



시맨틱 웹에서의 온톨로지 공학[†]

가톨릭대학교 양정진

1. 서론

온톨로지는 초기에 지식의 공유와 재사용을 목적으로 한 연구분야 특히 인공지능 분야에서 개발되었다. 이는 최근에 정보통합, 정보시스템, 정보검색, 전자상거래, 지식관리 등의 연구분야에서 정보를 능동적이고 연계적으로 추출하기 위하여 웹 정보에 관한 사람과 응용시스템 사이의 공유된 지식과 공통된 해석의 수단이라는 차원에서 더욱 관심을 받고 있다. 본 논문에서는 시맨틱 웹 기술 실현에 있어서의 온톨로지 공학을 다룬다.

먼저, 웹 상에 있는 온톨로지가 주축을 이루는 시맨틱 웹에 대하여 논하여 보자. 웹 페이지를 인터페이스로 하여 자연언어로 표현되어 있는 현재 웹에 존재하는 문서와 서비스는 사람의 지능을 통해 해석되어지고 사용되도록 되어 있다. 이는 역으로 웹 상에 존재하는 정보의 양이 방대해지면서 웹의 효율적인 사용을 저하시키고, 웹 서비스들 간의 부족한 상호운용성은 웹 성장의 장애적 요소로 드러난다. 이 문제의 해결책으로서 시맨틱 웹은 사람뿐만 아니라 컴퓨터도 웹 상의 내용을 이해하여 처리할 수 있도록 형식화된 지식과 데이터로 현재의 웹 형태를 강화하며 이로 인하여 웹 내용처리의 자동화를 꾀하고 웹의 효율적 자원공유를 높이려고 한다. 아직까지 이러한 형식화에 있어서 일치된 의견은 없지만, 그 범위는 더블린 코어와 같은 메타정보 스키마에서 논리적인 표현언어에 이르기까지 다양하고 넓다고 할 수 있다.

시맨틱 웹 기술의 구성은 크게 RDF와 같은 단일화된 데이터모델, 그 위에 규정 되어지는 DAML+OIL(OWL)과 같은 의미를 표현할 수 있는 언어, 웹

자원들을 나타내기 위한 표준화된 용어 규정의 온톨로지, 그리고 그러한 의미적인 것들의 생성과 처리를 지원하는 툴들로 이루어져 있다.

시맨틱 웹 기술을 개발, 적용함에 있어서 구성요소인 온톨로지의 역할을 나열해 보면

- 객체의 주체성(동일성) 관련정보를 온톨로지에 명시하여 시맨틱 웹이 자원을 식별하고 그 자원의 위치를 측정하는 추론을 가능케 함
- 시맨틱 모델들 간의 관련성을 파악하고 이질성을 줄이는 데 사용
- 시맨틱 웹 채택이 용이하도록 기초적인 온톨로지 개발방법, 잘 정돈된 온톨로지 라이브러리 제공, 온톨로지와 메타데이터 학습 가능성 등을 들 수 있다.

이러한 온톨로지의 역할 수행에 대하여 본 논문에서는 배경지식, 온톨로지 종류와 애플리케이션, 온톨로지 공학, 온톨로지 개발도구, 온톨로지 공학의 연구문제, 그리고 시맨틱 웹과 온톨로지 장래성의 순으로 다룬다.

2. 배경 지식

2.1 온톨로지

우리가 흔히 명칭하는 것들을 그룹이나 종류로 분류하여 놓은 것을 분류법(taxonomy)이라고 한다면, 온톨로지(ontology)는 개념(concept)들의 집합들을 나타내는 것으로 개념간에 존재하는 관계를 명백하게 기술하여 개념과 그러한 관계성립의 공유를 목적으로 사용된다. 온톨로지의 다양한 정의들 중에서 가장 본질적인 정의를 들어 온톨로지를 설명하자면 "온톨로지는 공유하기(shared) 위한 개념들의 개념화(conceptualization)를 형식적(formal)이고, 명백하

[†] 이 논문은 한국과학재단 지역대학 우수과학자 연구과제의 지원을 받음을 알립니다.

