



인 터 넷 진 화 의 열 쇠

온톨로지

웹 2.0에서 3.0으로

노상규·박진수 공저

god's toy business, 2007

SNU IDB laboratory

Ontology Contents

MODULE 1 온톨로지의 개념 및 응용

- Chapter 1 온톨로지 개요
 - 1. 온톨로지의 유래
 - 2. 분류와 개념화 과정
 - 3. 컴퓨터 온톨로지
- Chapter 2 온톨로지의 분류와 용도
 - 1. 온톨로지의 분류
 - 2. 온톨로지의 사용 목적과 중요성
 - 3. 온톨로지와 시맨틱 웹
- Chapter 3 온톨로지 구축 프로젝트
 - 1. 사이크(Cyc)
 - 2. 워드넷(WordNet)
 - 3. 전자거래문서
 - 4. 통합의학언어시스템
 - 5. 오픈 디렉터리 프로젝트
 - 6. 국제상품분류코드(UNSPSC)
- Chapter 4 온톨로지 적용 분야
 - 1. 전자상거래 분야
 - 2. 의료 분야
 - 3. 법률 분야
 - 4. 검색 서비스 분야
 - 5. 문화컨텐츠 분야

MODULE 2 온톨로지 언어와 구축도구

- Chapter 5 온톨로지 언어
 - 1. 온톨로지 언어의 발전 과정
 - 2. 인공지능 기반의 온톨로지 언어
 - 3. 온톨로지 마크업 언어
- Chapter 6 RDF(S): RDF와 RDF Schema
 - 1. XML과 RDF
 - 2. RDF
 - 3. RDF Schema
 - 4. RDF(S)의 한계점
- Chapter 7 OWL(Web Ontology Language)
 - 1. OWL의 기본 요소: 클래스와 속성
 - 2. OWL의 새로운 기능
 - 3. 세 종류의 OWL
 - 4. OWL 예제
- Chapter 8 토픽맵(Topic Maps)과 XTM(XML Topic Maps)
 - 1. 토픽맵(Topic Maps) 개념
 - 2. 토픽맵 구성요소
 - 3. XTM 예제
- Chapter 9 온톨로지 틀
 - 1. 온톨로지 틀의 분류
 - 2. 온톨로지 개발 틀
 - 3. 주요 온톨로지 틀 요약 정보

Chapter 2 온톨로지의 분류와 용도

■ 1. 온톨로지의 분류

- 1.1 형식성과 추론 메커니즘의 유무에 따른 분류
- 1.2 적용 범위에 따른 분류

■ 2. 온톨로지의 사용 목적과 중요성

- 2.1 시맨틱 상호운용성
- 2.2 표준화
- 2.3 커뮤니케이션
- 2.4 지식 관리 및 검색

■ 3. 온톨로지와 시맨틱 웹

1. 온톨로지의 분류

■ 1.1 형식성과 추론 메커니즘의 유무에 따른 분류

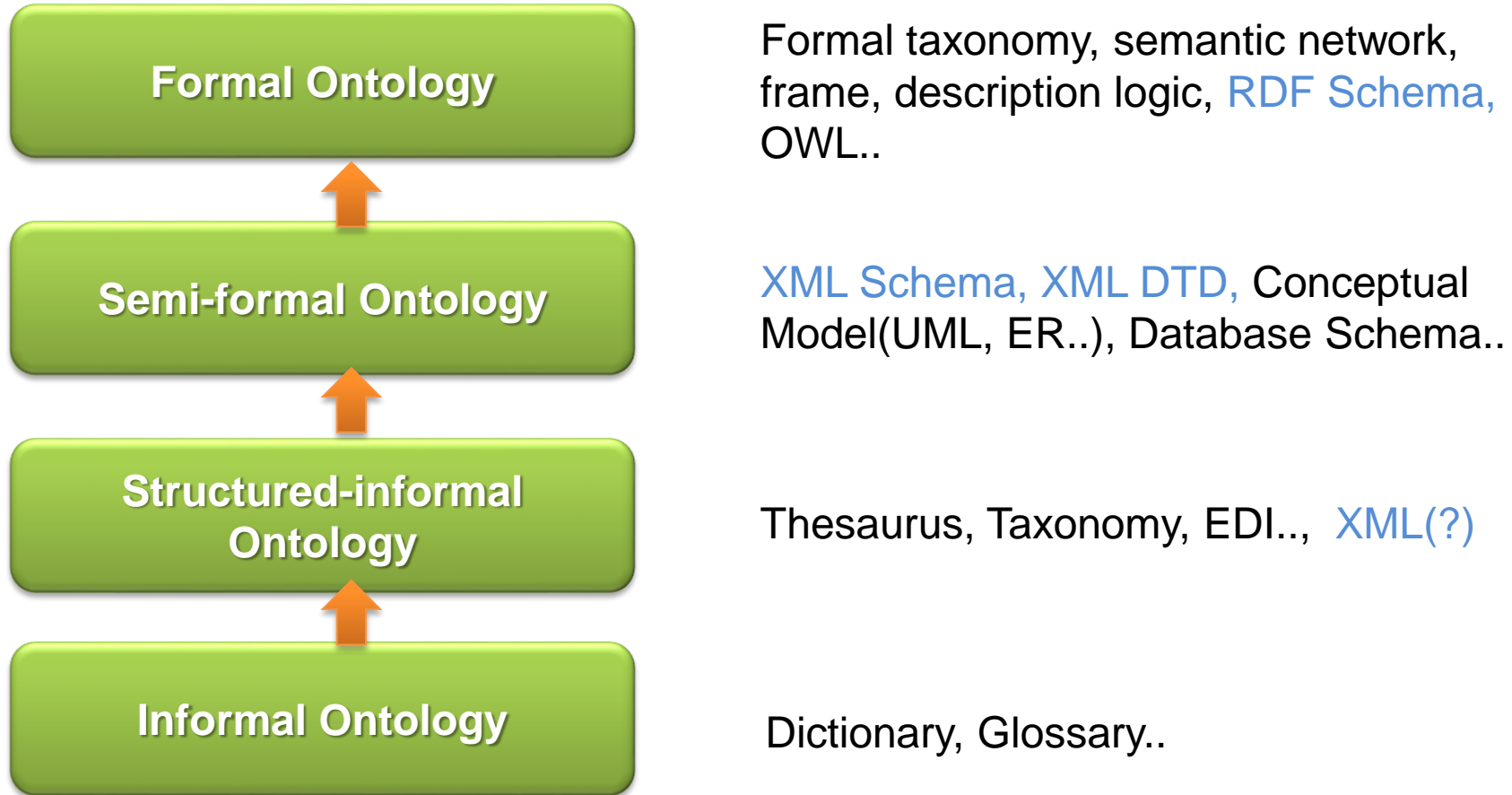
- 1.1.1 비형식적(Informal) 온톨로지
- 1.1.2 구조화된 비형식적(Structured-informal) 온톨로지
- 1.1.3 반형식적(Semi-formal) 온톨로지
- 1.1.4 형식적(Formal) 온톨로지

