

## 온톨로지(Ontology)



### 심 경

정보학박사  
한국도서관협회 평생회원  
(주)아이리스넷 대표  
shim@irisnet.co.kr

수년 전 온톨로지라는 말을 웹과 관련하여 처음 들었을 때 제일 먼저 생각이 난 것은 듀이십진분류법이었다. 그 중에서도 '100 철학'의 하위주제인 '형이상학' 아래 위치한 분야인 'ontology' 즉, '존재론'이었다. 그러나 이 용어가 우리 분야에서 언급되면서 철학에서 말하는 존재론을 의미하지는 않을 것 같았고, "대체 뭘까?"하는 호기심도 생겼지만, 선뜻 자료를 찾아서 공부할 마음이 생기지는 않았다. 그 이유 중 하나는 온톨로지는 거의 항상 시맨틱 웹(Semantic Web), RDF(Resource Description Framework: 자원기술구조), OWL Web Ontology Language 등과 같이 이해하려는 대상보다 더 어려운 다른 용어들과 같이 나오는 것이다. 그 이외에도 해석은 되지만 의미가 와 닿지 않는 생소한 용어들이 설명의 대부분을 차지하니, 어디부터 시작해야 좋을지 엄두가 나지 않았던 것이다. 또 어쩌다 큰 마음을 먹고 몇몇 자료를 살펴본 결과는 "한글이 이렇게 어려울 수도 있구나" 하는 좌절감이 들거나, "왜 글을 쓰면서 읽는 사람이 이해하기 쉽게 쓰지 않는 것일까?" 하는 애꿎은 저자에 대한 원망과 동시에 괜한 열등감에 빠지는 것이었다.

사실 국내 학술지에 수록된 대다수 연구논문들은 온톨로지가 무엇인가를 다루기보다는 "온톨로지를 적용한 결과가 이렇다" 또는 "온톨로지를 이러한 분야에 저렇게 응용해 보았다"라는 내용을 주로 다루고 있다. 말하자면 독자들이 온톨로지가 무엇인지는 알 것이라는 전제로 연구의 결과를 보고하는 것이기 때문에 교과서와 같이 일일이 그 개념을 설명해 주지 않는다. 그러다보니 실무에 종사하는 일반 독자들은 온톨로지의 기본 개념이 무엇인지를 정확히 알기 어렵고, 그런 논문을 아무리 열심히 읽어 보아도 그 내용을 이해하기 어려울 수밖에 없다. 하지만 천리 길도 한 걸음부터라고, 일단 기본이라도 이

해해 보겠다는 마음으로 여기 저기에서 찾아보니 제법 온톨로지에 관한 얘깃거리가 생겨나는 것 같다.

그 결과로 이 글에서는 온톨로지의 기본 개념을 이해하는 데 도움이 되는 몇 가지 주제를 다루어 보기로 하였다. 먼저 온톨로지라는 용어가 사용되기 시작한 역사를 살펴보고, 자주 인용되는 온톨로지의 몇 가지 정의를 비교하며 생소한 용어를 하나씩 풀어서 설명하였다. 그럼에도 개념을 이해하기가 쉽지 않으므로 우리에게 비교적 친숙하고 종종 온톨로지의 한 종류라고 일컬어지기도 하는 시소러스와 온톨로지를 비교하여 개념에 좀 더 다가가도록 시도하였다. 마지막으로 온톨로지 종류에 대한 몇 가지 분류를 소개하여 개별 온톨로지 간의 관계를 정리해 보았다.

## 온톨로지의 역사

온톨로지라는 용어는 상당히 오랫동안 사용되어 왔다. 예를 들면, 웹스터 사전은 온톨로지를 대략 1721년에 출현한 용어로 기록하며: 1) 존재의 본질과 관계에 관심을 가지는 형이상학의 한 분야; 2) 존재의 본질 또는 존재의 종류에 관한 특정 이론이라는 두 가지 정의를 제공한다. 여기에 제시된 추상적인 철학 개념의 온톨로지는 점차 좀더 수학적이고 세밀한 영역(domain)으로 움직였으며, 형식 온톨로지(formal ontology)라는 개념은 최소한 1900년 이래로 존재해 왔다.

온톨로지는 이처럼 원래 철학의 한 분야였지만 보다 광범한 분야에서 수용되어 왔다. 우리가 이 글에서 이해하려고 하는 세상(world) 또는 그 일부분에 대한 지식 표현(knowledge representation)의 한 형태라는 새로운 개념의 온톨로지는 인공지능, 시맨틱 웹, 소프트웨어 공학, 생명의학정보학(biomedical informatics), 도서관학, 정보설계(information architecture) 분야 등에 적용되고 있다. 또한 과리노(Guarino, N.)는 지식공학(knowledge engineering), 지식표현, 정성적 모델링(qualitative modeling), 언어공학(language engineering), 데이터베이스 디자인, 정보검색과 추출, 지식 관리와 조직 등의 분야에 온톨로지가 수용되었다고 한다<sup>1)</sup>.