계(explicit) 설명해 놓은 명세서(specification)" [1]라고 할 수 있다. 여기서 개념화(conceptualization)는 어떤 현상에 대해서 관련 있는 개념들을 식별하여 그 현상을 추상화한 모델로 설명하는 것이다. 명백하게(explicit) 표현한다는 것은 사용된 개념의 종류와 개념들이 추상화된 모델과 갖는 관계들, 그리고 그 개념들의 사용에 있어서 주어지는 제한점(constraints)들을 명백하게 정의한다는 것을 의미한다. 예를 들어, 의학분야에서 병이나 증상을 개념으로 볼 수 있고, 그들의 관계가 서로 인과관계에 있다면, 그 개념들의 사용 제한점은 병이 병 자체를 유발할 수 없는 것이 된다는 것이다. 형식적(formal)인 표현은 온톨로지가 기계 판독 가능한 형태이어야 한다는 것을 의미한다. 끝으로 공유된(shared) 개념은 개별적이거나, 사적인 용도의 온톨로지만이 아니라 그룹의 의견 수렴을 통해 그룹이 동의한 일치된 지식을 담고 있는 것을 말한다.

온톨로지 공학은 온톨로지 개발과정, 즉, 다루고자 하는 도메인에서 사용되는 용어들을 정의하고 그들 사이의 관계를 규정하여 그 과정을 구현하는 것을 의미한다. 이 과정에 포함되는 것으로 개념들의 클래스화, 그 개념들의 관계를 클래스의 계층적인 구조로 성립, 클래스들의 특성과 속성 슬롯화 및 그 특성과 속성에 존재하는 다양성과 제한요소 정의, 개체들의 규정을 들 수 있다. 이를 통해 얻어진 공동의 단어들과 공유된 이해를 바탕으로 상호운용시 의사 소통의 일관성을 추구한다.

2.2 온톨로지의 필요성

사람들이 웹을 통해 행정적인 또는 생명과학 분야 등에서 유용한 정보를 구하려고 할 때에는, 신속하고 정확하며 유용한 정보검출을 기대한다. 이러한 분야들은 여러 부분으로 나누어져 있고 관련 데이터들도 여러 사람들에 의해 계속적으로 모아지고 있는 실정이다. 사람들은 이러한 이질성의 데이터를 넘나드는 검색, 더욱이 부분적인 중복 또는 보완의 의미를 가진 데이터들을 섭렵한 정보추출을 요한다. 이를 위해서는 데이터의 표준화가 요구되고, 이 표준화를 통해 정보공유와 다른 정보의 올바른 이해가 가능하다. 온톨로지 개발은 정보의 상호 교통을 가능하게 하고 필요한 정보의 추출을 위해서 정보의 구조를 사람들과 소프트웨어 에이전트들 간에 서로 공동 합의된 형태

로 구성한다. 이러한 협력적 구성은 도메인 지식의 재사용을 또한 가능하게 한다. 온톨로지를 통해 도메인 내의 가설을 명시하여 차후 수정이 용이하도록 하고, 도메인 지식을 운영상의 지식과 구분하며 도메인 지식을 분석, 이해하기 위해서도 온톨로지는 필요하다. 이러한 역할을 수행하기 위해 제작되는 바람직한 온톨로지는:

- 특정 도메인의 전문적인 용어들의 관계를 나타내는 전문가가 사용 가능한 메커니즘
- 전문적인 지식 없이도 빠르게 정보를 찾을 수 있도록 용어들을 모아놓은 메커니즘
- 컴퓨터 프로그램을 통해 사용자가 원하는 정보를 같은 도메인의 다른 데이터 집합들의 일관성 있는 내용과 정보를 찾아 추출할 수 있도록 자동화된 추론이 가능한 메커니즘과 새로운 도메인과의 합병이 가능한 메커니즘

을 포함한 것이다. 온톨로지를 이용한 지식 모델링이 가져다 주는 기술적인 장점은 지식베이스의 일관성을 자동화하여 검사한다는 것, 특정하게 명시되지 않은 데이터를 유추할 수 있는 추론 메커니즘 사용 가능성, 사용자의 요구를 광범위하게 수용함과 검색을 용이하게 해 준다는 것 등을 들 수 있다. 더욱이 지식베이스를 디자인함으로써 얻을 수 있는 즉각적인 혜택은 필요로 하는 지식에 대해 설명을 하고, 그에 맞추어 지식을 구성하여 구조에 맞추는 과정을 통한 처리과정 자체를 모델링 한다는 데에 있다.

2.3 관련문제의 영역

웹 정보의 대량화는 기존의 검색문제를 분산된 정보의 의미를 고려한 일관성 있는 정보검색의 문제로 진화하게 하였다. 예를 들어, 생명 정보학 같은 분야에서 필요로 하는 정보검색은 더 이상 특정 유전자를 찾는 것에 국한된 것이 아니라 생명현상이 시스템적인 현상이라는 점에 착안하여 기능이 발견되어진 유전자가 어떠한 질병과 어떻게 영향을 주고 있는가 하는 관련성을 연구하는 문제가 해결 문제의 관건이다. 이러한 형태의 문제, 즉, 연계성을 찾는 문제는 생명 정보학 같이 정보의 양이 급격히 늘어나는 대량의 정보처리 문제뿐만 아니라 문제의 다양성 만큼이나 다양한 형태를 가진 데이터로 인하여 이러한 연관관계 또는 관련지식을 찾아내는데 더욱 어려움을 겪고 있다.