■ 1.2 적용 범위에 따른 분류

- 1.2.1 상위 온톨로지
- 1.2.2 하위 온톨로지

1.1 형식성과 추론 메커니즘의 유무에 따른 분류

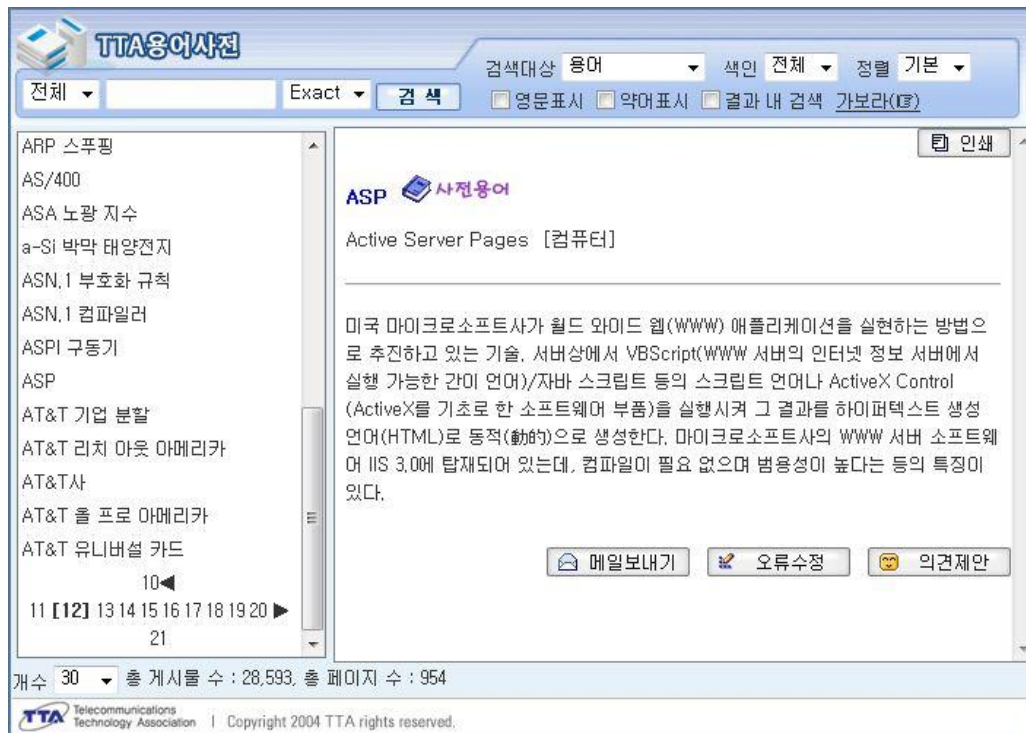
Formality 정도에 따른 Ontology 분류



1.1.1 비형식적(Informal) 온톨로지

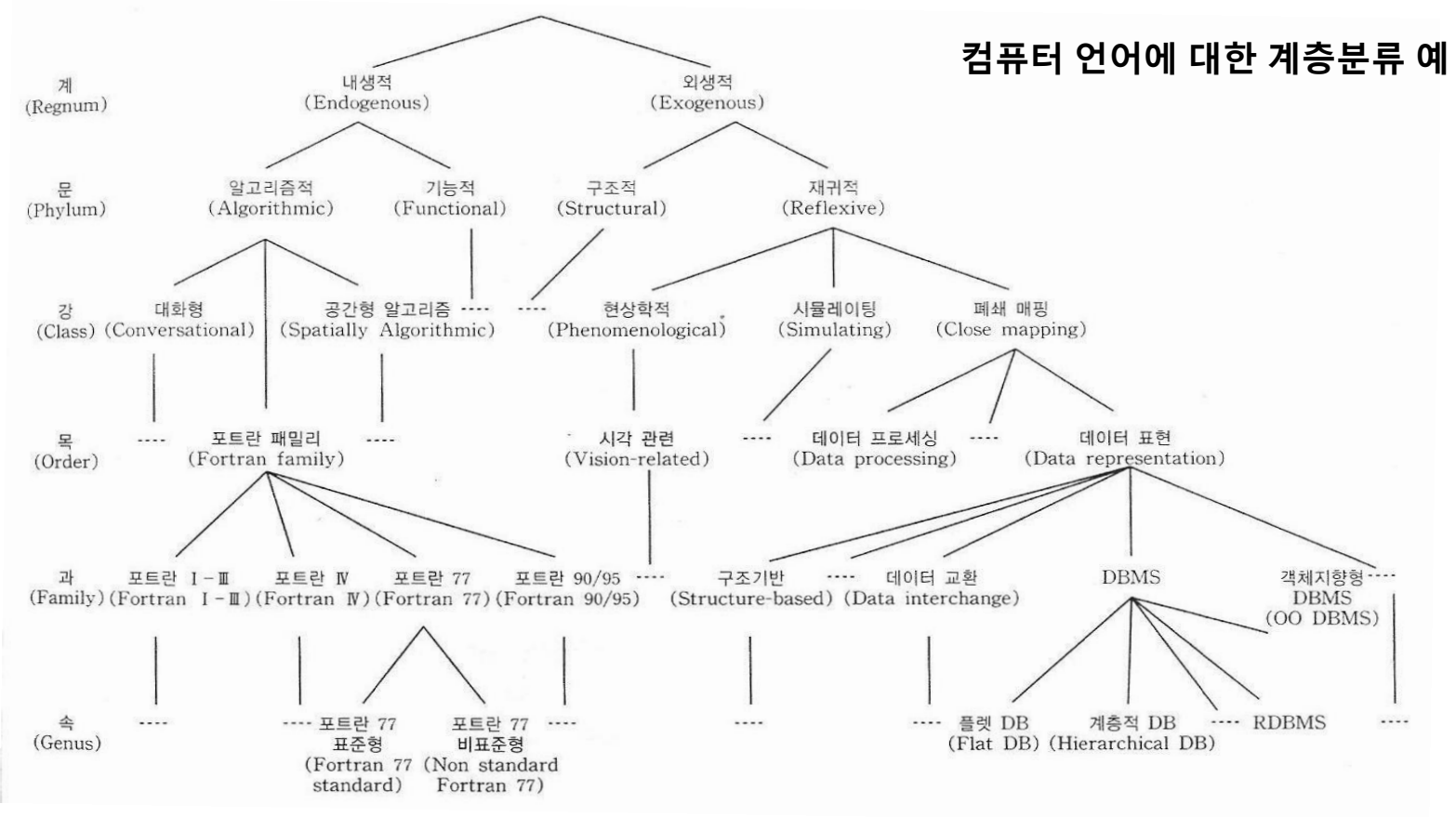
- 자연어로 자유롭게 표현, 최소한의 형식적 구조
- 인간이 이해하기 쉬우나 S/W Agent가 이해하고 처리하기 어려움
- 예) 사전(Dictionary), 용어사전(Glossary)

정보통신 용어사전 예 (<http://word.tta.or.kr/terms/terms.jsp>)



1.1.2 구조화된 비형식적(Structured-informal) 온톨로지

- 비형식적 온톨로지보다는 체계적인 구조를 갖춘 형태로 표현
- 대부분 자연어로 표현, 그러나 개념들 간의 관계를 구조적으로 연결
- 예) Thesaurus, Taxonomy, EDI(Electronic Data Interchange) Format, XML(?)



1.1.3 반형식적(Semi-formal) 온톨로지

- 컴퓨터가 이해할 수 있는 formal language로 표현한 온톨로지
- 엄격한 formal semantic을 제공하지는 않음
 - 그러나 명확한 formal syntax를 제공해 온톨로지 구축시 문법적 오류 방지
- 예) Database Schema, Conceptual Model(XML Schema, XML DTD, UML, ER model)

XML Schema 예

```
01 <?xml version="1.0" encoding="euc-kr" ?>
02 <schema smlns="http://www.w3.org/2001/XMLSchema">
03 <element name="노트북">
04     <complexType>
05         <sequence>
06             <element name="모델명" type="string" />
07             <element name="CPU" type="string" />
08             <element name="메모리" type="positiveInteger" />
09             <element name="DMB지원" type="boolean" />
10             <element name="가격" type="positiveInteger" />
11         </sequence>
12     </complexType>
13 </element>
14 </schema>
```


1.1.4 형식적(Formal) 온톨로지

- 주로 논리에 기반한 Formal Language로 표현한 온톨로지
- Formal Semantic을 적용
 - 표현력(expressivity)이 풍부해짐
 - 추론 규칙(inference rule)에 대한 정의 포함 가능
- 예) Formal Taxonomy, Semantic Network, Frame, Description Logic, PDF Schema, OWL..

OWL 예

```
01 <owl : Class red : ID="컬러 레이저 프린터">
02     <rdfs : subClassOf>
03         <owl : Class rdf : ID="컬러프린터" />
04     </rdfs : subClassOf>
05     <rdfs : subClassOf rdf : resource="#레이저 프린터" />
06 </owl : Class>
07 <owl : ObjectProperty rdf : ID="#대체하다">
08     <rdfs : domain rdf : resource="#컬러 레이저 프린터" />
09     <rdfs : range rdf : resource="#컬러 레이저 프린터" />
10     <rdfs : type rdf : resource="#http://www.w3.org/2002/07/owl#SymmetricProperty" />
11 </owl : ObjectProperty>
```

1.2 적용 범위에 따른 분류



1.2.1 Upper-level Ontology

1. 일반상식(Commonsense) 온톨로지

- 사물, 이벤트, 시간 같이 여러 분야에서 공통적으로 사용되는 개념들을 Generalization시킨 것
 - Bunge Ontology: 일반적 시스템을 모델링
 - OpenCyc Upper Ontology: 일반 상식을 담고 있는 지식베이스
 - ▶ 1985년, Doug Lenat, MCC AI project
 - ▶ 2001년 OpenCyc 공개
 - Sowa's Upper Ontology: 철학, 언어학, 논리학, AI 등에서의 개념
 - ▶ 1995년, John Sowa
 - SUMO: 컴퓨터 정보 처리 시스템의 기본이 되는 상위 온톨로지
 - ▶ Suggested Upper Merged Ontology, Teknowledge Corporation, Ian Niles & Adam Pease
 - ▶ Sowa Ontology + Russell&Norvig Upper Ontology

