

碩士學位論文

Delphi 技法을 利用한
Ontology 設計에 관한 研究
(美千窟 觀光地區를 모델로)



濟州大學校 經營大學院

經營情報學科

姜 漢 燮

2004년 12월

碩士學位論文

Delphi 技法을 利用한
Ontology 設計에 관한 研究
(美千窟 觀光地區를 모델로)



濟州大學校 經營大學院

經營情報學科

姜 漢 燮

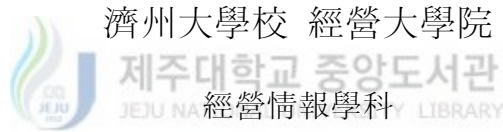
2004년 12월

Delphi 技法을 利用한
Ontology 設計에 관한 研究
(美千窟 觀光地區를 모델로)

指導教授 金 民 哲

이 論文을 經營學 碩士 學位論文으로 提出함

2004年 11月



姜 漢 燮

姜漢燮의 碩士 學位論文을 認准함

2004年 12月

審査委員長	_____	印
委 員	_____	印
委 員	_____	印

目 次

I. 서론

1. 연구의 목적	1
2. 연구의 배경	2
3. 논문의 구성	6

II. 이론적 배경

1. Web의 진화와 Semantic Web의 등장	7
2. 시맨틱 웹의 개요와 구조	8
3. 온톨로지의 개념과 정의	11
4. 온톨로지의 역할과 필요성	17
5. 시맨틱 웹에서 제공하는 온톨로지의 특징	19
6. 온톨로지 언어	
1) 메타데이터 스키마	21
2) 시맨틱 웹에서의 XML의 역할	22
3) 시맨틱 웹에서의 RDF의 역할	25
4) DAML+OIL	28
5) OWL	29
6) ezOWL	31

III. 지식표현을 위한 온톨로지 구축	
1. 지식표현의 정의와 방법론	34
2. 시스템 설계	38
3. 온톨로지의 개발과정과 라이프사이클	39
4. 온톨로지 설계 방법	42
5. 델파이 적용	45
6. 온톨로지 구축시 고려사항	46
7. 온톨로지 구축	48
IV. 온톨로지의 응용과 발전방향	
1. 응용사례	54
2. 자연어 이해	55
3. 지식관리	56
4. e-business	56
5. Ubiquitous Computing Business	56
V. 결론	58
참고문헌	60
부록 1 : 웹 온톨로지 모델링	63
부록 2 : 1차 설문지	66
부록 3 : 2차 설문지	69
ABSTRACT	72

그림 차례

그림 1. 웹의 진화 방향	8
그림 2. 시맨틱 웹의 계층적 구조	11
그림 3. ETRI의 'ezOWL'	32
그림 4. resolution refutation을 이용한 정리 증명의 예	35
그림 5. 의미망의 예 : 사무실 내 기계들	36
그림 6. Frame의 예: 호텔 방 내 비품들	37
그림 7. 시스템 구조	38
그림 8. 시스템 설계	39
그림 9. 협력적 온톨로지 설계방법	44
그림 10. 미천굴 관광지구의 클래스 계층구조도	49
그림 11. protege-2000에서의 클래스 계층 구조	50
그림 12. protege-2000에서의 클래스-슬롯과 속성결정	50
그림 13. 자동생성된 인스턴스	52
그림 14. Chart로 나타낸 조직기구도	52
그림 15. protege -2000에서의 구문형 Queries에 대해 검색된 결과	53
그림 16. 기존 방식과 온톨로지 기반의 자연어 처리방식	55

표 차 례

표 1. Possible interpretations of the term ‘Ontology’	14
표 2. Ontology Example	14
표 3. 온톨로지의 종류	16
표 4. 온톨로지 설계 방법	42



I. 서론

1. 연구의 목적

지식을 축적하고 활용하기 위한 계속된 시도들이 ‘시맨틱 웹(Semantic Web)’이라는 개념에 까지 이르게 되었는데 시맨틱 웹은 결국 인간의 지식을 정형화하려는 정보의 욕구에서 비롯된 산물이다. 이런 연구는 월드와이드웹(World Wide Web)의 탄생에 공헌한 팀 버너스리(Tim Berners-Lee)를 중심으로 한 웹 컨소시엄(www.w3c.org)과 학계 및 업계에서 활발히 전개되고 있다.