이러한 문제의 범위를 다루는 시스템은 대량의 정보를 저장하거나 검색하는 기능이 있어야 하고 도메인 지식을 기관, 개인들이 협력하여 연구, 공유할 수 있어야 하며 실험 및 결과를 통해 포착한 지식들은 그 질을 향상함과 함께 지식을 분류, 출간, 구성하는 서비스 차원의 기능도 갖추고 있어야 한다. 시맨틱 웹에서는 다양한 층의 온톨로지를 개발하여 연관관계 및 메타데이터 추출의 자동화를 추구한다.

3. 온톨로지 종류와 애플리케이션

온톨로지 사용의 또 다른 주된 목적은 효율적인 재사용에 있다. 데이터베이스 스키마 역시 개념화를 추구한 것이지만 오직 하나의 애플리케이션에 관한 것이라면, 온톨로지는 다양한 애플리케이션에 적용 가능하다는 것에 그 차이점을 둔다. 그러나 모든 온톨로지가 같은 사용 목적으로 개발되는 것이 아니고, 그 내용의 범위나 세부적인 내용에 많은 차이가 있으며 부분적으로 재사용 가능한 부분이 있는가 하면 재사용이 불가능한 부분들도 있다.

온톨로지를 종류별로 구분하여 보면

1. 도메인 기반 : 대장균과 같은 특정 도메인 또는 유전자 기능과 같은 종합적인 도메인 기반 온톨로지
2. 타스크 기반 : 주식 분석과 같은 특정 타스크 또는 문제 해결과 같은 보편적인 타스크를 위한 온톨로지
3. 일반적인 : 일반적이고 공통적인 개념들을 나타내는 온톨로지로서 자연언어 처리에 활용되는 온톨로지. 온톨로지를 재사용하는데 유용하게 사용될 상위개념의 또는 참조적인 온톨로지

로 나눌 수 있다.

온톨로지가 사용되는 애플리케이션의 종류는 커뮤니티의 지식을 참조, 공유하기 위한 온톨로지부터 디자인 처리를 위한 온톨로지, 서비스를 합성하기 위한 온톨로지에 이르기까지 용도별로 구분 지을 수 있다. 사용 가능한 온톨로지들의 예를 들자면, 온톨로지 라이브러리로는 Ontolingua ontology library (www.ksl.stanford.edu/software/ontolingua/), DAML (www.daml.org/ontologies/), Protégé(protégé.Stanford.edu/plugins.html)와 개괄적인 라이브러리인 UNSP SC(www.unspsc.org), Rosetta-Net(www.rosettanel.org), DMOZ(www.dmoz.org), WordNet

(www.cogsci.princeton.edu/~wn), 그리고 특정 도메인에 관한 UMLS Semantic Net, Gene Ontology (www.geneontology.org) 등이 있다.

다음은 온톨로지가 사용되는 가능한 애플리케이션들을 용도별로 그룹지어 분석한 것이다[2].

- 중립적인 저작 : 가공된 정보를 하나의 언어로 저술하여 후에 다양한 시스템에 맞게 변환, 사용되도록 하는 것. 지식 재사용, 지속적인 장기 지식 보유의 이점이 있다.
- 명세서로서의 온톨로지 : 주어진 도메인에 맞게 제작되어 소프트웨어 개발의 명세서적 역할을 하는 것. 문서화, 유지, 신뢰성과 지식 사용의 이점을 들 수 있다. EXPRESS와 같은 데이터모델링 언어가 데이터베이스 스키마 제작을 위해 사용된 경우, UML같은 정보와 시스템 모델링을 위한 언어를 통해 특정 소프트웨어를 위한 온톨로지 제작, 또는 메소드의 재사용과 클래스 상속이라는 관점에서 클래스의 분류를 통한 객체 지향적 모델링이 온톨로지를 명세서적 용도로 사용한 경우로 들 수 있다. 온톨로지의 초기적 응용분야로 많이 사용되지만 그동안 온톨로지 분야의 주된 관심분야는 아니었다.
- 정보의 일반적인 이용을 위한 온톨로지 : 사람이나 애플리케이션이 정보를 이용하려 할 때 정보들이 익숙치 않은 어휘로 구성되어 있거나 접근 불가능한 형태로 나타나 있는 경우들이 있다. 이러한 경우 온톨로지는 공유된 개념의 어휘로 또는 관련 지어지는 용어들을 찾아서 제공함으로써 정보를 이해할 수 있게 한다. 장점으로는 상호운용성, 지식 자원의 효과적인 사용과 재사용을 들 수 있다. 분자생물학 분야의 EcoCyc[3]가 이 부분의 대표적인 예로 공유된 온톨로지를 사용하여 다양한 형태의 데이터베이스를 사용할 수 있게 한다. 이 외에도 공유된 온톨로지 없이 매핑 규칙들을 통하여 특정 온톨로지의 개념이 다른 온톨로지에 사용되는 것을 중재하여 데이터를 공유하는 방식도 포함된다.
- 온톨로지 기반 검색 : 문서 또는 특정인의 웹 페이지를 찾고자 자원을 검색할 때 온톨로지를 사용하는 경우. 중요한 정보가 있는 자원을 빠르게 찾아 사용할 수 있다는 것과 자원을 찾는 정확도를 향상시키는 것이 대표적인 이점이다. 검색엔진이 온톨로지에 정의된 개념과 규칙들을 활용하면서, 온톨로지를 검색을 향상시키기 위한 추론의 기반으로