이렇게 온톨로지라는 용어가 철학 이외의 학문분야에서 본격적으로 사용된 것은 1970년 중반 이후 인공지능 분야의 연구자들이 대규모의 강력한 인공지능 시스템을 구축하기 위해서는 지식의 획득(capturing knowledge)이 필수적이라는 것을 깨닫기 시작하면서부터이다. 인공지능학자들은 일종의 자동추론을 가능하게 하는 전산모델(computational models)로서 새로운 온톨로지들을 생성할 수 있다고 주장하였다. 그리고 1980년대에 들어 인공지능 부문에서 “모델화된 세상에 대한 이론(theory of a modeled world)”과 “지식시스템의 한 구성요소(a component of knowledge systems)”를 지칭하기 위해 온톨로지라는 용어를 사용하기

1) Guarino, N. (1998). Formal ontology and information systems. In M. Guarino (Ed.), *Proceedings of FOIS '98*. Amsterdam: IOS Press. pp. 3-15.

시작하였다<sup>2)</sup>. 그러니 우리가 요즘 듣는 온톨로지의 개념은 인공지능 분야에서 시작된 것이라고 봐도 괜찮을 듯하다.

인공지능은 문헌정보학 분야에도 전혀 새로운 것은 아니다. 1970년대에 인공지능이 활기를 띠면서 우리 분야에서도 문헌색인을 위한 의미표현과 지능형정보검색시스템(intelligent information retrieval systems)이라는 주제와 더불어 인공지능 기법의 적용에 대한 연구가 유행하였으나, 1980년대 말경에 이르러 그 추세가 잠잠해졌다<sup>3)</sup>. 이처럼 우리 분야와 멀어졌다고 생각한 인공지능이 우리에게 새로운 개념인 온톨로지로서 문헌정보학 분야에 다시 부각되기 시작한 것은 근래 월드 와이드 웹과 관련하여 등장한 시맨틱 웹이라는 새로운 움직임과 관련이 있다고 할 수 있다. 이 시맨틱 웹이란 웹 상의 데이터나 콘텐츠를 응용 프로그램, 플랫폼 또는 영역에 관계없이 컴퓨터가 이해하고 처리하도록 하자는 것이다. 현재 웹에서 제공하는 콘텐츠는 검색엔진과 같은 컴퓨터가 검색해 오지만, 컴퓨터는 웹 상의 정보나 콘텐츠를 “이해”하지 못하기 때문에 그 검색결과가 부정확하거나 사람이 결과를 보고 적합성을 결정하고 연관자료를 찾는 등의 추가노력을 해야 한다. 따라서 과거 웹 상의 키워드 일치 방식을 의미 부여가 가능한 지식표현으로 컴퓨터가 웹 상의 데이터나 콘텐츠의 의미를 이해하도록 하여 웹 콘텐츠에 대한 이용자 요구를 만족시키는 움직임이다. 예를 들면, 뉴욕 여행에서 하나씩 사는 티셔츠에 인쇄된 “I ♥ New York”을 사람은 금방 이해하지만, 컴퓨터는 알 길이 없는데, 이러한 것을 시맨틱 웹으로 가능케 하자는 것이다. 좀더 우리 분야에 가까운 예를 들자면, 은행(銀行)을 검색한 이용자에게 은행(銀杏)에 관한 자료를 제공하지 않게 되는 것이다. 온톨로지는 이러한 시맨틱 웹에서 지식표현을 가능하게 해 주되, 인간보다는 기계가 이해하고 추론할 수 있도록 하는 요소기술 또는 핵심기술로서 근래 우리의 관심을 끌고 있는 것이다.

그럼 온톨로지가 어떻게 생겼길래 이런 일이 가능할까? 어쩌면 너무 과격한 단순화일 수도 있으나, 문헌정보학의 기법에 비유하자면, 온톨로지는 시소러스처럼 문자 그대로 특정 분야나 대상을 개념화해 놓은 명세이다. 여기 개념화라는 말의 의미는 뒤에서 상세히 설명한다. 물론 이와 같은 온톨로지가 그 자체로 웹 콘텐츠의 의미를 이해하도록 하는 것은 아니다. 우리가 시소러스도 미리 등록된 디스크립터를 이용하여 수작업으로 문헌을 색인하여 문헌주제의 상하관계 등을 표현한 후 검색 시 활용할 수 있듯이 온톨로지도 컴퓨터가 이해하도록 표현해 주는 언어인 RDF나 OWL 등으로 대상에 적용되어야 그 빛을 발한다.

위 내용의 이해를 돕기 위하여 RDF와 OWL에 대한 간략한 설명이 필요하다. RDF는 웹 자원에 관한 메타데이터를 표현하기 위하여 W3C가 개발한 언어로서 웹 상에서 “의미(semantic)” 정보를 손상 없이 응

2) Ontology (information science) (2009). In *Wikipedia*. Retrieved Aug. 1, 2009 from [http://en.wikipedia.org/wiki/Ontology\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science))

3) Sparck Jones, K. (1979). Problems in the representation of meaning in information retrieval. In *Informatics 5: The analysis of meaning* (pp. 193-201). London: Aslib을 보면 이유를 알 수 있음.

