온톨로지와 웹 온톨로지

Ontology and Web Ontology

고 영 만

성균관대학교 문헌정보학과

http://slisnet.skku.ac.kr/~ymko

Email: ymko@skku.ac.kr

온톨로지 (1/3) - 용어의 유래

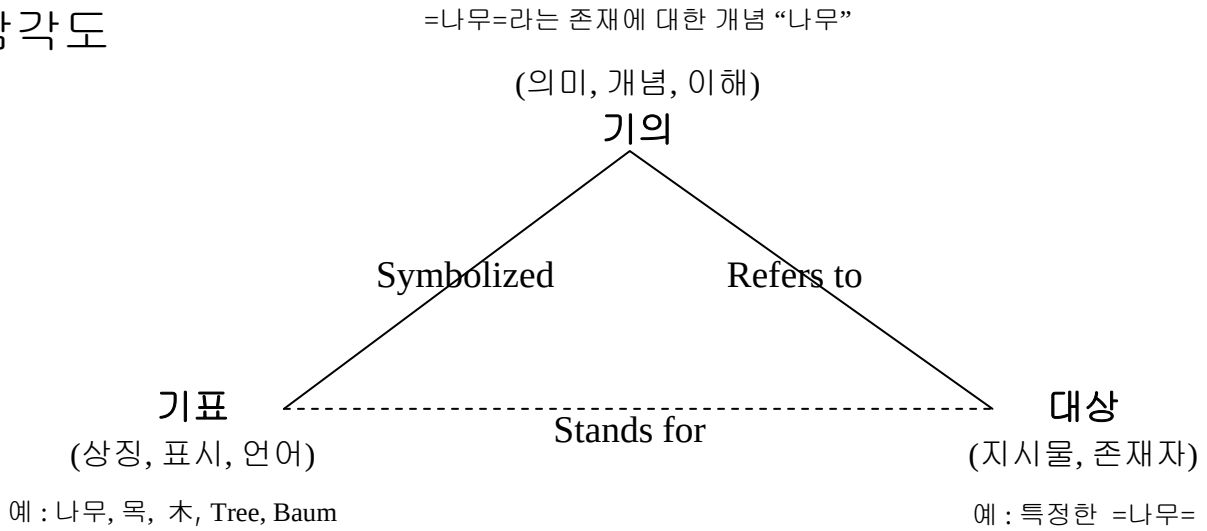
- 철학의 일부분으로 존재론 또는 존재학(存在學)이라고도 함
  - 라틴어로는 'ontoligia'라고 하며, 이것은 그리스어의 'on(**存在子**/존재하는 것)' 과 'logos(논)'로 이루어진 합성어로 데카르트파의 철학자 J.클라우베르크 (1622~1665)가 처음으로 사용
- 존재와 존재자(존재하는 것)의 본성을 연구하는 형이상학의 일부분,

  세상의 구성 요소에 대한 명확한 이해를 연구하는 철학의 일부분
  - 어떤 존재자에 대한 우리의 모든 이해와 태도를 이끌어가는 본질을 연구
  - 존재적 진리(Ontisch Wahrheit)
    - 존재하는 것에 대한 이해는 존재하는 것의 존재 방식에 대한 이해에 의해 유도됨
    - 예 : 곰보도 보조개 (얼굴의 흠에 대한 이해가 그 사람의 나에 대한 존재 방식에 의해 조종됨)

      - 이규호(1972)

온톨로지 (2/3) – 언어학적 전이

※ 의미삼각도



- 존재론(Ontology)의 학문적 연구결과물을 온톨로지로 전이해서 사용
  - 공유된 개념(화)의 정형적, 명시적 명세
    - 사람의 마음 속에 존재하는 내재적 생각이나 외재적 세계의 현상에 대하여 공유하고 있는 개념을 명확하고 명시적으로 정의하고 규정하는 것
    - 용어의 의미 관계와 연결 정보를 보다 유동적이고 상세하게 기술하기 위해 시소러스의 확장 개념으로 사용
      - 시소러스 : 계층적이고 고정된 형식으로 용어의 정의와 개념 간의 관계를 설정