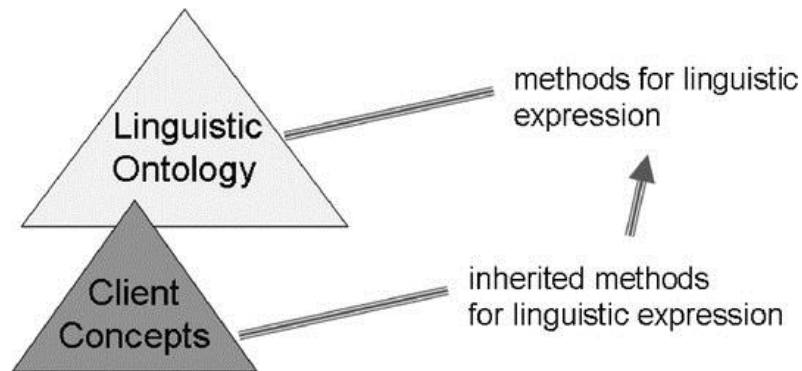
1.2.1 Upper-level Ontology: Release 4.0 of OpenCyc

- The core Cyc ontology whose domain is all of human consensus reality. The current release includes:
 - ~239,000 terms (up from ~177,000 terms in the previous release)
 - ~2,093,000 triples (up from ~1,500,000 in the previous release)
 - Select class (direct and indirect) instance counts:
 - ▶ ['place'](#): ~19,000
 - ▶ ['organization'](#): ~26,000
 - ▶ ['predicate'](#): ~22,000
 - ▶ ['business related thing'](#): ~28,000
 - ▶ ['person'](#): ~12,700
 - ~69,000 owl:sameAs links to external (non-Cyc) semantic data namespaces:
 - ▶ DBpedia: ~47,000 links, *including 696 links to the DBpedia ontology*
 - ▶ UMBEL: ~21,000 links
 - ▶ WordNet: ~11,000 links
 - ▶ Wikicompany: 1028 links
 - ▶ CIA World Factbook: 172 links
 - ▶ RDFAbout SEC company identifiers: 661 links
 - ▶ RDFAbout states and counties: 71 links
 - ▶ FOAF: 44 links
 - English strings (a canonical one and alternatives) corresponding to each concept term, to assist with search and display.

1.2.1 Upper-level Ontology

2. 언어(Linguistic) 온톨로지

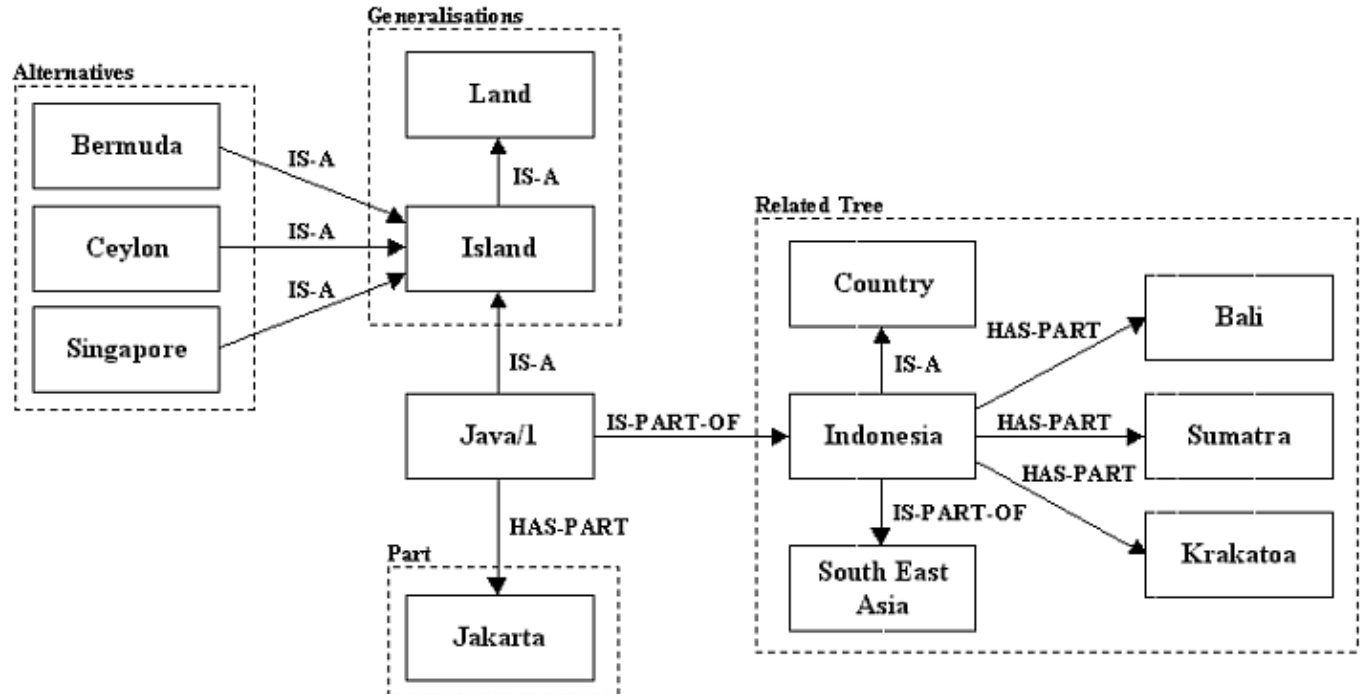
- 애플리케이션이나 특정영역에 상관없이 **자연어 온톨로지**를 이용할 수 있도록 지원하는 역할
- GUM(Generalized Upper Model), WordNet, Penman Upper Model ..



Traditional relation between the generalized upper model and domain ontologies

1.2.1 Upper-level Ontology: WordNet

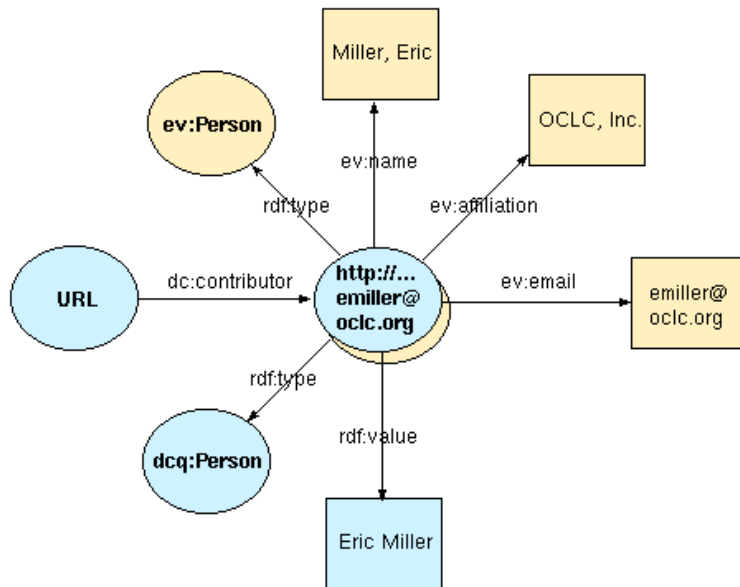
- **WordNet** was created in the Cognitive Science Laboratory of Princeton University under the direction of psychology professor George Armitage Miller starting in 1985
- As of November 2012 WordNet's latest Online-version is 3.1.
- The database contains 155,287 words organized in 117,659 synsets for a total of 206,941 word-sense pairs; in compressed form, it is about 12 megabytes in size
- WordNet includes the lexical categories nouns, verbs, adjectives and adverbs but ignores prepositions, determiners and other function words.