최근 많은 사람들에 의해 연구되고 있는 시맨틱 웹은 일반적인 웹과는 다르다. 일반적인 웹은 인터넷상에서 사용자에게 콘텐츠를 제공하는 정보 시스템을 말하는 반면에 시맨틱 웹은 사람뿐만 아니라 컴퓨터가 처리할 수 있는 데이터를 위한 웹을 의미한다.

Tim Berners-Lee에 의해 1989년에 처음 제안된 월드와이드웹(World Wide Web)은 널리 알려진 클라이언트-서버 개념과 쉽게 익힐 수 있는 간단한 HTML 언어를 이용하여 편리성을 추구한 덕분에 일반 사용자 누구나 쉽게 정보에 접근하거나 게시할 수 있게 되었고 결과적으로 폭발적인 정보의 증가를 가져왔다(최중민 외 2003).

1990년대 들어 정보통신의 발달과 인터넷 사용 환경의 편의성 제고 등으로 인해 웹의 사용이 일반화되고 각종 멀티미디어 정보들이 디지털화 되고 대용량의 데이터 교환이 가능해졌으며 단순한 텍스트뿐만 아니라 이미지, 음성, 동영상 등 다양한 형태의 자료를 공유할 수 있게 됨에 따라 다양한 정보의 공유와 그 방대한 양으로 인해 웹은 ‘정보의 바다’로 부각되었다(정인정 외 2003).

이런 단순성은 현재의 웹의 주요 성장요인이기도 하지만 한편으로 방대한 양으로 인해 사용자는 불필요한 정보를 걸러내느라 시간을 허비하는 문제점으로 작용하기도 한다. 현재의 웹은 사이트 개발자가 일방적으로 정해 놓은 검색필드 만을 검색할 수 있다는 문제점과 함께 데이터에 의미를 부여하여 사용자가 원하는 정보를 자동으로 추출하고 정보의 속성을 나타낼 수 없다는 단점을 가지고 있다. 즉 문서의 내용과 의미를 나타내는 시맨틱 정보를 표현하기가 어렵고, 사람이 아닌 프로그램 또는 소프트

웨어 에이전트가 자동으로 문서로부터 의미를 추출하기가 어렵다는 문제점을 해결하기 위해 1999년에 W3C(World Wide Web Consortium)에서 시맨틱 웹 (Semantic Web)을 제안하였다(S. A. McIlraith, T. C. Son, H. Zeng 2001).

인터넷은 정보소비자와 정보생산자의 구별을 없애고 정보획득 장벽을 철폐함으로써 정보유통의 민주화와 대중화를 가져왔다. 반면에 정보이용자는 자신이 원하는 정보만을 자신이 원하는 형태로 제공받기를 원한다. 지금처럼 거의 무한으로 확장되어 거짓정보와 진짜정보, 필요한 정보와 불필요한 정보가 뒤섞인 인터넷 환경에서 정보유통의 효율화를 이루면서 정보이용의 격차를 극복하고 정보이용을 극대화하려면 사용자가 원하는 정보만을 획득할 수 있는 방법론이 제공되어야 한다. 기존 인터넷은 정보과부하에 따라 한계에 도달했으며 이의 해결 방법은 의미에 의한 정보 접근이 될 수밖에 없다. 또 인터넷에 있는 정보의 양으로 볼 때 기계가 능동적으로 의미에 기초하여 사용자가 원하는 형태로 정보를 통합하고 가공하여 제공해야 한다. 이 문제를 극복하는 가장 실현 가능성 있는 해결방법은 시맨틱 웹이다(권혁철 2004).