온톨로지 (3/3) – 컴퓨터 분야에서의 차용

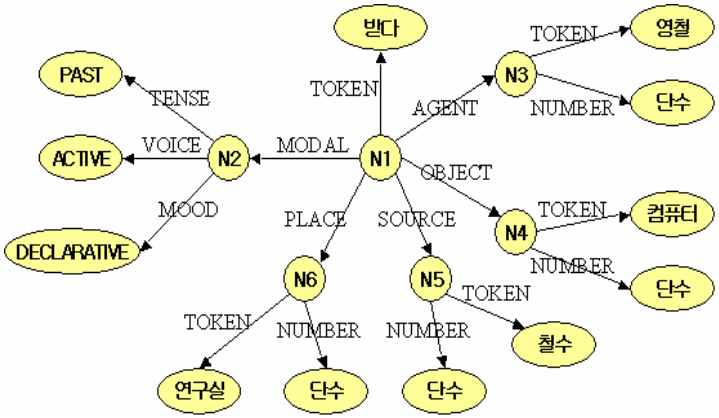
공유된 개념(화)의 정형적, 명시적 명세

- 사람의 마음 속에 존재하는 내재적 생각이나 외재적 세계의 현상에 대하여 공유하는 개념을 컴퓨터가 이해할 수 있는 형식으로 명확하고 명시적으로 정의하고 규정하는 것
- 연구분야
  - 데이터베이스 분야 : 메타데이터 온톨로지
    - 메타데이터 요소와 셋의 공유 및 표준화 → 메타데이터 레지스트리
  - 시맨틱 웹 분야 : 웹 온톨로지 (메타데이터 온톨로지의 일종)
  - 인공지능 분야 : 의미망 온톨로지 (Semantic Net Ontology)
    - 시소러스와 의미망의 결합을 통한 지식표현(Knowledge Representation)과 추론 기술 개발

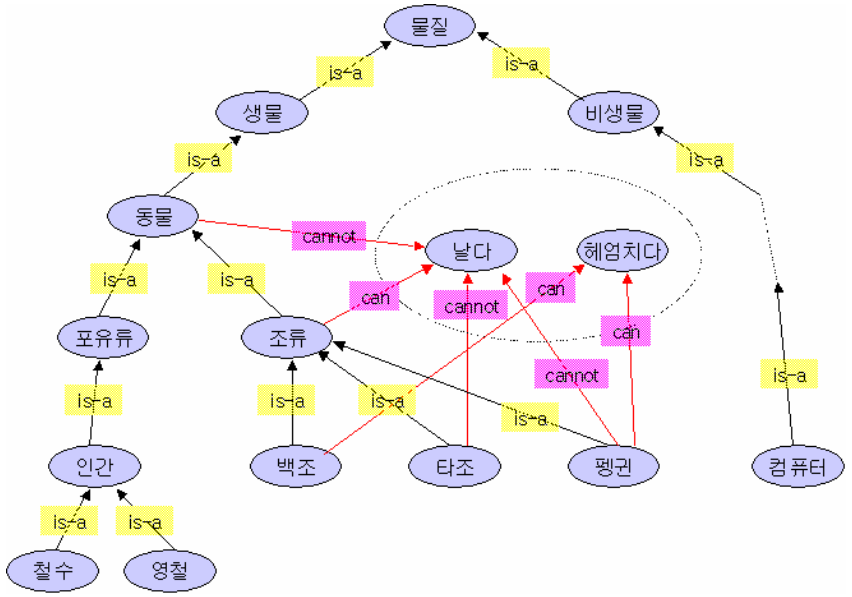
온톨로지 개발 (1/3) – 언어학 분야

- 시소러스(Thesaurus)
  - 의미망 구축의 기초
  - 문헌정보학의 전통적 연구 영역
- 의미망(Semantic Network)
  - 형식적인 면에서 시소러스와 유사하나, 한 어휘가 가지고 있는 다른 어휘들간의 관계를 망(network)으로 나타낸 것
    - 단어의 의미, 개념간의 연상관계 등을 표현하는 네트워크로서 노드와 링크, 관계 표시 등의 결합으로 이루어진 도식적인 표현(diagrammatic representation)으로 이루어짐
    - 지식표현의 관리 알고리즘을 이용한 다양한 추론 기술 및 데이터베이스와 같은 역할을 고려한 전산적 표현(computational representation)으로 이루어짐
  - 언어학(자연어 처리분야) + 전산학 => 인공지능 분야 연구 영역