이용한다. 예를 들어, 특정 기능을 같이 하는 단백질질을 찾아서 그와 관련된 다른 단백질들을 찾는 것과 같이 온톨로지에 표현된 관계에 따라 질의를 다시 적절하게 바꿀 수 있게 하여 검색을 한다. 상용 인터넷 포털들이 이 분야의 온톨로지 활용을 추진하고 있으며, 온톨로지 기반 시맨틱 검색에 대한 연구가 활성화 되고 있다.

- 데이터베이스의 주석과 기술 문헌의 이해를 위한 온톨로지 : 자연언어 처리를 돕기 위해 개발된 온톨로지로서 주석과 문헌에 있는 도메인 지식을 연결하면서 그 지식들이 문법이나 어휘 등의 언어학적인 구조와 어떻게 관련 있는지를 규정한다.

4. 온톨로지 개발의 온톨로지 공학

온톨로지를 개발하는데 있어서 정석이란 없다. 도메인 모델링과 마찬가지로 적절한 온톨로지 개발은 고려중에 있는 애플리케이션과 그 애플리케이션에서 예측되는 확장성에 달려있다.

4.1 온톨로지 라이프사이클

온톨로지 개발은 초기 안을 작성한 후 필수적으로 반복적인 과정을 거쳐 평가하고 수정해 나아간다. 관심 분야의 애플리케이션이나 문제해결 방법에 온톨로지를 적용하면서 또는 그 분야의 전문가와의 상의를 통해 세부적인 사항들을 채워가며 수정하고 가꾸가는 것이다. 그 과정 중에 디자이너는 장·단점과 다른 해결책의 보완을 고려하여 모델링 결정을 해 나가야 한다. 다음은 개괄적인 온톨로지 개발과정을 살펴본다[4].

4.1.1 초기 온톨로지 디자인

- A. 도메인의 특성과 온톨로지 사용용도를 고려하여 온톨로지 구성영역을 정한다.
- B. 온톨로지 영역: 온톨로지가 응답해야 하는 질의들을 나열하여 온톨로지 내에 포함되어야 할 내용의 세부 사항들을 정한다.
- C. 재사용 정도: 기존의 온톨로지 활용을 경제적으로 하기 위해서 그리고 특정 온톨로지를 이미 사용하고 있는 애플리케이션과 상호적인 처리를 하기 위해서 기존 온톨로지 활용은 중요하다.
- D. 온톨로지 학습: 언어학적인 분석과 기계 학습 관계를 이해하여 멀티미디어와 같은 다양한 자

료에서 자동적으로 학습될 수 있는 부분을 고려한다.

- E. 온톨로지 개발 목적에 따라서 메타데이터의 주석이나, 찾는 대상의 구조, 개인별 특징도 고려한다.

4.1.2 개념적인 개선

- A. 개념적 구조에서 시작하여 개념들의 확인, 수정, 증진을 시키면서 중요한 용어와 클래스의 계층구조, 개념의 내부적 구조인 특성(slot)과 특성들이 갖는 값들의 제한점(constraints)이나 다른 형태의 값들(facets)을 정의한다.
- B. 실제 적용대상인 콘텐츠에 맞추어 기존의 온톨로지를 수정, 재사용한다.
- C. 모델의 재공학: 기존에 써왔던 개념의 개요나 유의어 사전과 같은 것을 수정하여 활용한다.
- D. 개념들에 대한 일반적인 의견이나 다른 관점을 수용할 수 있는 톨과 방법들도 동시 개발한다.
- E. 최적의 모듈화를 고려하고 모듈간의 관계를 나타낸다.
- F. 특정 태스크 수행에 있어서 필요한 위임, 위탁, 제어, 최소한의 여론, 일치성을 검증하기 위해 필요한 관련 에이전트들 간의 상호작용 방법을 활용한다.
- G. 점층적인 수정 개발 작업을 가능케 할 추론을 고려한다.