1.2.1 Upper-level Ontology

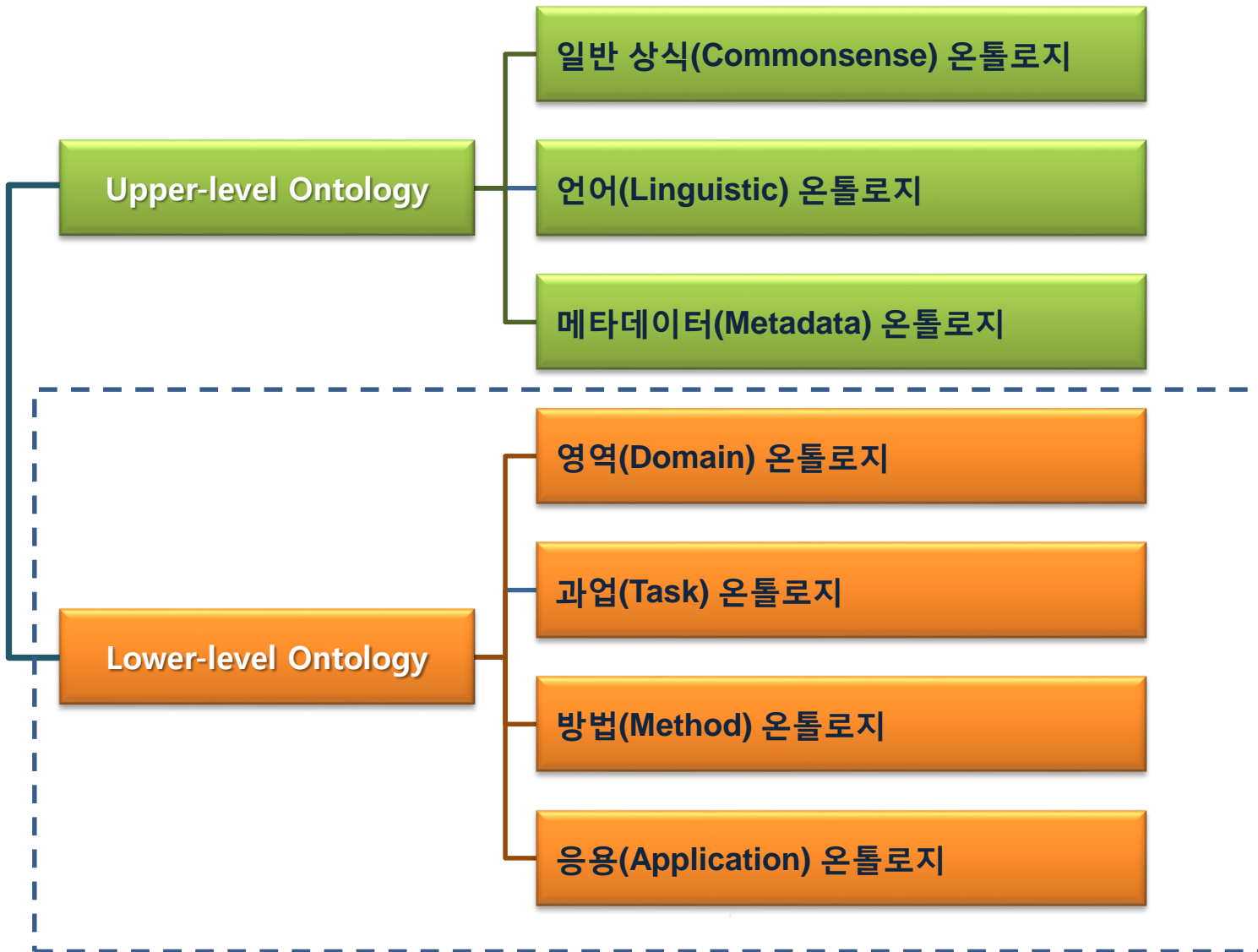
3. 메타데이터(Metadata) 온톨로지

- 온라인 및 오프라인 데이터 간의 Interoperability를 제공하기 위한 것
- 용이한 정보검색을 위해 메타데이터에서 사용하는 어휘 표준화



Dublin Core by 미국국립정보표준원 NISO

1.2 적용 범위에 따른 분류



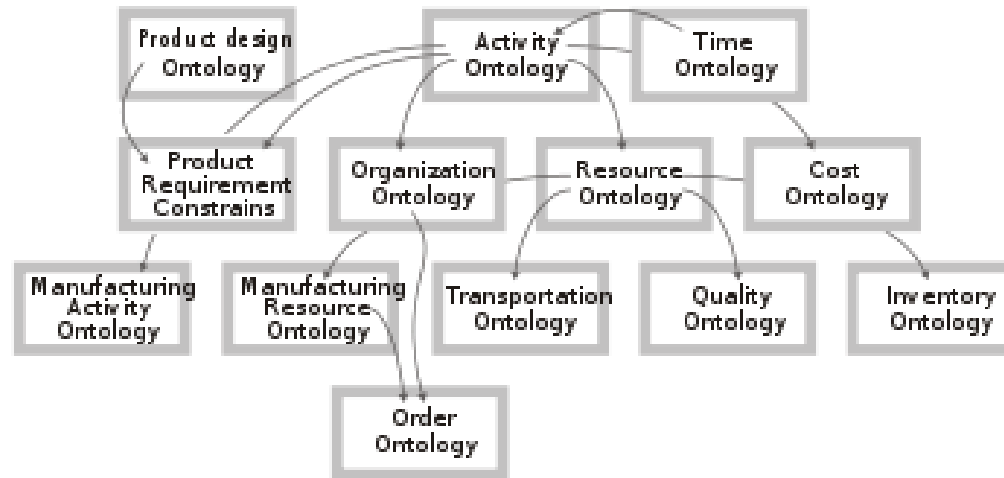
1.2.1 Lower-level Ontology

** 상위 온톨로지에 비해 구체적인 내용을 다룸

1. 영역(Domain) 온톨로지

■ 특정 도메인에서 사용하는 지식을 개념화한 온톨로지.

- 해당 도메인에서만 재사용 가능
- 예) Enterprise Ontology – 비즈니스 분야

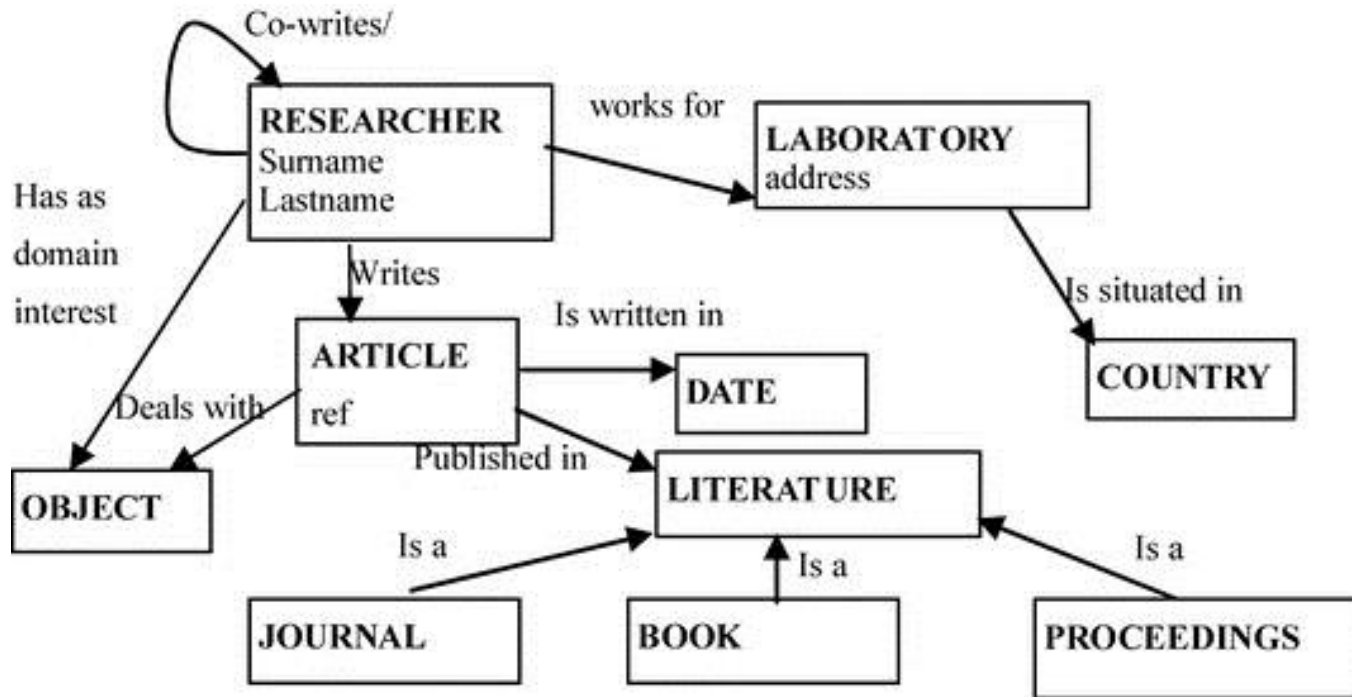


TOVE(Toronto Virtual Enterprise)