온톨로지 개발 (1-1/3) – 시맨틱 네트웍(의미망)



의미망에 의한 문장 표현



의미망에 의한 개념 표현

옥철영 (2004)

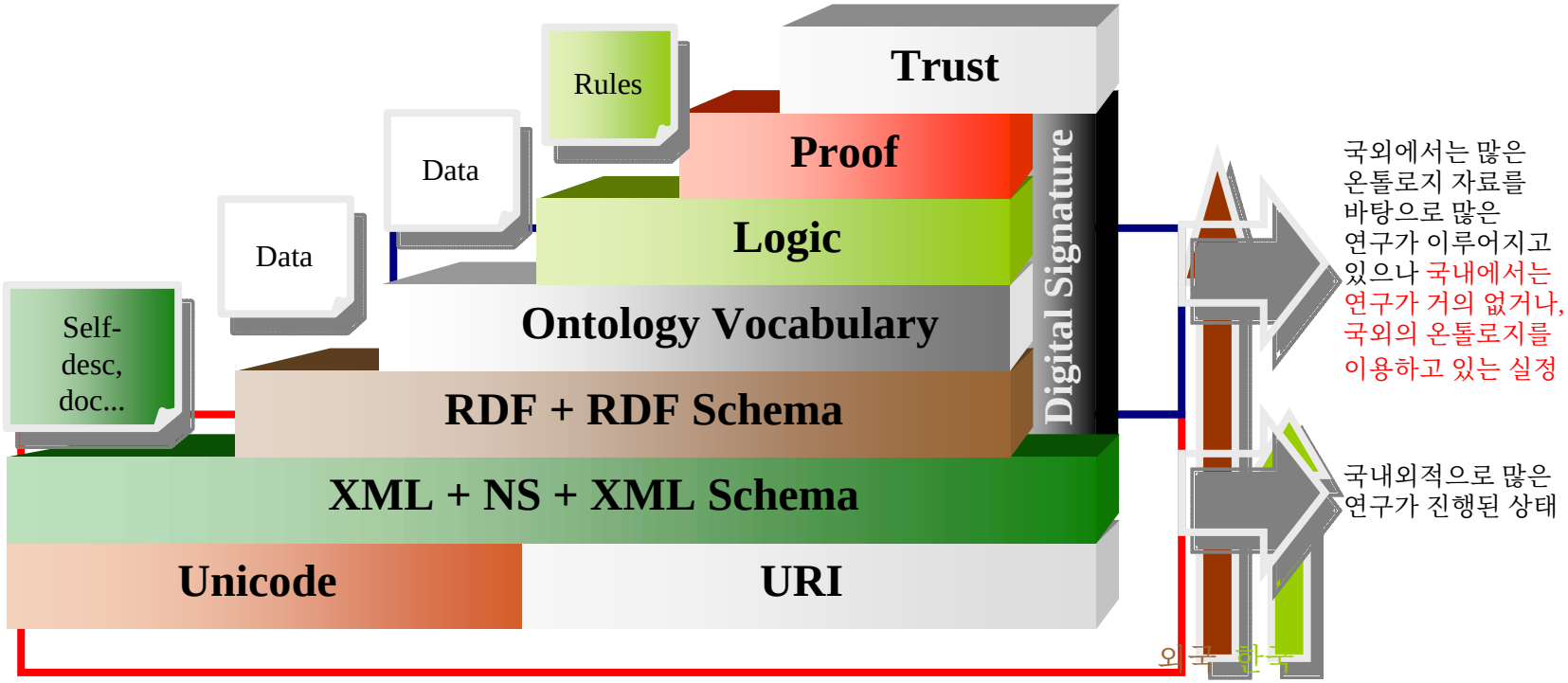
온톨로지 개발 (2/3) - 메타데이터 분야

- 메타데이터 온톨로지
  - 서로 다른 데이터베이스가 같은 개념에 대해 서로 다른 식별자(태그) 혹은 서로 다른 단어를 사용할 경우 이를 해결해 주기 위해 공유되는 개념화를 정형적, 명시적으로 명세화 한 집합체
    - postal code와 zip code를 사용하는 두 개의 DB에 들어있는 정보를 비교하거나 통합하려는 프로그램에서는 postal code와 zip code가 같은 의미를 지칭하는 메타데이터 요소/식별자라는 것을 알아야 함
  - MARC, 더블린코어와 같은 메타데이터 세트의 호환을 위한 메타데이터 레지스트리(MDR) 구축에 중점을 두고 있음
    - 사전 예방에 의한 문제 해결 지향
    - 최근 확장 메타데이터 레지스트리(XMDR)에 대한 연구가 진행중

온톨로지 개발 (3/3) – 시맨틱 웹 분야

- 시맨틱 웹
  - 1999년, Tim Berners-Lee (W3C)
    - 2001년 2월 공식적으로 Semantic Web Activity 착수
  - 웹상에 존재하는 자료에 의미를 부가하고 사람이 관여하지 않아도 컴퓨터가 자동으로 처리할 수 있는 차세대 지능적인 웹
    - 웹상에 존재하는 정보들을 사람뿐만 아니라 컴퓨터 프로그램 같은 기계들이 해독하고 작업하기 용이하게 표현
    - 정보간의 유기성까지 체계적으로 표현 : 정보공유 체제 마련
    - 자연어 위주의 기존 웹 문서와 달리 컴퓨터가 해석하기 쉽도록 의미를 부여한 계층을 가짐
  - 웹에 의미(Semantic)를 부여한다는 것