4.1.3 온톨로지 간의 연결

- A. 정렬, 제휴 방법
- B. 통합, 병합 방법이 있다.

4.1.4 평가단계

- A. 온톨로지가 요구조건에 맞는지 확인
- B. 다른 애플리케이션에 사용하여 실험적인 결과를 분석
- C. 온톨로지의 일관성 확인
- D. 표현능력, 정확도, 도메인, 인지학적 적절성 등 이론적인 방법을 통하여 온톨로지 비교

4.1.5 진화과정

- A. 변화가는 즉, 버전이 다른 온톨로지 간의 관계 성립
- B. 온톨로지 사용과 새로운 정보를 통해 온톨로지 변화의 필요성을 고려한다.

C. 온톨로지 변화가 메타데이터 주석에 미치는 영향을 고려한다.

4.2 지식표현 언어

온톨로지가 애플리케이션내에서 사용되기 위해서는 위의 개발단계의 기호화(encoding) 과정과 같이 컴퓨터가 처리할 수 있는 구체적이고 형식적인 표현으로 구현하여야 한다. 많은 언어들이 개념적인 모델들을 표현하기 위해 연구, 사용되어져 왔고 이들은 표현범위, 사용용이도, 계산의 복잡성들로 언어의 특색을 달리하고 있다. 이들 언어의 선택은 주로 온톨로지를 구성하고 있는 요소들을 얼마만큼 형식적으로 표현이 가능하고, 다루기 쉬우며, 명백하고 정확하게 나타낼 수 있는가를 말하는 인코딩 언어의 표현력(expressivity)과 온톨로지 내용표현의 만족도와 일관성의 정도를 말하는 언어의 엄밀함(rigour), 그리고 언어의 의미를 확실하게 표현할 수 있는 그 언어의 의미론(semantics)에 의해 결정된다[5].

온톨로지를 규명하기 위해서 사용되는 언어들은 크게 Gene Ontology와 같이 트리 구조의 상속관계를 나타내기 위한 어휘들을 정의하는 언어, 프레임(frame)과 같이 객체를 기반으로 한 지식표현을 위한 언어, 서술어(predicate)를 논리적으로 나타내는 서술논리(description logic) 언어로 나누어 볼 수 있다. 표현언어는 개념, 관계표현, 관계의 전이표현 등의 표현 가능한 범위(expressive power)와 계산처리 특징(computational property)에 준하여 구분할 수 있고 이들 언어의 선택은 표현하려는 시맨틱의 복잡도에 따라 위의 두 상관관계에 있는 표현능력과 계산처리 능력을 비교하여 결정한다. 이 외에도 추론엔진에서 사용되는 (전문가 시스템을 중심으로 한) 알고리즘 중심의 지식표현 언어도 포함된다. 시맨틱 웹 상의 온톨로지 언어의 자세한 내용은 이번 시맨틱 특집의 다른 논문과 참고 문헌들에서 찾아 볼 수 있다[5][6].

4.3 온톨로지 디자인 방법

온톨로지 공학은 원칙적인 디자인, 수정, 응용, 평가단계를 거쳐 온톨로지를 개발한다. 이러한 온톨로지 개발의 디자인 방법을 아래 다섯 분류로 나누어 볼 수 있으며 이들의 혼합적인 방법으로도 온톨로지를 개발할 수 있다[7].

• 고무적인 방법 : 개발자의 개인적인 관점을 기반으

로 온톨로지를 형성하는 방법

- 귀납적인 방법 : 관심분야의 특정 케이스를 관찰, 조사, 분석하여 얻어진 특징들을 그 분야의 다른 케이스에 일반화하여 적용시키는 방법
- 연역적인 방법 : 일반적인 원칙들로 이루어진 온톨로지를 특정한 케이스에 맞춰 점차적으로 수정해 가는 방법
- 통합적인 방법 : 이미 존재하는 온톨로지들로부터 부분적으로 관심분야를 나타내는 부분들을 합성하여 하나의 단일화된 온톨로지를 구성하는 방법
- 협력적인 방법 : 초기의 온톨로지를 기준으로 하여 여러 사람이 함께 협력하여 개발하는 방법