Tim Berners-Lee에 의해 제안된 시맨틱 웹은 컴퓨터(Intelligent Software Agents)가 웹상의 정보를 이해하고, 정보를 창출할 수 있는 웹 환경으로 정보의 탐색과 의사결정이 인간이 아닌 컴퓨터가 할 수 있도록 만들어진 Web 환경이다. 즉 Semantic Web은 기계를 이용한 정보의 탐색과 의사결정을 지원하는 진정한 의미의 인터넷 환경이라 할 수 있다(박재홍 외 2002).

본 논문에서는 델파이기법을 이용하여 미천굴 관광지구의 특징적 요소를 합의 도출하여 지식표현을 통해 해당 관광지를 지식체계화 하고 이를 통해 온톨로지를 구축하여 확장된 시스템으로 발전시키는데 그 목적이 있다.

2. 연구의 배경

대내, 외 환경 변화로 국내 관광산업이 사실상 무한경쟁시대로 접어들면서 제주관광산업은 일대 전환기에 직면하고 있다. 미시적으로는 관광산업에 국한된 문제이지만 거시적으로 보면 정부의 공공기관 지방 이전 등과 맞물려 지방분권 시대의 가속화도 풀이된다. 그만큼 지자체간 경쟁은 앞으로 더욱 심화될 전망이다. 자유경쟁시장 원리상 경쟁력을 갖추지 않는 지자체의 자연도태는 필연적이라는 사실을 명심해야 할

시점이다.

제주관광에 있어 최대 위기로인은 국내 관광환경의 변화속도가 예상보다 빠르게 진행되고 있는 데다 가격 경쟁력 등에서 밀리고 있다는 점이다. 제주관광의 위협 요인은 예전에도 적지 않았지만 현재는 직접적인 타격을 줄 수 있다는 면에서 예전과는 상황이 다르다는 게 전문가들의 지적이다.

첫째로 고속철도시대의 개막은 그동안 지역간 이동에 있어 제약이 되었던 시간과 접근성을 파괴, 이른바 ‘당일치기’관광시대를 열었다고 할 수 있다. 올 들어 지자체마다 경쟁적으로 관광 상품, 축제 등을 개발하여 인센티브까지 부여하면서 관광산업 육성에 나서는 것도 이와 무관하지 않다. 다른 지방 지자체와 관광업계도 고속철도 개통을 내세워 적극적인 관광객 유치활동에 뛰어들고 있다. 이와 함께 주말 여가시간이 늘어남에 따라 자동차로 강원도 등 내륙지방 여행이 한결 여유로워지는 등 제주관광이 지리적, 경제적인 면에서 다른 국내 관광지에 열세를 보이고 있다(제민일보 2004).

둘째, 고속철도 개통은 2004년 7월1일부터 시행된 ‘주 5일 근무제’ 본격 시행과 맞물려 국민들의 여행 패턴을 획기적으로 뒤바꿀 요인임을 누구도 부인하지 않는다. 그만큼 잠재된 파괴력은 큰 셈이다. 제주의 경우 고속철과 카페리를 연계한 관광 상품이 개발되었으나 아직까지 이렇다 할 반응이 없어 더욱 다각적인 연계상품 개발이 요구되고 있다. 특히 고속철을 통해 부산항에서 일본 여행도 가능해진만큼 여행객들의 선택 폭은 그야말로 광범위해진 상태다. 여기에 현대아산에서 추진하는 금강산관광도 1일 육로관광 코스로 개발된 만큼 앞으로 제주관광 시장에 큰 위협요소로 작용할 것으로 보인다(제주관광신문 2004).