    - 사용자 인터페이스를 위한 자연어 처리 기능을 부가하고자 하는 것이 아니라 컴퓨터가 처리하기 용이하게 하고자 선언적인 추가 정보를 부여하는 것을 뜻함
- 웹 온톨로지
  - 메타데이터 온톨로지의 일종으로서 시멘틱 웹의 중심 개념 중 하나
  - 웹 문서를 생성하는 마크업 언어에서 정의된 동일 의미의 다른 명칭 식별자(태그) 또는 같은 내용을 다른 구조로 정의하는 식별자 등에 의해 발생되는 호환상의 문제를 해결하기 위해 공유되는 개념화를 정형적, 명시적으로 명세화한 집합체
  - 온톨로지 언어(XML 기반의 RDF, DAML+OIL, OWL, Topic Map 등)의 개발에 중점을 두고 있음 : 사후 처방에 의한 문제 해결 지향

시맨틱 웹에서의 온톨로지 (1/7) - 시맨틱 웹의 계층구조



- W3C (재인용, 옥철영 2004)

시맨틱 웹에서의 온톨로지 (2/7) - 시멘틱 웹과 XML

- HTML에 대한 XML의 장점
  - 출력형식만을 정의하는 HTML에 비해 잘 정의된 구조화 문서 작성
  - 의미정보를 나타낼 수 있는 태그 이름을 자유롭게 정의하여 사용
- XML의 표현방법이 시멘틱 웹 구현에 어려운 점
  - 문서 요소들의 구조적 정의만을 제공하며 정보 리소스 사이의 의미적 관계를 정의하지는 않는다.
  - 태그의 사용에 대한 규칙은 DTD 혹은 XML 스키마에 의해 정해질 수 있으나 이들의 의미를 컴퓨터가 이해하지는 않는다.
    - 서로 다른 사람이 같은 문서를 작성할 때 같은 의미를 뜻하는 다른 이름으로 태그를 정의하여 사용할 수 있음
    - 같은 내용에 대해 다양한 구조의 XML 문서를 사용할 수 있음 (예 ; 구조를 세분화 해서 사용)
      - <date>12/03/59</date>
      - <date>
        <day>12</day>
        <month>03</month>
        <year>59</year>
        </date>