감화적인 방법은 이론적인 바탕이 부족하고 비현실적일 수 있으나, 독특하고 혁신적인 온톨로지를 개발할 수 있다. 귀납적인 방법은 일반성을 결여하기 쉽고 연역적인 방법은 특정케이스를 끌어내기 위해 선택할 수 있는 일반성을 지닌 스키마가 존재한다는 가정에 준한다. 통합적인 방법은 앞의 세가지 방법을 통합적으로 적용하나 개발자의 해석이나 통합기술에 의존적이다. 앞의 네 가지 방법에 비해서 협력적인 방법은 다수의 관점을 통해 지속적인 평가와 제안을 통해 온톨로지를 개발, 수용하는 반면, 개발자들의 계속적인 참여가 요구되고 그들의 개발기술에 영향을 받는다. 일치된 정보를 공유하고 재사용 한다는 차원에서 협력적인 방법이 주로 시도되고 있는 실정이며, 이 방법은 표현되어진 개념과 그러한 개념과 적절하게 관련지어진 정보들이 공동집단의 의견과 일치되도록 계속 변화할 수 있다는 것이 큰 장점이다. 온톨로지를 통해 얻으려고 하는 지식관리 또는 경영은 계속적으로 변화하는 것이고 여러 전문 분야에 걸친 주제를 다룬다는 점에 비추어 볼 때 다양한 분야에 관한 온톨로지 개발을 위해서 연구의 폭, 디자인의 기준, 평가기준이 현실적이고, 견고하며 다양해야 한다.

5. 온톨로지 개발도구 및 제품들

공유되는 온톨로지 개발의 병목현상은 기술보다는 상호적인 교류에 있다는[8] 점에 비추어 온톨로지 개발 도구들은 온톨로지 작성과 복수의 온톨로지를 합병하는데 필요하고 특히 협력적인 온톨로지를 구성하는데 있어서는 절대적으로 필요하다. 만들려고 하는 대부분의 개념적인 모델들은 복잡하고 다루기 힘든 것이다. 이를 돕는 온톨로지 제작 툴들은 대

체적으로 결과물을 시각적으로 보고 검증하도록 다차원의 그래프로 표현할 수 있게 되어 있고 특정한 모델의 완성도를 테스트 할 수 있는 논리적인 방법도 갖추고 있다. 본 논문에서는 대표적인 툴 몇 가지를 소개한다. 다른 툴과 툴 간의 특성 비교에 대한 자세한 내용은 [9]와 [10]에서 찾아볼 수 있다.

- Ontolingua(<http://ontolingua.stanford.edu/>)

Ontolingua는 웹 상에 분산되어 있는 온톨로지들을 사용자가 생성하고 관리할 수 있도록 도와주는 시스템이다. Ontolingua가 지원하는 두 가지 중요한 기능은 여러 개의 온톨로지들을 통합하거나, 여러 개의 온톨로지나 단일 온톨로지를 해석할 수 있도록 해준다. 지식베이스를 각기 다른 형태로 가져와서 taxonomy들을 재구성 하고 이름 충돌 문제를 해결할 수도 있다. 온톨로지를 브라우즈 할 수 있으며, 용어들을 편집할 수도 있다. 이 외에도 유용한 기능들이 많이 포함되어 있다.

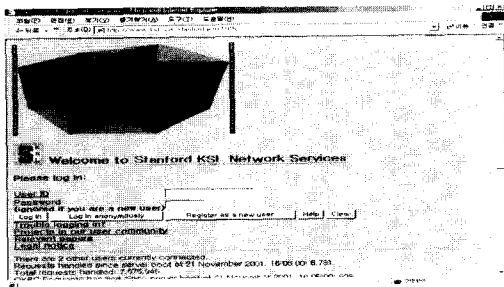


그림 1 Ontolingua

- OntoEdit(<http://www.ontoprise.de>)

응용 시스템의 지식 모델을 가져와서, 적용, 설계하기 위한 개발 환경을 제공한다. 계층 구조의 개념적 정의와 다중 상속이 가능하고 지역적이고 전역적인 관계를 정의할 수 있다. OntoEdit에 포함된 다른 특징들로: 다국어 사용, 개념들 간의 분리, 대칭적 관계의 기본적인 axiom 정의, RDF import/export, DAML+OIL import/export, 추론 엔진인 OntoBroker의 데이터 포맷인 Flogic에 대한 import/export 등이 있다.