1.2.1 Lower-level Ontology

2. 과업(Task) 온톨로지

- 일반적 문제 혹은 특정 도메인에서 발생하는 문제 해결의 어휘(명사, 동사, 형용사 등)를 제공

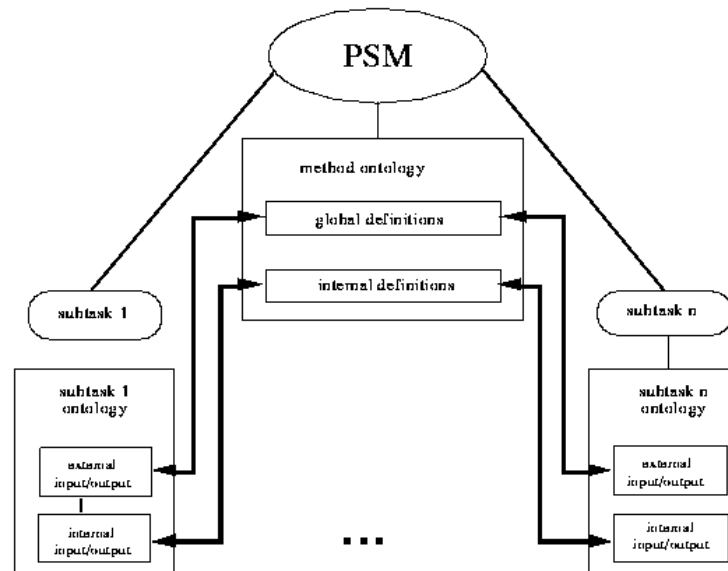


Science monitoring task ontology

1.2.1 Lower-level Ontology

3. 방법(Method) 온톨로지

- 해당 업무를 성공적으로 마무리할 수 있도록 돕는 데 사용
- '추론과정을 지원할 때 적용될 수 있는 특정업무와 관련된 개념'과 '그 개념들 간의 관계에 대한 정의'
- 영역(Domain) 온톨로지와 비슷하나, 일을 처리하는 **방법**을 기준으로 온톨로지를 구축한 것 (Chandrasekaran et al., 1998)

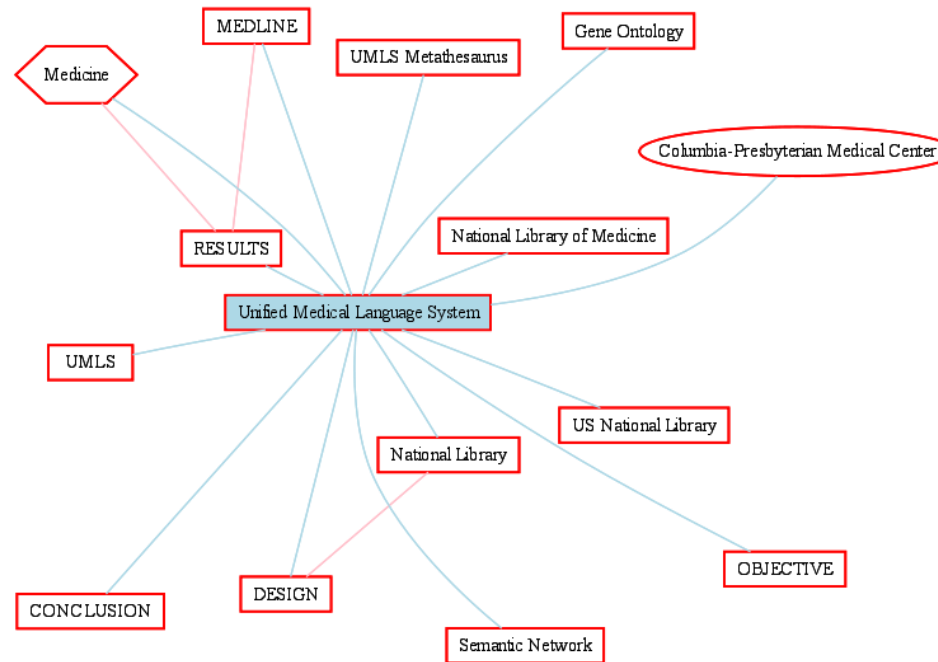


The method ontology and the related subtask ontologies

1.2.1 Lower-level Ontology

4. 응용(Application) 온톨로지

- 해당 Application에서 다루는 지식을 모델링하는데 필요한 개념 모두 포함
- 해당 Application에서의 활용도 ↑, 다른 Application에서의 재사용 ↓
 - 예) 통합의학언어시스템



Copyright 2009, Research Foundation, Stony Brook

UMLS(Unified Medical Language System)

Chapter 2 온톨로지의 분류와 용도

■ 1. 온톨로지의 분류

- 1.1 형식성과 추론 메커니즘의 유무에 따른 분류
- 1.2 적용 범위에 따른 분류

■ 2. 온톨로지의 사용 목적과 중요성

- 2.1 시맨틱 상호운용성
- 2.2 표준화
- 2.3 커뮤니케이션
- 2.4 지식 관리 및 검색

■ 3. 온톨로지와 시맨틱 웹


2. 온톨로지의 사용 목적과 중요성

■ 온톨로지의 중요성

- 특정영역의 지식을 모델링하고 처리하여 구성원 간의 지식 공유 및 재사용을 가능
- 다양한 용도로 사용
 - ▶ 지식표현, 지식베이스, 지식검색, 지식관리,
 - ▶ 정보시스템 개발, 비즈니스 프로세스 모델링, 표준화,
 - ▶ 기업 정보시스템 통합, 정보시스템 평가, 시맨틱 웹 구축 등

■ 온톨로지의 사용 목적

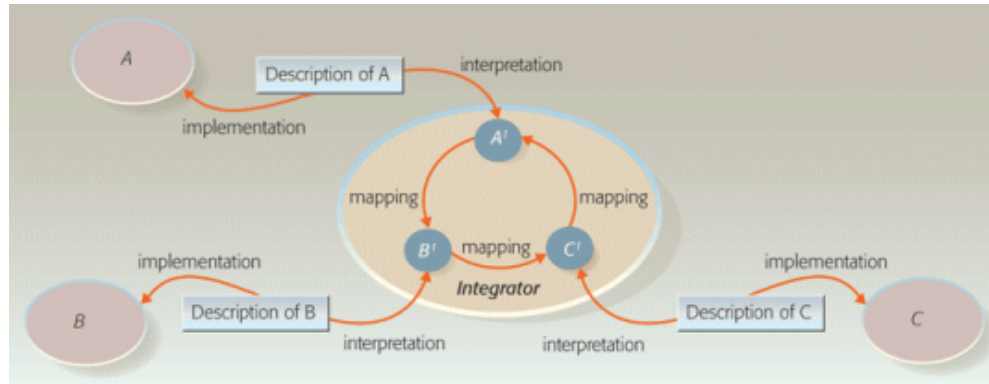
- (2.1) 시맨틱 상호운용성
- (2.2) 표준화
- (2.3) 커뮤니케이션
- (2.4) 지식관리 및 검색



지식의 공유와 재사용의 관점에서 볼 때
목적은 서로 밀접한 관련성을 가짐

2.1 시맨틱 상호운용성(Semantic Interoperability)