용프로그램 간에 상호 교환할 수 있는 공통의 틀을 제공하는 것이다. 한편 OWL은 온톨로지 선언문을 작성하는 언어로서 의미를 표현하는데 XML, RDF, RDFS보다 적합하며, 따라서 이들보다 웹 상의 기계가 독형 콘텐츠를 표현하는 능력이 탁월하다고 여겨진다. 이 언어들은 우리가 알고 있는 일반 프로그램 언어와는 달리 복잡한 논리적 관계를 표현할 수 있는 능력을 가지며 주로 XML 구문을 활용한다. 위에서 OWL이 RDF보다 의미표현에 적합하다는 것은 복잡한 온톨로지의 개념표현 및 논리적 관계를 표현하는데 좀더 최적화되어있다고 이해하면 된다.

## 온톨로지란 무엇인가?

온톨로지의 정의들 가운데 가장 자주 인용되는 것은 “온톨로지는 공유된 개념화(shared conceptualization)의 정형화되고 명시적인 명세(formal and explicit specification)”라는 그루버(Gruber, T.R.)의 것으로서, 그는 나아가 “지식의 공유(knowledge sharing)와 재사용(reuse)이 가능케 할 목적으로” 온톨로지를 만든다고 하였다<sup>4)</sup>. 또 위키피디아의 정의를 살펴보면, “전산학과 정보학에서 온톨로지는 한 영역 내에서 일련의 개념과 그 개념들 간의 관계에 대한 정형화된 표현(formal representation)이다. 이는 그 영역의 속성에 대한 추론(reason)에 사용되며, 그 영역을 정의하는데 사용될 수 있다 ... 하나의 온톨로지는 공유된 어휘(shared vocabulary), 즉 한 영역을 모델화 하는데 사용될 수 있는 어휘를 제공한다. 다시 말하여, 존재하는 객체(objects) 및/또는 개념(concepts)의 유형, 그리고 그들의 속성과 관계를 말한다”<sup>5)</sup>라고 기술한다.

두 정의는 설명방법의 함축성에 차이가 있을 뿐, 논리학, 언어학, 철학과 인공지능분야에서 유래한 용어가 많아 이들에 대한 사전 지식 없이 그 의미를 이해하기는 어렵다. 우선 “공유된(shared)”이란 대상 커뮤니티 및 그 구성원은 물론 컴퓨터 간에도 합의가 되었다는 의미로서 우리 분야의 MARC포맷을 생각해 보면 이해가 쉽다. 이는 온톨로지의 중요한 속성으로 이를 통하여 지식의 공유와 재사용이 가능할 것이다. “개념화(conceptualization)”는 특정 목적을 위하여 표현하고자 하는 대상세계(world)에서 일어나는 현상에 연관된 개념들을 파악하기 위한 추상적 모델(abstract model)<sup>6)</sup>을 의미한다. “정형화된(formal)”이란 컴퓨터가 읽고 처리할 수 있는 기계가독형이어야 한다는 것을 뜻하고, “명시적(explicit)”이란 개념의 종류와

4) Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43, 907-928. Retrieved on June 4, 2009 from <http://tomgruber.org/writing/onto-design.pdf>

5) Ontology (computer science) (2009). In *Wikipedia, the free encyclopedia*. Retrieved Aug. 1, 2009 from [http://en.wikipedia.org/wiki/Ontology\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science))

6) 일반적으로 우리는 추상화(抽象化)가 추상화(抽象畫)처럼 특정 대상을 사실적 표현이 아닌 원가 이해하기 어렵게 기술 또는 묘사한 것으로 받아들이기 쉬움. 하지만 여기서 추상화란, 예를 들면, 우리가 ‘문화(culture)’를 설명할 때, 종교적 측면, 경제적 측면 등 하위 범주로 나누어 기술하는데, 이런 분류는 실제 문화라는 개념의 하위구분이라기 보다는 우리가 편이를 위하여 문화를 ‘추상화’한 것이라 할 수 있음.

그들 간의 관계 및 그 개념들을 사용하는데 적용되는 제한점들(constraints)을 명백하게 정의하는 것을 의미한다<sup>7)</sup>. 그리고 “공유된 어휘”란 이러한 영역 내의 개념은 주로 어휘로 표현되기 때문이다.

결국 위 정의를 종합해 보면 온톨로지란 특정 영역이나 세계를 개념(즉 어휘)과 개념 간의 관계로 표현한 것이다. 여기서 영역 또는 세계는 우리가 도서관 환경에서 주로 다루는 특정 주제분야일 수도 있고, 특정 업무분야일 수도 있으며, 이를 확장하여 이론적으로는 세상 전체가 될 수도 있다<sup>8)</sup>. 이런 대상을 표현하는 방법으로 개념과 이들 간의 관계를 사용하는데(개념화), 이 표현은 사람이 아닌 컴퓨터가 이해하도록 구현되어야 한다(formal). 또한 이러한 관계는 컴퓨터가 이해하고 추론을 하기 위하여는 애매하지 않고 명백하게 정의해 주어야 하는 것이다(explicit). 마지막으로 이와 같은 절차를 통하여 온톨로지가 구성되어도 이들이 반드시 가져야 하는 속성은 “공유”이다(shared)<sup>9)</sup>.