주 5일 근무제의 확대 실시로 인해 직장인들의 여가시간이 대폭 늘어났지만 제주관광시장의 특수로는 직접적으로 이어지지 않고 있다. 실제로 다른 지방의 경우 종전 1박2일로는 관광이 불가능했던 홍도, 흑산도, 거문도 상품 등이 쏟아지고 있다. 고속철도로 관광지 접근이 쉬워져 전국이 근거리 관광지로 좁혀진 것도 제주관광에 부정적인 영향을 미치고 있는 것으로 분석된다.

셋째, 매년 이어지는 항공요금 인상은 접근성과 경제성 면에서 제주관광의 경쟁력을 저하시키는 요인으로 작용하고 있다. 항공료 인상으로 인한 적지 않은 여행경비는 제주관광의 발목을 잡는 걸림돌이 될 것으로 관광업계는 내다보고 있다. 실제로 제주도관광협회 수도권 홍보관이 지난 6월 서울 코엑스에서 열린 제17회 한국국제관광전

에서 설문조사를 실시한 결과 주 5일 근무제로 가장 가고 싶은 여행지를 묻자 49.7%가 제주도라고 대답했지만 이들 응답자 중 58.1%는 항공료가 인상될 경우 국내외로 여행지를 변경하겠다고 답했다(제민일보 2004).

이제 주 5일근무제 실시로 200만 명에 가까운 직장인들이 새로운 경험의 세계를 맞게 되고 4-5인 가족 기준 최대 1000만 명에 이르는 새로운 관광수요가 발생한다고 한국관광공사에서 예측한 바 있다. 그러나 제주도의 관광정책은 취약한 조직, 인력 구조상 장기적인 대책을 갖고 추진하기보다 일회성이나 단편적인 사업 추진에 그치면서 한계점을 노출하고 있는 상황이다. 결국 제주관광은 무한경쟁시대에서 살아남기 위한 전략 마련과 장기 정책 수립, 이를 위한 조직 개편, 업계 자율적인 경쟁력 강화 방안 모색 등이 결합되지 않는 한 총체적 난국에서 쉽게 벗어나기는 어려울 것으로 보인다. 소득 증가와 맞물려 관광패턴은 이미 가족 단위 중심으로 급속하게 바뀌고 있고 씹씹이도 단순 소비형에서 체험, 웰빙, 휴식, 음식 등 특화형으로 빠르게 전환되고 있다. 특화형 관광패턴은 주 5일 근무제 본격시행과 맞물려 더욱 확산될 것이라는 전망이 지배적이다. 이런 추세와 맞물려 제주관광도 이미 ‘보는 관광’의 한계성이 누차 지적되어 ‘즐길거리, 먹거리, 볼거리’를 복합적으로 아우른 새로운 관광형태가 대안으로 제시되고 있다. 특히 국제자유도시 추진과 맞물려 일부 특화형 관광개발사업이 추진되고 있으나 제주관광은 아직까지 이렇다 할 변신에는 성공하지 못하고 있는 게 현실이다. 변화하는 고객 패턴에 발맞추어 제주도관광협회는 최근 제주여행 패턴이 가족단위 등 개별여행 추세로 변화함에 따라 개인별 성향에 맞는 맞춤형 정보를 제공하고 외국인 잠재 관광객 유치 증진을 위해 인터넷을 통한 홍보활동을 강화하고 있다. 이에 따라 제주도관광협회는 (주)다음커뮤니케이션을 이용한 내국인 대상 인터넷 홍보활동 외에 한국관광공사의 인터넷 ‘Tour2Korea.com’을 활용해 외국인 네티즌들을 대상으로 다양한 인터넷 홍보활동을 펼치기로 했다. 또한 2005년 2월까지 제주전용 콘텐츠를 구성하고 영어, 일본어, 중국어(간체, 번체) 등 주요 어권 사이트 메인페이지에 배너 광고를 게재하고 제주관광 뉴스란을 신설하여 한국관광공사 일본, 중국 지사 등과 연계한 온라인 홍보활동을 강화할 계획에 있다.