- 여러 정보시스템들이 각각의 고유한 자율성과 다양성을 유지하면서 마치 하나의 시스템처럼 운용되는 것



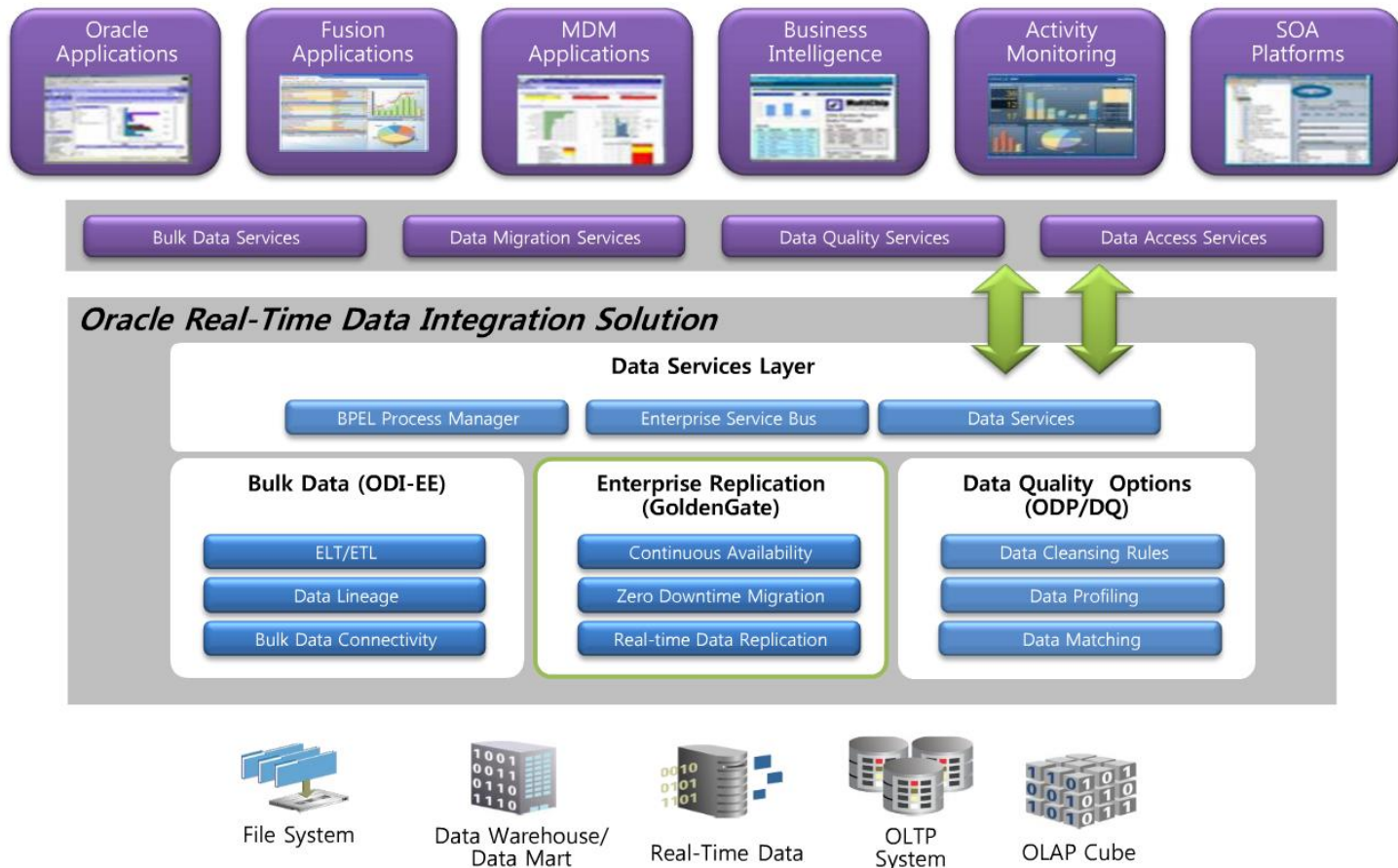
Any-to-Any Centralized Semantic Interoperability Model

- 온라인 시장 진출, 합병 등을 하는 기업의 증가
 - 이종 시스템간의 통합, Interoperability 문제가 발생하고 있음
 - 기업의 정보가 대부분 호환되지 않는 다양한 형태로 기존시스템에 있기 때문
 - 예) → 서로 다른 기업/부서가 동일 제품을 다른 제품명/제품번호로 사용
 - 도량단위 표준(킬로그램, 온스, 미터, 마일, 인치 등) 사용의 불일치
 - 각 시스템 별 정보형식/포맷/내용이 달라 정보 공유/교환시 문제발생
- 시스템 간 상호운용성은 다양한 정보시스템에 저장되어 있는 정보에 접근하려는 기업들에게는 매우 중요한 이슈

2.1 시맨틱 상호운용성

■ Forrest Research 보고

- 98%의 회사가 IT전략 중 시스템 통합이 '매우 중요'하다고 함
- 예 : RTDI(Real Time Data Integration 이종 시스템 간 실시간 정보 공유)



2.1 시맨틱 상호운용성

■ 이종 시스템간 통합시스템 구축

- 개별시스템의 독립성과 자율성을 보장
- 개별 시스템 간의 데이터 비호환성 (Semantic Conflict)을 실시간 탐지/해결
- 그러나 정보 공유를 위해선 대부분 기존 시스템의 수정이 필요

■ 바람직한 시스템 환경 인프라

- 주요한 통합구조 변경 없이 개별 시스템이 쉽게 추가/제거되는 유연성 필요

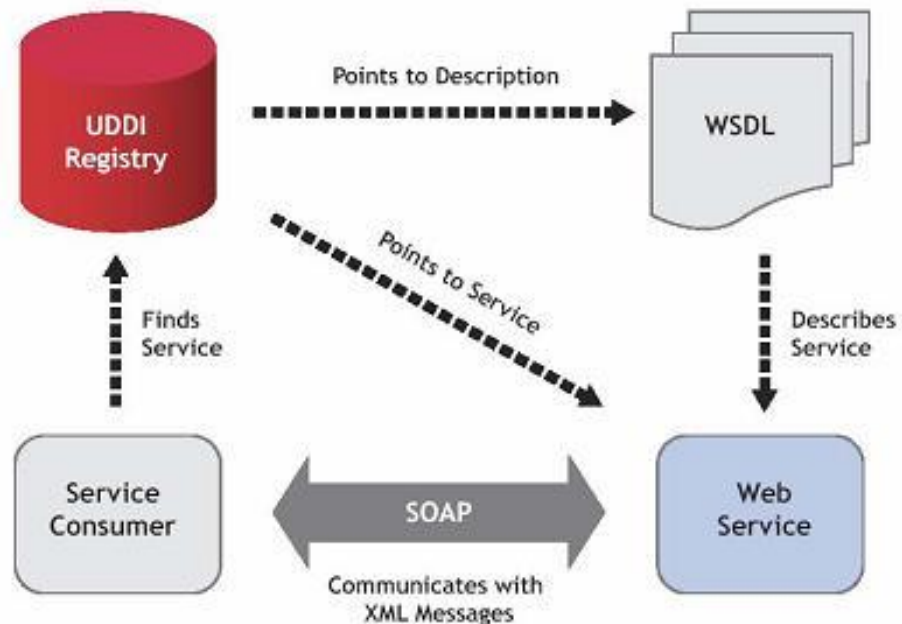
■ 2 Resolutions

- Syntactic Interoperability
- Semantic Interoperability

2.1 시맨틱 상호운용성 : 2 Resolutions

■ 1. Syntatic Interoperability

- 개별 SW 컴포넌트간에 message passing으로 시스템 상호운용성을 제공
 - ▶ 표준화된 프로토콜을 사용
 - ▶ 예) XML 기반의 웹 서비스
 - WSDL(Web Service Description Language)
 - SOAP(Simple Object Access Protocol)



2.1 시맨틱 상호운용성: 2 Resolutions

■ 2. Semantic Interoperability

- 정보의 암시적 의미 (Implicit Knowledge)나 내재하는 규칙(Rules)을 이용
- 이종 시스템 간에 정보의 의미까지도 공유할 수 있는 능력
 - ▶ 예) 대학학점 평점 3.8 (Univ Minnesota - 4.0만점, SNU - 4.3만점)
→ 같은 3.8이지만 의미는 다르다
- 여러 정보 시스템에서 정보검색이나 시스템통합시에 항상 직면하는 문제