시맨틱 웹에서의 온톨로지 (3/7) - 시맨틱 웹과 RDF

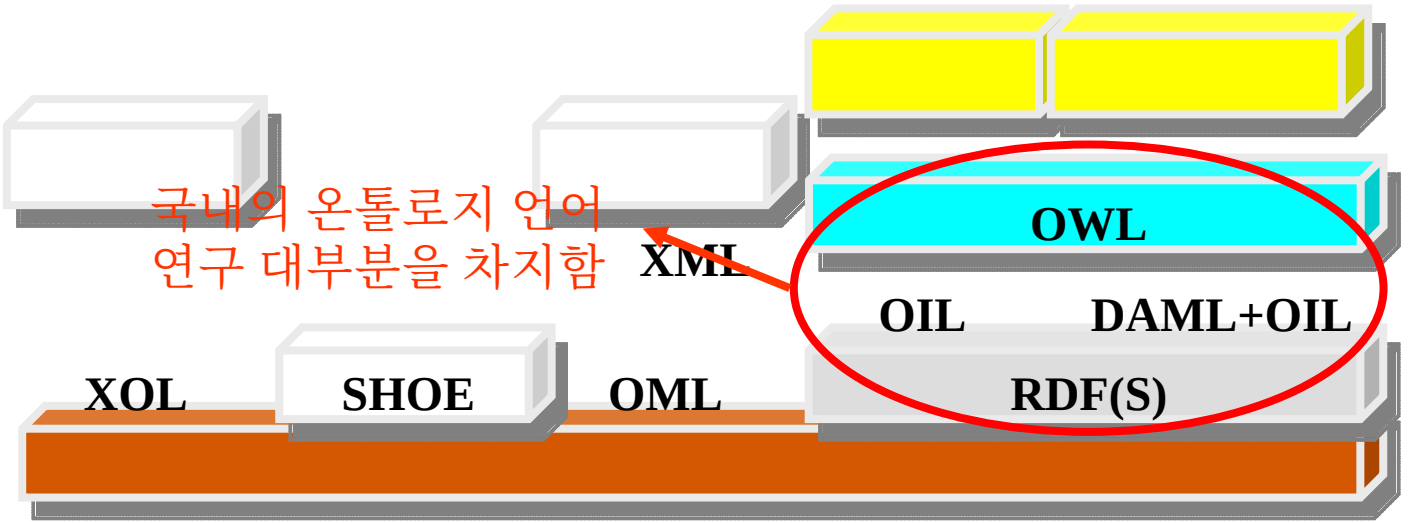
- XML에 대한 RDF의 잇점
  - 구조 측면에서 XML에서와 같이 서로 다른 구조를 가진 여러 가지 표현방법이 존재하지 않으므로 같은 내용에 대해서는 해석이 하나로만 귀결됨

- RDF가 시멘틱 웹 구현에 부족한 점
  - 의미 측면에서 XML의 문제점인 태그 이름의 중첩성과 모호성이 여전히 존재
    - XML에서와 마찬가지로 서로 다른 태그이지만 실제로는 같은 의미일 수 있음
    - XML에서와 마찬가지로 같은 태그이지만 사용자에 따라 다른 의미로 사용될 수 있음