- OilEd(<http://oiled.man.ac.uk/>)

OilEd는 DAML+OIL을 사용하여 온톨로지를 개발할 수 있는 에디터로, 현재로는 대규모의 온톨로지 제작이나 온톨로지 합병 등에 치중하지 않고 개발한 온톨로지의 일관성 검사를 FcCT를 통해 추론할 수 있는 정도의 기능을 제공한다.

- Protege-2000(<http://protege.Stanford.edu/>)

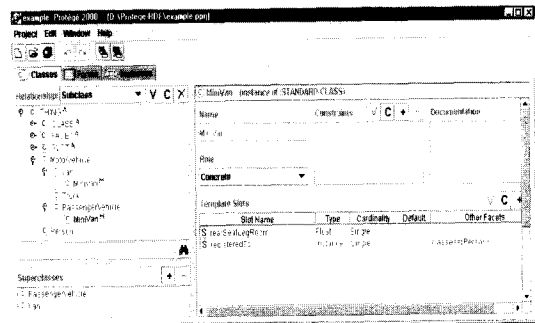


그림 2 Protégé-2000 Editor

온톨로지와 지식 기반을 수정하고 생성하기 위한 확장이 용이하며, 플랫폼 독립적인 환경이다. 이 제품의 특징은 직관적이고 사용하기 쉬운 GUI, 필요할 때 데이터베이스 back-end를 가져와 기억장치에 캐싱하여 사용한다. 확장 가능한 플러그인 구조를 지원한다.

- OntoWeb(<http://ontoweb.semanticweb.org/>)

OntoWeb은 지식 경영과 전자상거래를 위한 온톨로지 기반의 정보교환을 위한 것이다. 다양하고 이질적인 단체 사이에서 통신 병목현상을 피하는 것이 OntoWeb 네트워크의 목표이다. OntoWeb은 European commission에서 내는 특정 주제의 네트워크 제안의 대표적인 예이다.

- Network Inference(<http://networkinference.semanticweb.org/>)

영국의 Manchester에 기반을 두고 있는 사이트로 FaCT(Fast Classifier of Terminologies) 시스템의 개발로 인해 지식 표현 공동체의 선두가 되었고 W3C의 시맨틱 웹 Activity의 주요 요소로 주목 받았다. Network Inference에서 개발된 CerebralM은 기계가 이해할 수 있는 메커니즘으로 기술된 논리나 추론 엔진을 사용한다. 이것은 온톨로지의 한 형태로 복잡한 개념을 표현하고 논리구조를 제약할 수 있게 한다. 최근 메타데이터를 연결짓는 시맨틱 그리드(SemanticGrid, <http://www.semanticgrid.org>): 시맨틱 웹 + 웹 서비스 + 그리드(grid), 프로젝트 중의 하나인 myGrid(<http://www.mygrid.org.uk/>)에서 사용될 추론 엔진으로 채택되었다.

6. 온톨로지 공학에서 해결해야 할 연구문제

온톨로지 공학의 목적은 온톨로지 개발과 사용을 온톨로지 라이프사이클 디자인, 평가, 검증, 유지, 전개, 매핑, 통합, 공유와 재사용 과정을 거쳐 효과적으로 지원하는 것에 있다. 온톨로지 개발은 시간과 비용이 많이 들고 쉽지 않은 작업이다. 디자인의 목적이 자동화된 추론에 적용되기 위한 것이라면 형식화하는 과정을 포함하여 더욱 그러하다. 이에 온톨로지 공학에서 해결해야 하는 문제들을 알아본다.

- 온톨로지에 함축된 가정

용도에 맞는 온톨로지를 개발하는 데 있어, 또 다른 어려움은 온톨로지에 내재되어 있는 가정들을 식별해 내는 것이다. 이러한 어려움은 협력적인 작업을 통해서도 쉽게 해결되지 않으며 온톨로지가 공유되고 재사용 되기 위해 개발되는 근본적인 목적을 달성하는 데 있어 장애요인이 된다. 디자인 과정에서 이러한 함축적인 가정을 피하여야 하고 이러한 가정들을 식별할 수 있는 메커니즘이 필요하다.

- 이론적 분석 vs. 실험적 결과

현재 온톨로지 공학에서는 문제분류에 따라서 적용한 온톨로지의 적절성이나 성능을 다른 온톨로지와 평가, 비교할 수 있는 벤치마킹이 없다. 이론적으로는 논리적 언어의 원리(axioms)들로 온톨로지의 명세서를 나타내어 온톨로지들을 비교, 분석할 수 있는 부분도 있지만, 온톨로지의 형식적인 특징을 나타내는 기술적인 면과 실제적인 시맨틱 표현능력과의 상호적인 관계 같이 어떠한 부분은 애플리케이션 내에서의 실험적인 분석을 통해서만 평가가 가능한 부분도 있다. 개발된 온톨로지는 기술적으로 공유된 지식시스템 안에서 유사개념들을 적절하게 식별해 내야 하기 때문에 이러한 분석은 중요하다.