이러한 설명을 종합해 보면 온톨로지는 우리가 익숙한 시소러스와 비슷한 구조 또는 특성을 가지는 것을 알 수 있고, 실제 그 모습도 시소러스와 매우 유사하다.

## 시소러스는 온톨로지의 한 종류인가?

우리가 새로운 대상을 이해할 때 이미 알고 있는 대상과 비교하여 유사점이나 차이점을 살펴보면서 다가가는 것도 좋은 방법이다. 온톨로지는 개념과 그들 간의 관계를 명시한다는 점에서, 그리고 개념을 어휘로 표현한다는 점에서 그 구조가 정보검색분야에서 60여년 전부터<sup>10)</sup> 사용해 오던 제어어휘집(controlled vocabulary)인 시소러스와 유사하다. 그렇다면 시소러스는 온톨로지의 한 종류이고, 온톨로지는 시소러스를 조금 복잡하게 만든 버전일 뿐일까? 이에 대한 답변은 단순하지 않다. 실제 웹에서 ‘thesaurus’와 ‘ontology’를 함께 검색하면 적지 않은 문헌을 발견할 수 있으며, 그 수만큼이나 의견도 분분하다. 근본적으로 이러한 혼돈은 전산 및 다른 분야에서 온톨로지라는 용어를 무분별하게 사용한 데 기인한다고도 할 수 있다.

어쨌든 앞서 언급한 온톨로지의 정의를 기반으로 <그림 1>의 온톨로지 스펙트럼을 보면서 결론을 내리지는 못하더라도 차이점이 무엇인지 살펴보자. 먼저 온톨로지는 개념과 개념과의 관계를 통하여 지식표

7) Fensel, D. (2001). *Ontologies: silver bullet for knowledge management and electronic commerce*. Berlin: Springer Verlag, p. 8을 주로 참조함.

8) 이와 같은 온톨로지는 “특정 영역(particular domain)”과는 상반되게 “일반 엔티티(general entities)”를 기술하는 것임.

9) 공유를 위하여 온톨로지는 여러 사람에 의하여 ‘동의’를 얻어야 한다는 전제조건이 있음. 그리고 한 온톨로지에 동의하고 받아들이기로 하는 것을 이 분야에서는 그 온톨로지에 ‘commit’ 한다는 표현을 씀.

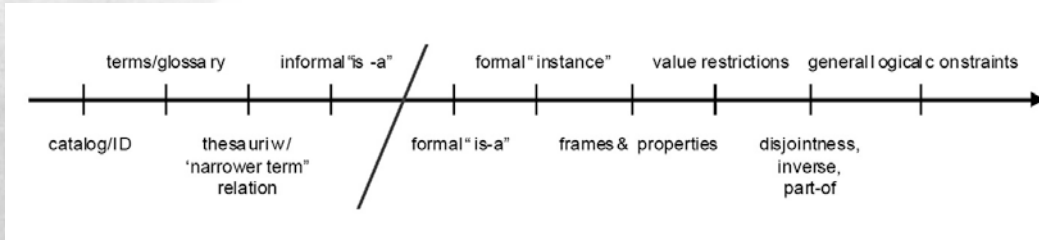
10) Aitchison, J. & Clarke, S. D. (2004). The thesaurus: a historical viewpoint with a look to the future. In S.K. Foe & A.R. Thomas (Eds.), *The Thesaurus: review, renaissance, and revision*. (pp. 5-21). Binghamton, NY: Haworth Information Press.

현을 한다고 하였는데, 개념은 흔히 용어(term)로 표현된다. 따라서 온톨로지는 용어의 의미를 명확히 용어명과 용어 의미의 구체적 명세를 제공하는 용도로 사용될 수도 있다. 이와 같이 온톨로지를 용어의 개념화 명세라는 관점에서 보면 <그림 1>과 같은 스펙트럼 상에 온톨로지의 다양한 종류를 개념에 따라 나열할 수 있다. 이 스펙트럼의 양극 중 왼쪽은 의미관계(semantic relations)가 미약하고, 오른쪽 극은 의미관계가 강한 것을 나타낸다. 여기서 의미관계가 강하고 약하다는 것은 기계가 이해하고 추론하도록 표현한 개념 및 논리관계의 상세성 또는 정확성의 정도라고 보면 이해하기가 쉬울 것이다.