시맨틱 웹에서의 온톨로지 (4/7) – 웹 온톨로지 언어

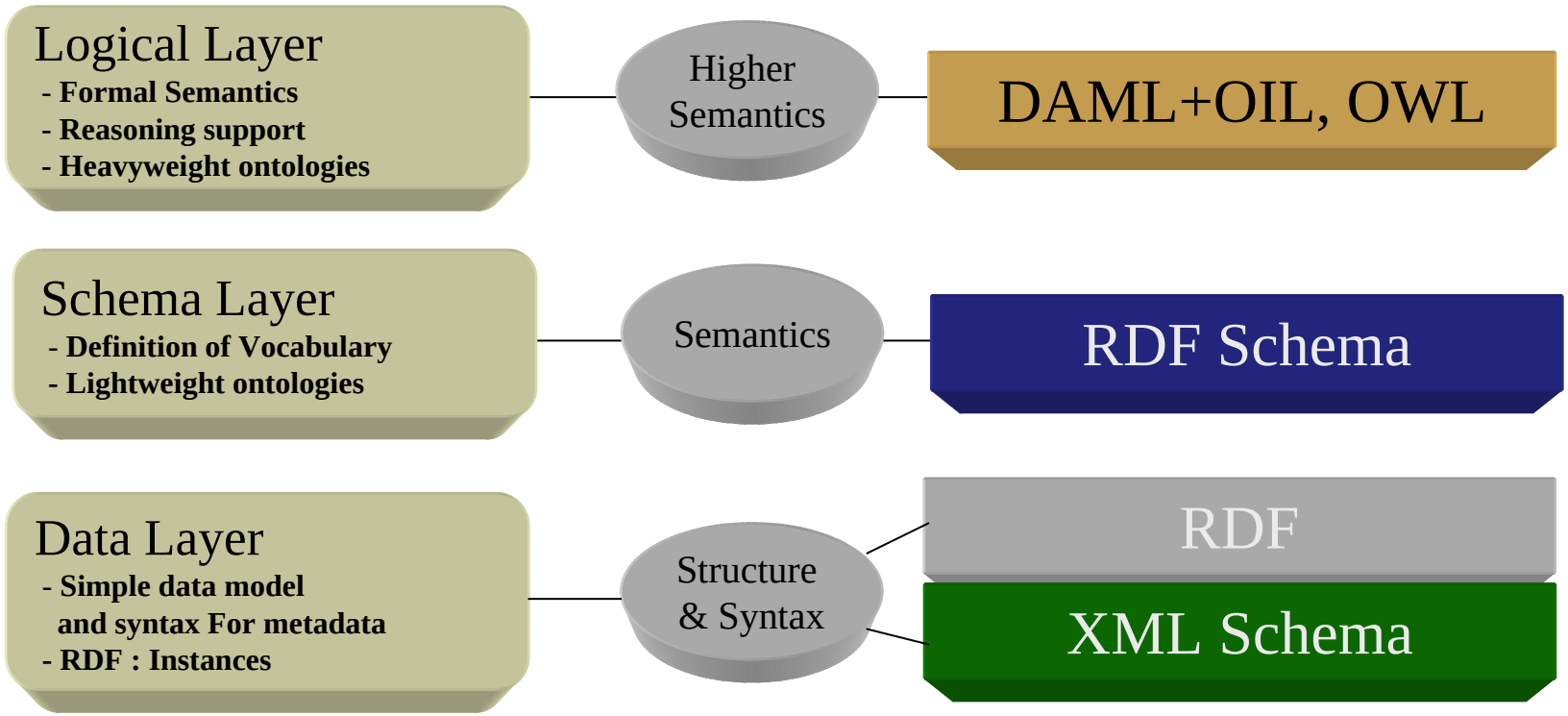
- 주로 XML 기반의 온톨로지 언어를 이용한 웹 온톨로지 연구에 집중
  - RDF 스키마와 유사, 그러나 보다 확장된 개념

– 옥철영 (2004)



- XOL(Xml-based Ontology Exchange Language)
- SHOE(Simple HTML Ontology Extension)
- OML(Ontology Markup Language)
- RDF(Resource Description Framework)
- DAML(DARPA Agent Markup Language)
- OIL(Ontology Inference Layer)
- OWL(Ontology Web Language)

시맨틱 웹에서의 온톨로지 (5/7) – 웹 온톨로지 아키텍처



– 김홍기 (2002)

시맨틱 웹에서의 온톨로지 (6/7) - 온톨로지 영역

- 시맨틱 웹의 완성에 필요한 온톨로지 영역
  - **데이터베이스 메타데이터 온톨로지**
    - 데이터베이스의 메타데이터 요소의 의미 중첩성과 상이한 메타데이터 셋에 따른 모호성 및 태그 구조의 다양성 해소
    - XML과 RDF를 기반으로 하는 데이터베이스의 연계성
  - **웹 온톨로지**
    - 상이한 형식의 웹 문서 태깅에 수반되는 의미의 중첩성과 구조의 다양성 해소
  - **(용어)온톨로지**
    - 시소러스와 의미망의 연계를 통한 의미망 온톨로지 완성
    - 궁극적으로 웹 온톨로지가 (용어)온톨로지에 포함되는 온톨로지 구축