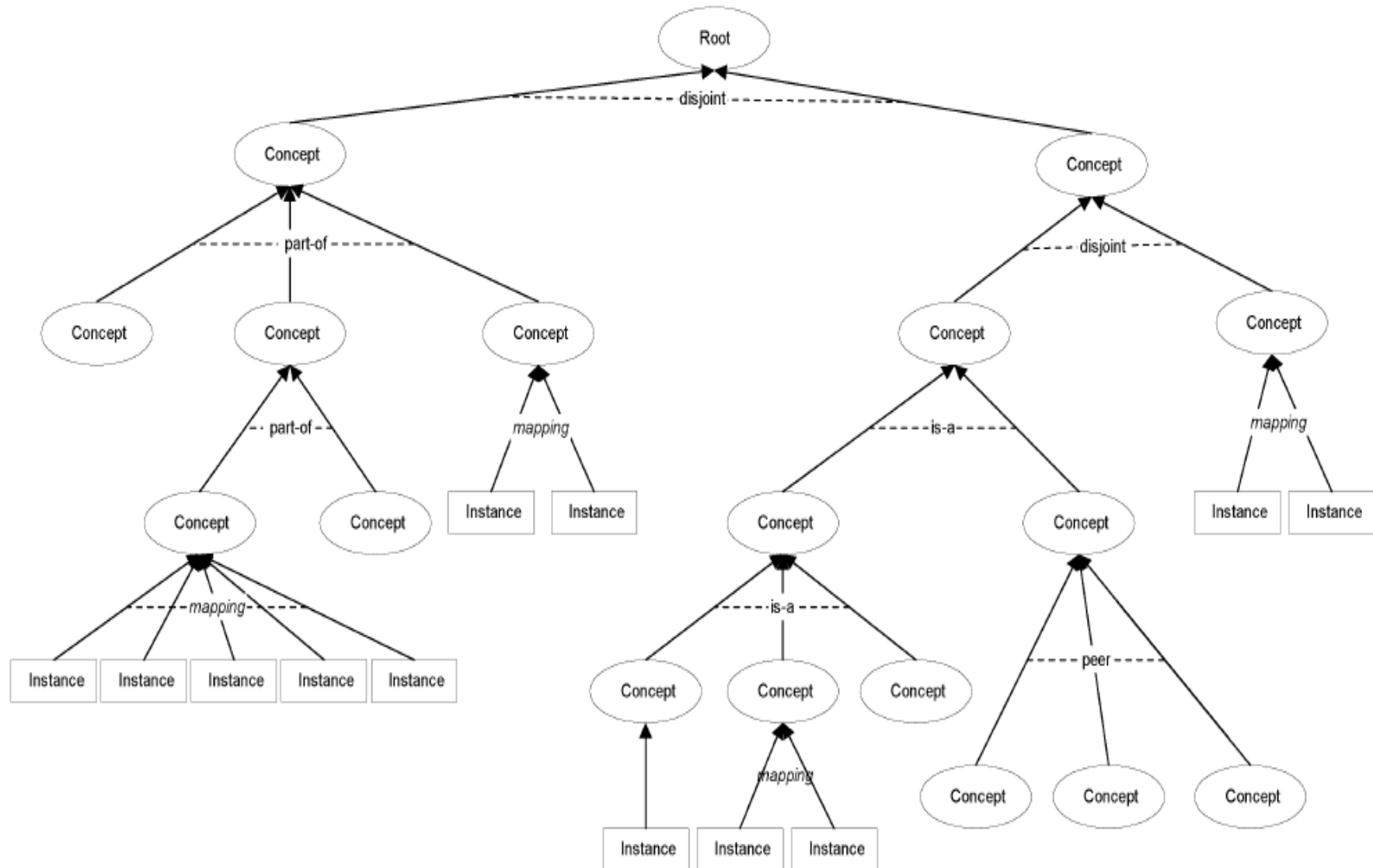
■ 문제 해결은? → **Ontology!**

■ 관련 집단 구성원들의 합의로 도출된 개념화

- 데이터의 의미를 기술할 수 있음
- 다양한 형태의 의미 충돌을 해결 가능

■ 이종시스템들의 상호이해와 Semantic reconciliation을 위한 기반

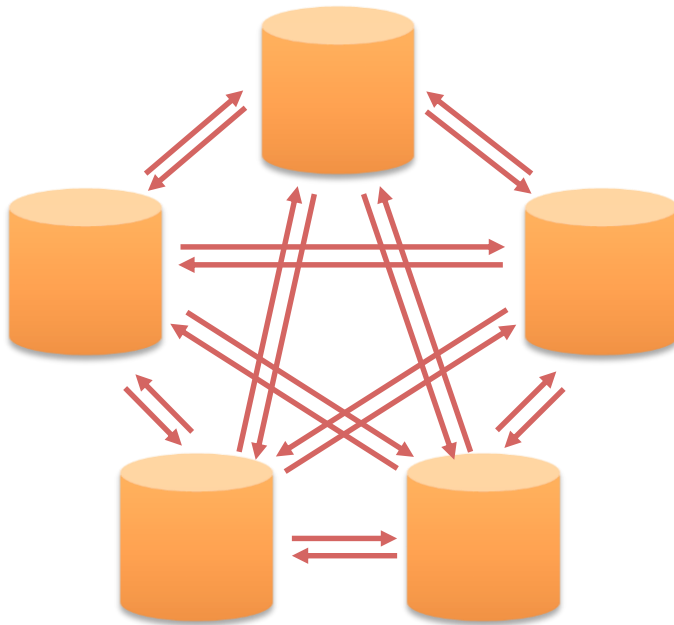
2.1 시맨틱 상호운용성: By Ontology



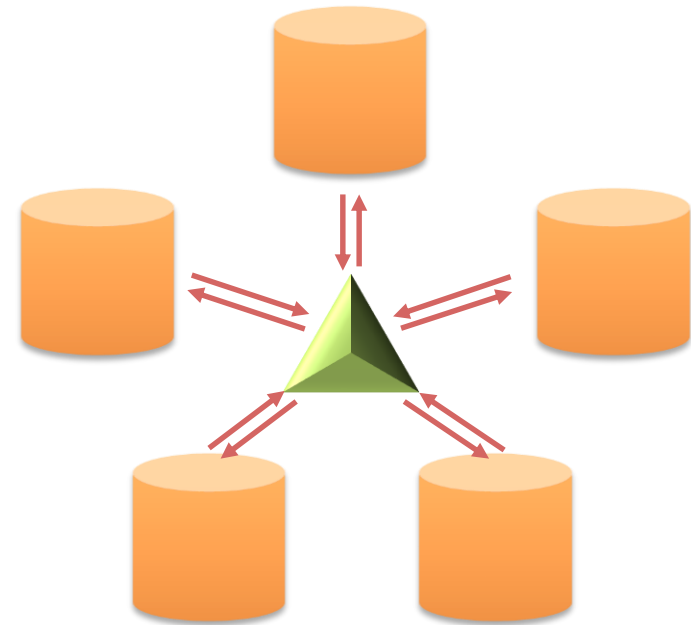
SCROL(Semantic Conflict Resolution Ontology)

2.1 시맨틱 상호운용성: By Ontology

시스템 간의 데이터 변환이 필요한 경우의 수



(a) 온톨로지없이 시스템간 직접 정보공유



(b) 온톨로지를 매개로 한 시스템 간의 정보공유

2.2 표준화

■ 표준화된 온톨로지

- 특정영역에 개념을 표현하는 단어들과 그 관계들을 계층적 구조로 표현
- 구성원 모두가 사회적 합의하에 구축
- 응용 프로그램 사이의 정보/지식 공유 수월화
- 특정영역의 지식을 문서화하고 재사용을 가능하게 함

■ 전자상거래에서 글로벌 인프라를 위한 표준화된 온톨로지

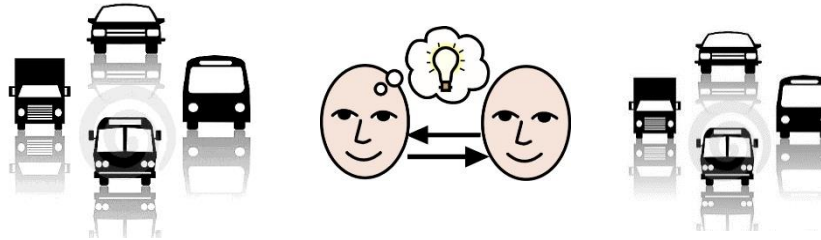
- 예) 전자상거래 개방형 표준인 ebXML, 로제타넷

■ 표준화는 Semantic Interoperability와도 깊게 연관

- EMR(진료정보를 전자차트에 담는 전자의무기록) 문제
 - ▶ 의사들의 진료기록/처방내역이 각기 달라 통일하기 어려움
- 온톨로지로 이런 진료관련 용어를 표준화하여 환자정보를 쉽게 공유가능

2.3 커뮤니케이션

- 온톨로지는 개념적/용어적 혼돈이 가능한 부분들을 단일화된 구조로 명시
 - 서로 구성원들 사이에 이해와 커뮤니케이션을 촉진
 - 합의를 통해 하나의 통일된 용어를 선택하고, 그 단어의 개념을 명확히 정의
 - ▶ 예) '자동차'에 '트럭'과 '버스'도 포함하는 것인지 아닌지?
 - ▶ '자동차'의 범위를 규정해 의미를 정확히 정의



- Application Program 혹은 S/W Agent들이 특정 분야 개념의 의미를 정확히 이해
 - SW간 커뮤니케이션을 가능하게 함
 - 예) 전자상거래용 S/W Agent 기반 구매협상 시스템(Automated Negotiation System)을 구축
 - ▶ 협상을 위한 커뮤니케이션 어휘
 - ▶ 협상 프로토콜 관점에서 각 어휘의 의미를 제공
 - ▶ 예) 가격협상 시 '가격' 용어의 의미에 따라 협상 내용 영향 (화폐단위, 가격이 포함하고 있는 내용)