- 실험가능한 애플리케이션의 부재

현재의 온톨로지 디자인 방법과 기술이 발전하기 위해서는 실제적이면서 시험 가능한 산업적 문제들이 많이 있어야 한다. 개발된 온톨로지의 적합성을 실험해 보기 위한 시험적인 애플리케이션은 특히 개별적으로 제작된 작은 규모의 온톨로지의 집합체인 대규모의 온톨로지 활용을 필요로 하는 것이 적절하다.

7. 시맨틱 웹과 온톨로지의 장래성

시맨틱 웹 구현의 어려움은 웹 상에 존재하는 다양한 형태의 표현들을 이해할 수 있는 프레임워크 구성

에 있다. 즉, 시맨틱 웹 적용의 어려움은 표현하려는 데이터 자체가 아니라 컴퓨터가 정확하고 빠르게 이해, 처리할 수 있는 형태의 데이터로 그 규칙과 의미에 있다는 것이다. 반면에, 시맨틱 웹의 이점은 다양한 형태의 정보들을 컴퓨터가 이해, 연관 지을 수 있을 때에야 충분히 언어될 수 있으므로, 시맨틱 웹의 미래는 시맨틱 웹 상에 존재하는 온톨로지에 달려 있다고 해도 과언이 아니다. 온톨로지의 성장은 곧, 시맨틱 웹의 성장으로 이것을 통해 팀 버너스 리가 예견하듯이 지금의 웹이 진화되어 컴퓨터가 웹 상의 콘텐츠를 이해할 수 있게 되는 것을 의미하게 된다. 온톨로지, 시맨틱 웹을 통해 소프트웨어, 컴퓨터, 사람이 함께 이해할 수 있는 웹으로의 진화는 이미 웹의 범용성을 통해 시작된 생태계를 보다 더 협력적인 관계로 형성하기 위한 사람이 소프트웨어와 컴퓨터에 공동 생명력을 부여하는 우호적인 초대인 것이다.

참고문헌

- [1] T. Gruber, A Translation Approach to Portable Ontology Specifications, in Knowledge Acquisition Journal, Vol. 5, pp. 199-220, 1993.
- [2] R. Jasper and M. Uschold, A Framework for Understanding and Classifying Ontology Applications, IJCAI-99 Ontology Workshop, Stockholm, Sweden, July, 1999.
- [3] P. Karp et al., EcoCyc: Encyclopedia of e.coli genes and metabolism. Nucleic Acids Res., vol. 24, pp. 32-40, 1996.
- [4] N. Noy and D. McGuinness Ontology Development 101: A Guide to Creating Your First Ontology 2002, http://protege.stanford.edu/publications/ontology_development/ontology101.html
- [5] A. Gomez-Perez and O. Corcho, Ontology Languages for the Semantic Web, IEEE Intelligent Systems, pp. 54-60, 2002.
- [6] R. Fikes, Providing Defaults in the Frames Portion of First-Order Logic Knowledge Bases, June, 2000.
- [7] C. Holsapple and K. Joshi, A Collaborative Approach to Ontology Design, IEEE Intelligent Systems, February, 2002.

- [8] V. Benjamins, The Ontological Engineering Initiative (KA), Formal Ontology in Information Systems, pp. 287-301, IOS Press, Amsterdam 1998.
- [9] 시맨틱 웹 기술을 적용한 전자상거래 표준 운용체계, ECIF, 2002.
- [10] A. Duineveld, M. Weiden, et al, Wonder Tools?:A comparative study of ontological engineering tools, International Journal of Intelligent Systems, Vol. 16, No. 3, pp. 377-390, 2001.



양 정 진

1985 이화여자대학교 전자계산학과(학사)
1992 University of Connecticut 공학석사
1998 University of Connecticut 공학박사
1999 IDIS Lab. at Uconn Post Doc.
1999~2000 University of Hartford 조교수
2001~현재 가톨릭대학교 컴퓨터정보공학부 조교수
E-mail : jungjin@catholic.ac.kr

● 제30회 춘계학술발표회 ●

- 일 자 : 2003년 4월 24~26일
- 장 소 : 제주대학교
- 문 의 처 : 한국정보과학회 사무국 Tel. 02-588-9246/7
<http://www.kiss.or.kr>, E-mail:kiss@kiss.or.kr