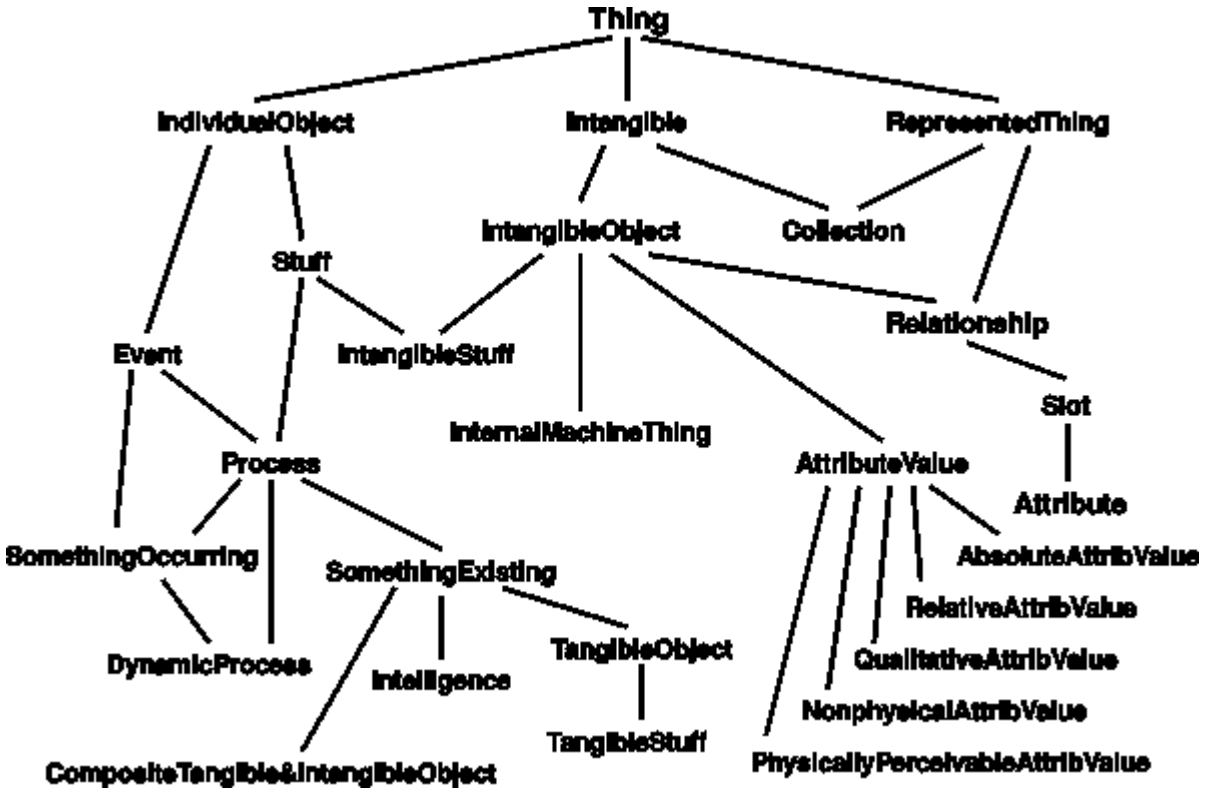
시맨틱 웹에서의 온톨로지 (7/7) - 웹 온톨로지 표준화

- W3C
  - RDF
  - DAML-OIL (DARPA Agent Markup Language-Ontology Interface Layer)
  - OWL
- ISO/IEC
  - Topic Maps

온톨로지의 종류 (1/3) – 시스템 측면

- 언어온톨로지 (Linguistic Ontology)
  - 텍스트, 말뭉치 등에서 추출된 데이터나 정보에 대한 자연어 인터페이스를 지원하는 온톨로지
    - CYC 온톨로지, EDR, WordNet
- 공리온톨로지 (Axiomatized Ontology)
  - 정보시스템과 관련된 영역의 부분적 이해를 표현하는데 이용되는 규칙, 이론, 제한점 등의 스키마나 어플리케이션을 자동으로 생성시키는 것을 지원하는 온톨로지
    - 웹 온톨로지
- 언어온톨로지 + 공리온톨로지
  - 인공지능 온톨로지
    - 시소러스 + 의미망(semantic net) => 의미망 온톨로지(?)

온톨로지의 종류 (1-1/3) - 언어온톨로지 사례 (CYC)



J.F. Sowa (2001)

온톨로지의 종류 (1-2/3) - 공리온톨로지 사례

FORALL X, Y

X : Researcher[cooperate With->>Y]<->

Y : Researcher[cooperate With->>X].