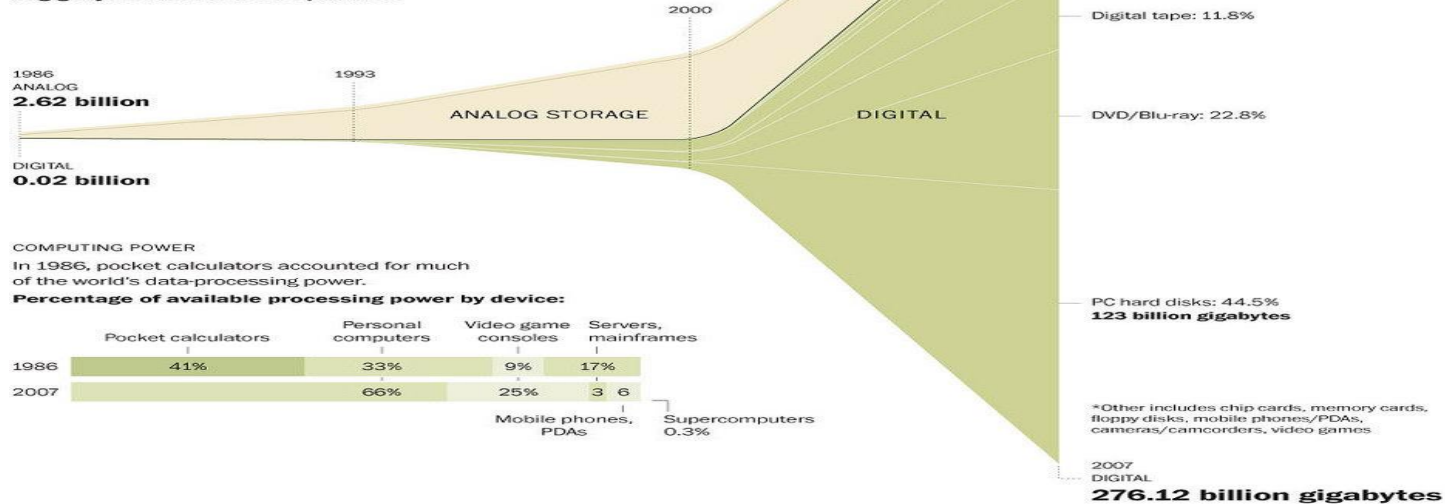
2.4 지식 관리 및 검색

정보생산량 폭발적 증가

THE WORLD'S CAPACITY TO STORE INFORMATION

This chart shows the world's growth in storage capacity for both analog data (books, newspapers, videotapes, etc.) and digital (CDs, DVDs, computer hard drives, smartphone drives, etc.)

In gigabytes or estimated equivalent



2.4 지식 관리 및 검색

- 짧은 시간 내에 필요한 지식을 찾는 것이 중요한 문제
 - 관련 정보를 어떻게 효율적으로 처리하느냐가 아닌,
 - 어떤 정보가 관련이 있고, 그 정보가 어디에 있는지를 정확하게 찾아내는 것
- 온톨로지
 - 광대한 정보 공간에서 지식관리 및 검색능력을 향상시킴
 - 온톨로지 기반 지식 검색시 특정용어와 관련된 다른 지식의 검색도 가능
 - 예) '오사마 빈 라덴' 검색 ➔ 알카에다, 탈레반, 관련 인물들의 검색도 가능
- 온톨로지 기반 검색
 - 단순 Keyword Matching 보다 지능적인 시맨틱 기반 검색과 Filtering가능
 - 각 개인이 사용하는 용어의 차이에 관계없이 정확히 필요한 정보 검색 가능
 - 동일한 검색어지만 다른 의미를 지닌 관련 없는 정보의 검색을 피할 수 있다

소금 검색! Salt? or Small Flute?



Chapter 2 온톨로지의 분류와 용도

- 1. 온톨로지의 분류
 - 1.1 형식성과 추론 메커니즘의 유무에 따른 분류
 - 1.2 적용 범위에 따른 분류
- 2. 온톨로지의 사용 목적과 중요성
 - 2.1 시맨틱 상호운용성
 - 2.2 표준화
 - 2.3 커뮤니케이션
 - 2.4 지식 관리 및 검색
- 3. 온톨로지와 시맨틱 웹

3. 온톨로지와 시맨틱 웹

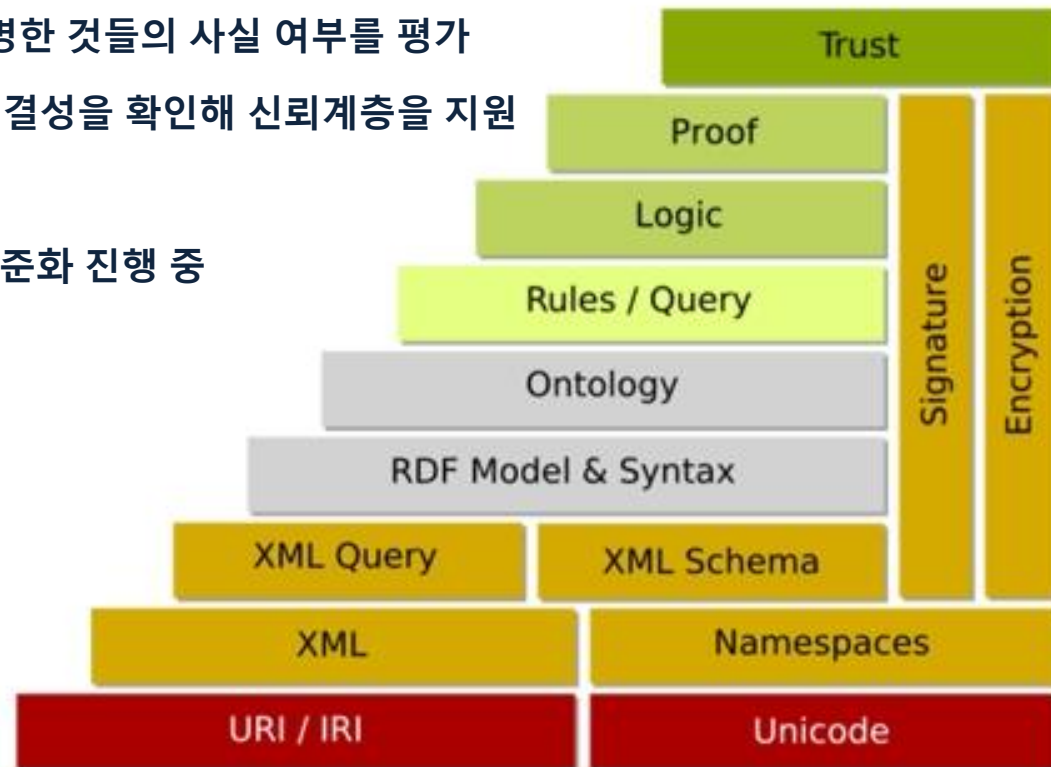
- 시맨틱 웹은 기존 웹(WWW)에 온톨로지를 확장된 개념
 - 웹 정보에 잘 정의된 의미를 추가
 - Tim Berners-Lee의 비전에서 출발 (Web of Data)

현재의 웹	시맨틱 웹
HTML 기반	XML, RDF, RDF Schema에 기반
인간의 노동력으로 움직이는 웹	기계의 노동력으로 움직이는 웹
인간만 이해 가능 컴퓨터가 자동으로 이해, 처리, 전달 어려움	인간&기계 모두 이해하고 해석 가능한 형태로 표현

- 다양한 기업, 커뮤니티. 여러 응용 프로그램을 지원하는 정보 중개자
 - 정보를 공유하고 재사용
- 여러 표준과 관련기술로 웹 상에 있는 데이터를 표현하는 Abstract Model
- 전세계의 컴퓨터 간의 Semantic Interoperability를 용이하게 할 것임

3. 온톨로지와 시맨틱 웹: 시맨틱 웹의 계층구조

- 웹 데이터를 표현하는 가장 기본 계층
 - Unicode, URI 및 XML 기반
- 온톨로지 계층 - RDF와 RDF Schema , OWL로 구성
- Logic 계층 - 논리적 규칙(rule)들로 구성
 - S/W Agent가 일을 처리하는 데 필요한 논리적 추론을 가능하게 함
- Proof 계층 - 논리적 규칙들이 올바르게 실행되었는지 증명하는 역할 담당
- Trust 계층 - 가장 상위계층, 논증계층에서 증명한 것들의 사실 여부를 평가
- 전자서명 계층 - 여러 계층에 걸쳐 웹문서의 무결성을 확인해 신뢰계층을 지원
- Logic 계층, Proof 계층, Trust 계층은 아직 표준화 진행 중



3. 온톨로지와 시맨틱 웹

■ 시맨틱 웹 성공의 조건

- 쉬운 접근성이 인증된 온톨로지와 다양한 콘텐츠의 확보에 의해 결정
- 온톨로지, RDF 기반 웹 콘텐츠 제작을 위한 쉬운 툴의 등장

■ 시맨틱 웹의 사용

- 웹 서비스, 지식관리, 지능형 검색,
- e-마켓플레이스, 전자상거래, 유비쿼터스 컴퓨팅,
- 자연어 처리, 가상 커뮤니티, 지능형 정보통합, 실시간 정보 공유

■ Ontology-based Semantic Web

- 다양한 추론이 가능한 광범위한 지식 기반 시스템 네트워크 구성