 시소러스를 구성할 수 없음을 알 수 있다.



〈그림 2〉 ICCD의 다국어 시소러스의 일부

4. 4. 1 온톨로지 적용 다국어 시소러스 구성

온톨로지 적용 다국어 시소러스의 구축은 두가지 관점에서 실험적으로 실시한다. 하나는 한글 디스크립터를 영미계 개념구조에 적용하는 것과 영어 디스크립터를 한글 개념구조에 적용하는 것이다.

- 한글 디스크립터의 영미계 개념구조

다음과 같은 시소러스 구조를 영미계 개념구조로 적용하는 것은 현재의 시소러스 구축 지침범위 밖에 있어 번역에 어려움이 있다. 이와 같은 어려움은 도깨비에 해당하는 대응 영미계 시소러스가 존재하지 않는 것이다. 문화재 시소러스의 경우 처용의 얼굴이 들어있는 도깨비문양의 막새기와는 외부로부터의 액을 막아주는 부적의 개념이기 때문에 좋은 의미를 갖는 개념이다. 이에 비해 영미계의 유사용어는 괴물이나 귀신과 같은 불길한 의미를 갖기 때문에 이에 대한 정확한 대치가 어렵다.

이와 같은 문제점을 극복하기 위해서는 패싯의 개념이 포함된 온톨로지 문법을 이용하여

국내 개념구조	영미계 개념구조
건축재료	building materials
.기와	.tile
..막새기와	..finishing tile
...도깨비문양기와	...??? pattern tile

시소러스를 구축한다. 세계의 과정으로 이루어진다. 도깨비가 속해있는 계층구조를 선언하고, 해당 용어에 대한 속성적 특징을 배정한 후에, 해당용어가 포함된 도깨비문양기와를 계층구조에 배열하는 것이다. 다음은 온톨로지 어휘를 사용하여 재구조화한 예이다.

```

<owl:Class rdf:ID="도깨비">
  <rdfs:subClassOf rdf:resource="# 鬼類"/>
</owl:Class>
...
<owl:ObjectProperty rdf:ID="hasCharacter">
  <rdfs:subPropertyOf rdf:resource="# has도깨비Descriptor"/>
  <rdfs:range rdf:resource="# 도깨비Character"/>
  ...
  
```

</owl:ObjectProperty>

즉, 복합명사 레이블의 특성을 받아들여 도깨비가 갖고 있는 특성을 인지하고, dokebi pattern tile이라는 대응 언어로도 한글을 이해하지 못하는 외국 이용자들이 이 문양의 기와가 갖는 의미를 인지할 수 있다.

- 영미계 디스크립터의 한글 개념구조

다음과 같은 영미계 시소러스 구조를 한글 개념구조로 적용하는 것은 현재의 시소러스 구축 지침범위 밖에 있어 번역에 어려움이 있다. 이와 같은 어려움은 영미계 문화와 언어적 특성에 대한 이해 부족과 언어의 의미적 변형이 빈번하게 발생함으로 인한 것이다.

국내 개념구조	영미계 개념구조
도자기	china
①	porcelain
②	ceramics

위의 표에서 도자기라는 우선어와 외국어 비우선어로서 china, porcelain, ceramics를 링크할 수 있으나, AAT(Art and Architecture Thesaurus)에서는 별도의 디스크립터로 배정하고 있다. 즉, 한글에서는 하나의 개념어로 표현되는 디스크립터를 사용하고 있으나, 영어에서는 복수의 개념어로 세분되어 전개하고 있다.

이와 같은 문제점을 극복하기 위해서는 패시 개념이 포함된 온톨로지 문법을 이용하여 시소러스를 구축하는 것이 보다 효율적이다. 구축의 기준은 영미계 용어에 포함된 공통 속성과 서로를 구분할 수 있는 속성을 이용한다. 이 때

구조화 과정은 두개 단계로 구분할 수 있다. 하나는 각 용어가 갖고 있는 속성을 선언하는 것이고, 다른 하나는 각 용어들이 서로 다르다는 것을 선언하는 것이다. 다음은 온톨로지 어휘를 사용하여 China클래스와 PeriodJin 클래스의 개체들이 'beMadeIn'이라는 속성을 갖고 있으며, china와 porcelain, ceramics의 개념은 서로 다르다고 선언한 예이다.

```

<owl:ObjectProperty rdf:ID="beMadeIn">
<rdfs:domain rdf:resource="#China"/>
<rdfs:range rdf:resource="#PeriodJin"/>
</owl:ObjectProperty>
....
<owl:AllDifferent>
<owl:distinctMembers rdf:parseType="공
예품">
<vin:도예류 rdf:about="#china"/>
<vin:도예류 rdf:about="#porcelain"/>
<vin:도예류 rdf:about="#ceramics"/>
</owl:distinctMembers>
</owl:AllDifferent>
    
```

5. 결론

본 연구에서는 다국어 전자정보원을 검색하기 위한 검색도구로서 시소러스 기반의 온톨로지 개념을 이용하는 이론적 방안을 연구하였다. 기존의 시소러스는 외국의 전자자원을 검색하기 위해서는 해당 디스크립터에 병렬로 배정된 영어 대응 디스크립터를 사용하였다. 한편 국내 한글 시소러스는 문화와 언어

적 차이 때문에 모든 디스크립터에 대한 대응 외국어 디스크립터가 없었으며, 이는 외국의 다국어 시소러스에서도 동일한 현상을 확인할 수 있었다.