이 스펙트럼에서 카탈로그는 의미관계가 제일 약하게 표현된 가장 단순한 형태의 온톨로지라 할 수 있다. 카탈로그의 특정 용어(예, 자동차)는 동일한 식별자(예, 25)를 사용하도록 한다. 예를 들면, 상품 카탈로그에서 식별자 1은 주방용품, 2는 전자제품 ... 25는 자동차인 셈이다. 이 관계는 그 카탈로그 내에서는 구분이 뚜렷하지만 이 관계로 기계가 의미를 해석하거나 단순한 것일지라도 추론을 하기에는 역부족임을 쉽게 알 수 있다. 그 다음으로 용어사전(glossary)은 용어와 그 의미를 수록한 리스트로서 자연어 설명은 사람들에게 용어에 대한 설명을 제공하므로 카탈로그 보다는 좀더 많은 정보를 제공한다. 하지만 이러한 자연어 설명은 사람에게는 이해가 쉽지만 컴퓨터에게는 애매하여 기계처리를 위한 명세로는 적합하지 않다. 시소러스는 이들 보다는 용어 간의 관계에 약간의 추가적인 의미(semantics)를 제공하는데, 광의어와 협의어 관계를 통하여 계층관계를 추론할 수 있지만 일반적으로 명확한 계층관계를 제공하지는 않는다<sup>11)</sup>. 그 다음 비정형(informal) “is-a”<sup>12)</sup>도 시소러스와 유사하게 계층관계를 제공하지만, 비정형이란 표현이 시사하듯이 논리적으로 명확한 “is-a” 관계는 아니라는 뜻이다. 다시 말하여, 명확한 “is-a” 계층체계에서는 만약 A가 B의 상위클래스(superclass)일 때, 한 객체가 B의 사례이면 이는 그 객체가 반드시 A의 사례여야 하는데 이 비정형 “is-a” 관계에서는 이 원칙이 항상 준수되지 않는다는 뜻이다. 결과적으로 시맨틱 웹을 위한 온톨로지의 핵심은 앞서 설명한 ‘정형적(formal)’이란 특성의 기계가독성을 가져야 하는데, 비정형 “is-a” 관계는 논리적 일관성의 결여로 기계가 이해하기 어렵게 된다. 이와 같이 엄격한 하위클래스(subclass) 계층구조는 프로그램에서 상속(Inheritance) 원리를 활용하는데 반드시 필요하며, 추론을 가능케 하므로 시맨틱 웹을 위한 온톨로지가 필요로 하는 특징이다.

11) Art and Architecture Thesaurus 초기 버전에는 Hagiography(聖人研究) BT Saints(聖人)로 정의되었는데, 성인연구는 성인의 한 종류가 아님. 즉, 일반적이거나 정확한 계층관계를 갖고 있지 못함. ‘명확한 계층관계’에 대한 좀 더 상세한 설명은 Fischer, D. H. (1998). *From thesauri towards ontologies?* Retrieved Aug. 11 from [http://www.ipsi.fraunhofer.de/orion/pubFulltexts/Fischer\\_1998.pdf](http://www.ipsi.fraunhofer.de/orion/pubFulltexts/Fischer_1998.pdf)를 참조할 것

12) 이는 instance of a (class)를 의미하며, subtype 또는 subclass라고도 함.