이밖에 관광외적인 면에서 볼 때 제주도와 정보통신부가 공동으로 추진하고 있는 IT(정보통신) 신성장동력의 대표적인 사업인 텔레매틱스 시범도시 사업이 확정되어 우선 1차 연도(2004년 8월 ~2005년 7월)에 50억원(국비 20억원, 도비 15억원, 민자 15

억원)을 투자할 계획이다. 2005년 7월까지 1년간 1차 연도 예산으로 정부 예산 30억원이 지원되는 등 2006년 7월까지 총 100억원(정보통신부 40억원, 제주도 및 민간기업 각각 30억원)이 투입된다(정통부 2004).

제주도의 텔레매틱스 시범도시 사업 추진은 ‘유비쿼터스 제주’ 건설의 대명제 아래에서 세계적인 관광도시인 제주도가 명실 공히 정보 관광도시로 거듭난다는 데 커다란 의미가 있다. 텔레매틱스는 원격통신(telecommunication)과 정보과학(infor-matics)의 합성어로 위성위치 확인시스템(GPS)과 지리정보시스템(GIS)을 활용해 차량운전자 휴대전화나 개인휴대단말기로 각종 정보를 제공하는 종합정보시스템이다. 이에 따라 관광객들과 운전자들은 이 서비스를 통해 맞춤형 여행 정보, 외국어통역서비스, 오락 등을 제공받을 수 있다.

제주도는 우선 올해 말까지 1차 연도에 제주도내 1000대의 렌터카에 관광, 교통정보 등을 제공하는 시범서비스를 실시하고 여행정보, 문화행사정보, V-Shop 정보, 레저생활정보, 엔터테인먼트 정보, Safe 제주정보서비스 등 텔레매틱스 관련 정보의 공동 활용을 위한 텔레매틱스 정보 센터를 올해 말까지 구축할 계획이다. 제주가 시범도시로 지정된 것은 지리적, 경제적으로 다른 지방보다 유리한 여건을 갖고 있기 때문이다. 우선 제주가 텔레매틱스의 주 고객층인 렌터카 운용량이 전국 최고인데다 국제적 관광지로 관광객이 많아 이 서비스의 활용과 홍보가 쉽다는 점이 이점으로 지적됐다. 또한 제주도가 지리적으로 독립된 공간이어서 서비스 및 대상지역 선정 시에 주변 지역에 미치는 영향을 최소화할 수 있는 ‘전과 청정지역’이라는 점도 시범도시 선정에 중요한 요인으로 작용했다.

전술한 관광시장의 변화에서 가족단위 관광객의 증가는 ‘맞춤 여행 시대’의 도래를 알리는 것이고 이로 인해 필연적으로 웹 서비스의 중요성은 더욱 강조될 것이다. 인터넷을 통해 제주관광을 검색하면 Yahoo에서 775개, Daum에서 346개의 웹 사이트가 검색되고 다른 검색사이트에서도 대동소이한 결과를 보여준다. 대부분의 여행자는 관광지나 여행 목적지 검색에서 원하는 정보를 얻기 위해 불필요한 정보들을 걸러내느라 시간적, 경제적 낭비를 감수하고 있다. 이를 해결하기 위한 방법이 바로 기계가 정보를 탐색하고 의사결정을 지원하는 인터넷 환경, 즉 온톨로지인 것이다.

3. 논문의 구성

본 논문의 구성을 보면 2장에서는 시맨틱 웹의 전반적인 개요와 함께 메타데이터, RDF 스키마 및 온톨로지 언어들을 기존에 발표된 연구문헌을 참고로 하여 온톨로지에 대한 전반적인 이론적 배경을 소개하였다. 3장에서는 미친굴 관광지구를 모델로 Delphi 기법을 이용하여 본격적인 지식표현과 이를 통한 온톨로지를 설계하였으며 Protege-2000을 기본적인 RDFS 구축을 위한 보조 에디터로 이용하였다. 4장에서는 온톨로지가 산업의 각 부문에서 어떻게 응용될 수 있는지를 살펴보면서, 온톨로지가 전자 도서관, 정보통합, HCI(human computer interfaces), 비즈니스 프로세스 모델링, 의료정보, 바이오 인포매틱스 등 다양한 영역에서 적용될 수 있는 가능성을 타진해 보았다.