이러한 시소러스의 한계를 극복하기 위해 국내의 시소러스 구축의 주요 지침서에 있는 관계들을 조사하고, 가장 공통적인 형식은 용어간의 대등관계를 비롯하여 연관관계, 계층관계로 용어간 관계설정방법이 고정화되어 있음을 확인하였다. 각각의 관계는 보다 정교한 검색이 가능하도록 사례중심으로 세분화되어 있었으며, ANSI의 지침은 기존의 모든 관계설정이론을 수용하고 전자자원용 검색도구에 가장 적합한 사례를 갖고 있었다. 이렇게 조사된 사례를 온톨로지 언어 가운데 W3.org에서 표준화한 OWL중 OWL Lite에 나와 있는 언어를 분석하여 시소러스의 구조를 수용할 수 있는 지 여부를 조사하였다.

조사결과를 요약하면 다음과 같다.

OWL Lite 언어를 시소러스의 기본관계에 대입해 본 결과, 계층관계의 경우는 subClass

Of와 subPropertyOf 등이, 대등관계는 equivalentClass를 비롯하여 equivalentProperty, sameAS 등이, 그리고 연관관계는 ObjectProperty을 비롯하여 DatatypeProperty, inverseOf 등이 적합한 것으로 나타났다. 이러한 언어를 기반으로 실제 AAT에 배정된 디스크립터와 ICCD에서 개발한 다국어 시소러스의 디스크립터의 한글과 영미계 디스크립터의 번역화 작업은 2개 이상의 개념 선언화 과정이 필요하였다. 또한 한글시소러스를 온톨로지 로 변환에는 단순히 대등관계의 온톨로지 언어보다 실제로 속성관련 온톨로지 언어로 해당 용어에 대한 선언과정과 구조화 과정이 주로 사용되었다. 이 연구는 최소한의 데이터로써 온톨로지로의 변환을 시행한 것이기 때문에 본 연구의 결과가 모든 한글시소러스에 동일하게 적용될 수는 없을 것이다. 따라서 다국어 시소러스의 온톨로지로의 변환에 대한 실험데이터를 극대화한 후속 연구와 Protege수준의 다국어 시소러스 개발용 프로그램의 개발도 필요하다.

참 고 문 헌

- 김은경, 남영준. 2004. 시맨틱 웹을 위한 온톨로지 구축방법에 관한 비교연구. 『정보관리연구』, 35(2): 57-85.
- 김은경. 2004. 『시맨틱 웹을 위한 온톨로지 구축방법에 관한 비교연구』. 석사학위논문, 중앙대학교 대학원, 문헌정보학과.
- 남영준. 2004. 대등관계 설정의 확장 지침에 관한 연구. 『정보관리학회지』, 21(2): 1-21.
- 남영준. 2003. 시소러스의 대등관계에 관한 연구. 『문헌정보학보』, 중앙대학교 문헌정보학회, 6: 143-175.
- 오삼균. 2002. 시맨틱웹 기술과 활용방안. 『정보관리학회지』, 19(4): 297-320.

- 정도현, 김태수. 2003. 시소러스를 기반으로 한 온톨로지 시스템 구현에 관한 연구. 『정보관리학회지』, 20(3): 155-176.
- Aitchison, Jean, and Alan Gilchrist, David Bawden. 2000. *Thesaurus Construction and Use: a Practical Manual*. 4th. ed. Aslib.
- Galinski, Christian, and Goebel Jurgen W. 1996. *Guide to Terminology Agreements*. Inforterm, Termnet.
- ICCD. 1999. *Thesaurus Multilingue del Corredo Ecclesiastico*. ICCD.
- ISO. 1985. *Documentation - Guidelines for the Establishment and Development of multilingual Thesauri*. 2nd edition (ISO 5964-1985(E))
- ISO. 1986. *Documentation - Guidelines for the Establishment and Development of Monolingual Thesauri*. 2nd edition (ISO 2788-1986(E)) 정동열 역. 단일언어 시소러스 제정 및 개발 지침. 시소러스 개발지침. 문헌정보처리연구회. 1994.
- Lassi, Monica. 2002. *Automatic thesaurus construction, the GSLT course Linguistic Resources*. Graduate School of Language Technology Swedish School of Library and Information Science, University College of Borås. [online]. [cited 2004.8.13] <<http://www.adm.hb.se/personal/mol/gslt/thesauri.pdf>>
- NISO. 2003. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri: ANSI/NISO Z39.19 - 2003*, the American National Standards Institute.
- NISO/ASI/ALCTS. 1999. *Workshop on Electronic Thesauri: Planning for a Standard*, Washington, DC. [online]. [cited 2004.9.10] <http://www.niso.org/news/events_workshops/thesau99.html#issues>
- Seetharama, S. 1975. *Term-Concept Relationship in an Information Retrieval Thesaurus*. Seminar on thesaurus in Information Systems. DRTC and INSDOC. Bangalore. : A80-A87.
- Smith, Michael K., Chris Welty, and Deborah L. McGuinness. *OWL Web Ontology Language Guide, W3C Recommendation*. [online]. [cited 2004.9.13] <<http://www.w3.org/TR/owl-guide/>>