FORALL X,Y

X : Person[Publication->>Y]<->

Y : Publication[author->>X].

FORALL X,Y

X : Person[editor->>Y]<->

Y : Publication[editor->>X].

FORALL X,Y

X : PhDStudent[supervisor->>Y]<->

Y : AcademicStaff[supervises->>].

FORALL X,Y

X : JournalArtical[journal->>Y]<->

Y : Journal[containsArticle->>X].

온톨로지의 종류 (2/3) - 구축 범위에 따른 구분

- 일반온톨로지(generic or common-sense ontology)
  - 우리 주위를 둘러싸고 있는 세상을 구성하는 일반적 개념들을 대상으로 구축한 온톨로지
  - 공간, 상태, 사건, 시간과 같은 일반적이고 포괄적이 지식에 대하여 의미론적 연관관계를 구축한 개념의 집합체
    - CYC http://www.opencyc.org/
    - EDR http://www.iijnet.or.jp/edr/, http://www2.nict.go.jp/kk/e416/EDR/index.html
    - WordNet http://wordnet.princeton.edu/
- 영역온톨로지(domain ontology)
  - 특정 (주제)영역에서 유효한 지식들을 대상으로 구축한 온톨로지
  - 구체적인 사물 혹은 특정 학문영역과 같은 제한된 영역에서 의미론적 연관관계를 구축한 개념의 집합체

온톨로지의 종류 (3/3) – 구축 대상에 따른 구분

- 메타데이터 온톨로지
  - 데이터베이스 정보의 의미론적 연결을 위한 기술 어휘와 기준의 집합
- 웹온톨로지
  - 넓은 의미에서 메타데이터 온톨로지의 한 종류
  - 웹 문서 작성 언어의 태깅과 정의에 필요한 기준과 기술 방식의 의미론적 연결
- 표현온톨로지 (representational ontology)
  - 특정 영역(domain)에 해당하지 않으면서 무엇을 표현해야 하는지에 대한 언급 없이 프레임, 슬롯, 제한을 기술함으로서 개체(entity)를 표현하는 온톨로지
- 업무온톨로지 (method/task ontology)
  - 특정한 업무에 관한 정보를 제공하는 온톨로지

온톨로지와 웹 온톨로지의 비교 – 옥철영 (2004)

	Ontology	Web Ontology
정 의	공유된 개념화에 대한 형식적이고 명백하게 상세화하는 것	웹 문서에 나타난 지식을 표현하고, 공유하며, 재사용할 수 있도록 하는 온톨로지
활용분야	지식 또는 정보의 체계화와 응용화 등에 관련된 모든 분야	Semantic Web, E-Commerce, Agent, Web Service 등 웹 관련 분야
구축언어	특정한 온톨로지 구축 언어 없음	Web Ontology Language (RDF, DAML+OIL, OWL..., Topic Maps)
구축대상	단어, 어휘(개념), 전문용어...	웹의 정보를 공유할 수 있는 단어, 어휘, 전문용어 등
구축사례	WordNet, CYC, UMLS, EDR, Lexical FreeNet, ETRI CoNet...	실험적 수준 (기존 온톨로지의 변환)

주요 참고문헌

김홍기, 2002. 보이지 않는 공간의 혁명 : 시맨틱 웹 - 개념과 이해. ZDNet Korea, 2002.4.10.
<http://www03.zdnet.co.kr/techupdate/trend/0,39024505,10047655-8,00.htm>

옥철영, 2004. 한국어 정보처리와 온톨로지. 한국어정보처리연구회 동계튜토리얼, 성신여대, 2004. 2. 13.

이규호, 1972. 앎과 삶. 서울; 연세대학교출판부.

최중민, 2003. 시맨틱 웹의 개요와 연구동향. 한국어정보처리연구회 동계튜토리얼, 성신여대, 2004. 2.13.

Capurro, Rafael, 2002. Beiträge zu einer digitalen Ontologie
<http://www.capurro.de/digont.htm>

Sowa, John F. 2001. Building, Sharing, and Merging Ontologies.
<http://www.jfwowa.com/ontology/ontoshar.htm>

W3C, 2004. OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004.
<http://www.w3.org/TR/owl-features/>