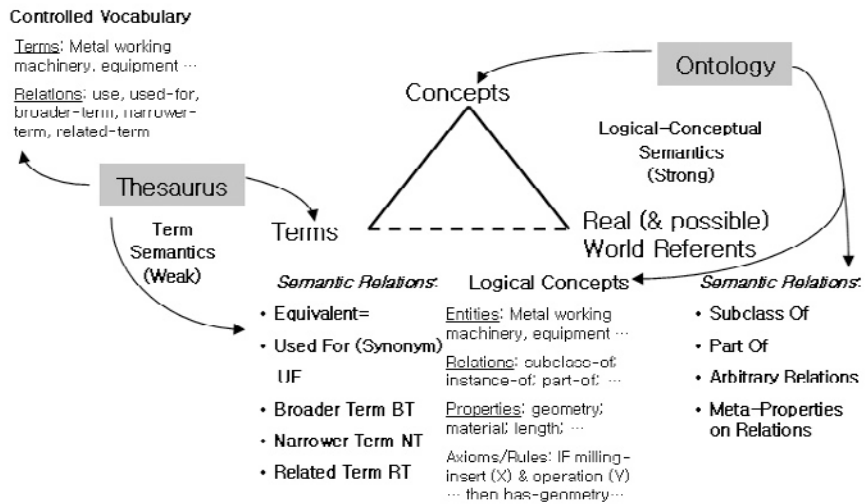
〈그림 1〉 온톨로지 스펙트럼<sup>13)</sup>

보다 확실한 이해를 위하여 계층구조에서 비형식적이라는 뜻을 예를 들어 살펴보자. 야후에서 ‘apparel(의복)’은 하위에 ‘women(숙녀)’를 포함하고, 이 범주는 다시 하위범주로 ‘accessories(액세서리)’와 ‘dresses(드레스)’가 있으며, ‘accessories(액세서리)’ 아래 ‘fragrance(향수)’가 있다. 여기서 ‘apparel(의복)’이 ‘women(숙녀)’를 포함하지만, 이는 정확히 ‘women’s apparel(숙녀복)’이어야 하며, ‘women(숙녀)’는 ‘apparel(의복)’의 한 사례가 아니다. 이어서 ‘accessories(액세서리)’와 ‘dresses(드레스)’도 ‘women(숙녀)’의 사례가 아니기는 마찬가지이며, 숙녀 ‘accessories(액세서리)’의 한 사례인 ‘fragrance(향수)’는 ‘apparel(의복)’의 사례가 아니다. 이와 같이 전문화(specialization)된 계층관계로 일반화(generalization)될 수 없는 관계를 비형식적이라 하며, 이런 계층관계는 기계가 이해하기 어려울 뿐 아니라, 그 온톨로지를 공유하는 것 또한 문제가 있다.

지금까지 설명한 〈그림 1〉의 슬래쉬("/") 앞부분에 속하는 대상을 온톨로지로 간주하는 이는 소수이며, 일반적으로 온톨로지로 간주되기 위하여는 명확한 계층관계(explicit hierarchy)를 가지는 것을 선호한다. 따라서 시소러스는 온톨로지의 한 종류인가라는 초기 질문에 대한 답은 각자가 결정할 수 있겠으나, 흔히 말하는 시맨틱 웹에서 활용하는 온톨로지의 한 종류로는 적합하지 않다고 하겠다.

그럼 시소러스에 존재하는 계층관계가 명확하지 못한 것이 우리가 주로 말하는 온톨로지와의 유일한 차이일까? 〈그림 2〉가 이 둘의 차이점을 좀 더 상세히 보여주고 있다. 온톨로지는 개념표현을 단순한 용어와 그들과의 관계보다 다양한 사실을 표현할 수 있는 장치를 가지고 있다고 할 수 있다. 여기서 다양한 사실이란 개념의 속성, 명확한 논리적 관계, 그리고 이들에 대한 공리와 원칙을 정의하여 복잡한 의미정보를 포함할 뿐 아니라, 컴퓨터의 자동 추론과 계산을 가능하도록 해 준다.

13) Lassila, O. & McGuinness, D. (2001). *The role of frame-based representation on the Semantic Web*. Stanford, CA: Knowledge Systems Laboratory. Retrieved Aug. 5, 2009 from <http://www.ida.liu.se/ext/epa/ej/etai/2001/018/01018-etabody.pdf>



〈그림 2〉 시소러스 vs. 온톨로지<sup>14)</sup>

## 온톨로지의 분류

온톨로지라는 용어가 다양하고 폭넓은 대상(referents)에 무차별하게 사용되어 그 정확한 의미전달이 어렵자 “온톨로지” 앞에 수식어를 부가하여 지칭하는 대상을 구분하였다. 그러나, 이런 명칭의 온톨로지들 간의 차이도 구분하기 어렵자 이들 간의 특성을 비교함으로써 차이점을 제시하기 위한 분류가 시도되었다. 소와(Sowa, J. F.)는 〈그림 1〉과 유사하게 의미(semantics) 표현의 강도에 따라 온톨로지를 다음 세가지로 구분하였다<sup>15)</sup>.

- 온톨로지(Ontology): 특정 도메인 또는 주제분야에 존재하거나 존재 가능한 엔티티 종류의 명세이다. 형식적으로 온톨로지는 개념명칭(names for concept)의 집합과 유형-하위유형 관계(type-subtype relation)에 기반한 일부분에 대한 순위화(partial ordering)로 조직된 관계유형들로 명시된다. 한 온톨로지 내에 존재하는 일부 또는 모든 개념과 관계에 형식공리(formal axiom)와 정의를 연관시킬 수 있다.

14) *Ontology spectrum*. Retrieved May, 2005 from [http://www.xml.saic.com/icm1/main/ic\\_thesaurus/IC-thesaurus.CONOPS-1.1.doc](http://www.xml.saic.com/icm1/main/ic_thesaurus/IC-thesaurus.CONOPS-1.1.doc)

15) Sowa, J. F. (1996). *Ontologies for knowledge sharing*. Manuscript of the invited talk at TKE '96, Vienna 1996(From thesauri towards ontologies? / Fischer, D. H. 1998에서 재인용)



- 용어 온톨로지(Terminological ontology): 모든 개념과 관계가 이들의 활용을 위한 필요충분조건을 결정하는 공리(axioms)<sup>16)</sup>나 정의에 의하여 명시되지 않는(not fully specified) 온톨로지이다. 개념은 하위유형-상위유형(subtype-supertype) 또는 부분-전체(part-whole) 관계에 의하여 부분적으로 명시될 수 있으며, 이러한 관계는 개념 사이에서 특정 개념의 상대적 위치를 결정할 뿐, 완전하게 정의를 하지는 않는다.
- 공리 온톨로지(Axiomatized ontology): 개념과 관계가 논리(logic) 또는 논리로 자동 번역되는 어떤 컴퓨터지향 언어(some computer-oriented language)로 선언된 관련 공리와 정의를 가진 용어 온톨로지이다. 공리와 정의를 선언하기 위하여 사용되는 논리의 복잡성에 대한 아무런 제한이 없다.

결국 소와의 분류에 포함된 온톨로지의 종류는 앞선 <그림 1>에 구분된 온톨로지들의 특정 범위에 명칭을 부여한 것으로 볼 수 있다. 정확하지 않을 수 있고 해석의 차이가 존재할 수 있지만, 위 분류에서 온톨로지는 <그림 1>의 카탈로그와 용어사전에 가깝고, 용어 온톨로지는 시소러스와 비정형 “is-a” 범주에 속하는 온톨로지를 의미하고, 공리 온톨로지가 슬래쉬 우측에 위치한 범주의 온톨로지를 의미한다고 이해할 수 있다.

반면, 팬젤(Fensel, D.)은 온톨로지의 일반성 수준(generality level)<sup>17)</sup>이 낮은 것에서 높은 순으로 다음 네 가지로 구분하고, 이외의 두 가지 다른 형태를 제시하였다.

- 영역 온톨로지(Domain ontology): 전자, 의학, 기계, 디지털 영역과 같이 특정 영역에만 통용되는 지식을 표현한 온톨로지이다.
- 메타데이터 온톨로지(Metadata ontology): 더블링크어와 같이 온라인 정보원의 내용을 기술하기 위한 어휘를 제공하는 온톨로지를 지칭한다.
- 일반 온톨로지(Generic or common sense ontology): 시간, 공간, 상태, 사건 등과 같은 사물의 기본적인 관념과 개념을 제공하는 세상에 대한 일반 지식을 포함하는 온톨로지이다. 결과적으로 이 형태의 온톨로지는 여러 영역에 공통으로 적용될 수 있다. 예를 들면 부분전체론(mereology)에 관한 온톨로지는 많은 기술적 영역에 적용 가능하다.
- 표현 온톨로지(Representational ontology): 영역 온톨로지처럼 특정 영역을 대상으로 하지 않는다.

16) 공리(axiom): 논리학 용어로 본래의 특성이나 자명함에 의거하여 이제껏 일반적으로 받아들여왔거나 또는 일반적으로 받아들일 만하다고 보이는 증명이 불가능한 제1원리·규칙·준칙 등을 말함.

17) 일반성이란 특수성과 대비되는 개념으로 이론을 예로 설명하면 일반이론(general theory)은 어떤 가정이나 제약없이 모든 경우에 적용되는 이론을 의미하는 반면, 특수이론(special theory)은 그 이론이 성립하려면 전제조건을 만족하는 경우를 칭함.

무엇을 표현해야 하는지 언급하지 않고 표현개체(representational entities)를 제공하는 온톨로지이다. 유명한 표현 온톨로지는 그루버의 Frame Ontology으로서 이는 지식표현이 개체지향 또는 프레임 기반 방식으로 표현가능하도록 프레임, 슬롯(slots)과 슬롯 제약과 같은 개념을 정의한다.

이와 더불어 팬젤은 방법 온톨로지(method ontology)와 과업 온톨로지(task ontology)<sup>18)</sup>를 제시하였다. 전자는 특정 문제해결방법(Problem Solving Methods)에 제한된 용어(terms)를 제공하고, 후자는 특정 과업에 제한되는 용어를 제공한다.

그 밖에 일반성을 기준으로 맨달(Mandal, C.)은 최상위 온톨로지(Top-level ontology), 영역 온톨로지, 과업 온톨로지와 응용 온톨로지(Application ontology) 순으로 나누었으며<sup>19)</sup>, 애셔프(Aschoff, F. R.)는 표현 온톨로지, 일반 온톨로지, 영역 온톨로지와 응용 온톨로지로 구분하였다<sup>20)</sup>. 위에 제시된 온톨로지 형태 중 이 글에서는 처음 나온 최상위 온톨로지와 응용 온톨로지에 대한 설명이 필요하다.

- 최상위 온톨로지: 이는 상위 온톨로지(upper ontology) 또는 기초 온톨로지(foundation ontology)라고도 하며, 모든 영역에 걸쳐 동일한 아주 일반적인 개념을 기술한 것이다. 정확한 대칭관계를 이루지는 않겠지만, 영역 온톨로지를 우리가 아는 시소러스 종류 중 특정 주제분야의 시소러스인 마이크로 시소러스(micro thesaurus)라고 한다면 최상위 온톨로지는 모든 주제를 망라하는 매크로 시소러스(macro thesaurus)와 유사하다고 하겠다. 그러나 매크로 시소러스가 모든 분야의 제어어휘를 제공한다면, 최상위 온톨로지는 어느 하나의 특정 문제 영역에 속하지 않는 일반 개체를 기술하는 개체의 계층과 연관법칙을 기술한다는 것이 차이점이다. 다시 말하여 최상위 온톨로지의 가장 중요한 기능은 모든 영역을 지원하는 온톨로지 자체를 제공하는 것이 아니라, 많은 온톨로지들이 최상위 온톨로지 “아래에서(under)” 매우 포괄적인 의미 상호운용성을 가질 수 있도록 지원하기 위한 것이다. 이러한 온톨로지의 구축이 가능한가에 대한 논쟁은 아직도 진행 중이다<sup>21)</sup>.
- 응용 온톨로지: 일반성에 있어 가장 협의적(specific) 온톨로지로서 예를 들면 의학분야 중 급성 방사선 증후군 치료와 같이 특정 응용분야에 대한 지식을 표현한다.

18) 지식공학에서 task와 method의 구분으로 전자는 문제의 유형(type of problem)을 의미하며, 후자는 문제를 해결하는 수단을 의미함.

19) Mandal, C. (2005). *Statutes on the semantic web*. Retrieved on Aug. 23, 2009, from <https://www.cs.tcd.ie/courses/cs11/mandal/c0405.pdf>

20) Aschoff, F. R. (2004). *Knowledge Mediation: A Procedure for the Cooperative Construction of Domain Ontologies*. Unpublished doctoral dissertation, Ruprecht-Karls-University, Heidelberg.

21) Upper ontology (information science) (2009). In *Wikipedia, the free encyclopedia*. Retrieved July 1, 2009 from [http://en.wikipedia.org/wiki/Upper\\_ontology\\_\(information\\_science\)](http://en.wikipedia.org/wiki/Upper_ontology_(information_science)) 을 참조하면 이들 논쟁의 쟁점이 요약되어 있음.

이와 같은 온톨로지에 대한 구분은 주장하는 이에 따라 다소 차이가 있으나 옳고 그름을 판단하기 위한 것이 아니다. 그러므로 어느 분류가 가장 옳은 것인가를 고민하기 보다는 앞서 지적한 바와 같이 다양하고 때로는 무분별하기까지 한 온톨로지 관련 용어에 대하여 개념정리를 시도한 것으로 받아들이면 적절할 것이다.

## 결언

온톨로지라는 용어를 피상적으로 접한 초기에는 “조금 복잡한 시소러스구나”라는 느낌과 더불어 수많은 전문용어로 인하여 혼란스러웠다. 그래서 가능하면 문헌정보학 배경을 가진 사람이 그 기본개념만이라도 이해할 수 있도록 설명하려 하였지만, 제한된 필자의 지식과 할당된 지면 탓에 그것마저 무모한 시도였는지도 모르겠다. 물론 이 과정에서 지나친 단순화, 부족한 용어설명 등 만족스럽지 않은 점이 있겠지만, 온톨로지라는 개념에 대하여 독자들이 “가까이 하기엔 너무 먼 당신”은 아니라는 느낌을 가졌으면 한다.

이 글을 읽고 온톨로지에 대한 이해와 판단은 독자마다 차이가 있을 수 있겠지만, 필자는 그루버가 언급한 온톨로지 구축의 가장 기본적인 목적인 “지식의 공유와 재사용”을 생각하면, 문득 10여 년 전 국내에서 XML이 유행하던 시기가 떠오른다. 당시 XML도 지금 온톨로지와 상당히 유사한 기치아래 정보의 호환성, 재활용성 등을 외치며, 여러 분야에서 유행처럼 퍼져 나갔다. 하지만 XML 그 자체는 이를 위한 기반을 갖추고 있는 것이지 XML로 구축된 모든 데이터, 문서 등이 재활용성은 물론 호환성을 무조건 가지는 것은 아니다. 당시 우리 분야에서 많은 XML 문서와 메타데이터들이 구축되었는데, 이들의 문제는 통일된 DTD<sup>22)</sup>가 아닌 독자적으로 개발하여 원래 목표인 재활용성과 호환성 모두를 상실한 것이 대부분이었다. 온톨로지도 각자가 구축한다고 지식의 공유와 재사용이 가능한 것은 아니라는 점을 주지할 필요가 있다.

이 글을 통하여 짧게나마 우리는 온톨로지와 수인사(修人事)(?)를 나눈 셈이다. 이제 온톨로지가 우리 분야에 어떤 영향을 미치며, 이를 적용하였을 때 이점이 무엇인가, 또 우리가 사서로서 할 일이 무엇인가 등 보다 현실적인 문제를 생각해 보아야 할 것이다. (㉞)

22) DTD는 문서유형정의(Document Type Definition)의 약자인데, 원래 SGML에서 사용하던 기술로서 태그와 속성, 요소간의 상하관계 등 구조를 정의하는데, XML에서도 새로운 문서구조를 만들 때 DTD를 사용하였음. DTD는 SGML과 호환성 측면에서는 장점을 가지나 XML 문법을 따르지 않을 뿐더러 몇 가지 문제가 존재하여 XML에서는 문서의 구조를 정의하기 위하여 XML Schema라는 기술을 개발하여 사용하고 있음.