본 논문에서는 지식표현을 통한 시스템 설계와 온톨로지 구축 작업까지 다루었고 시스템 구현은 추후 연구과제로 남기고자 한다.



II. 이론적 배경

1. Web의 진화와 Semantic Web의 등장

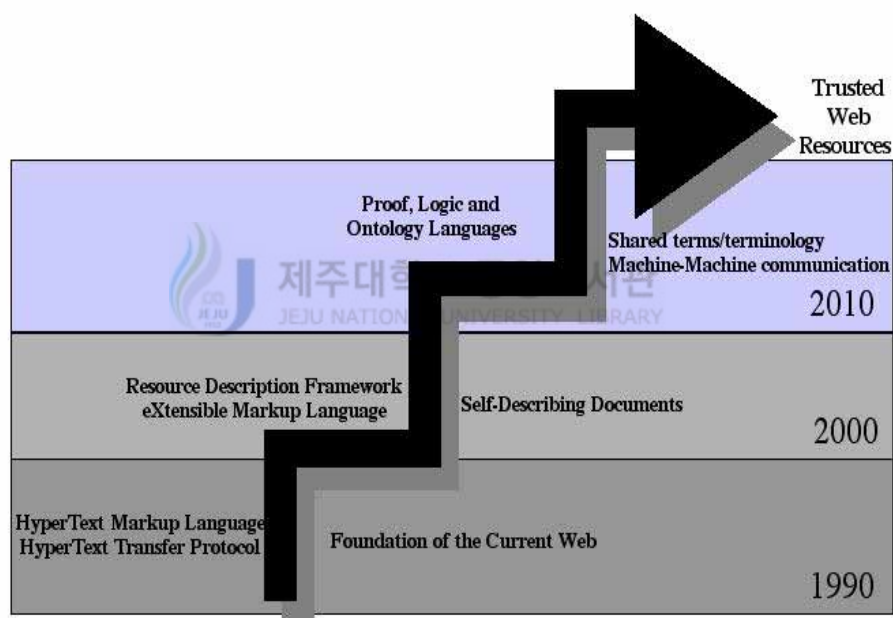
우리말로 ‘거미줄’을 뜻하는 단어인 웹(Web)을 제안한 Tim Berners-Lee는 그의 책 ‘Weaving the Web’에서 웹에 대한 비전은 ‘무엇이든지 서로 연결되는 공간을 지구상의 컴퓨터 하나하나에 실현시키는 것’이라고 했다. 수많은 하이퍼텍스트 링크에 의해 거대한 양의 정보가 서로 연결되는 꿈은 웹 기술과 함께 현실화되었다. 웹(the Web:고유명사로 단일한 정보 공간)에서 정보들 사이의 연결은 문서 형태로 되어 있는 개별정보에 고유한 주소를 부여하고 일정한 규칙에 따라 컴퓨터가 찾아 올 수 있게 함으로써 가능하게 되었다. 웹에 대한 비전을 현실화한 세 가지 중요한 기술적 요소로는 URI, HTTP, HTML이라고 할 수 있다(김홍기 2002).

URI(Uniform Resource Identifier)는 정보 공간에서 길을 찾기 위한 일종의 표시라고 할 수 있다. 팀 버너스리가 처음 IETF(Internet Engineering Task Force : 인터넷 운영 프로토콜의 표준을 정의하는 기구)에서 제안했을 때의 명칭은 UDI(Universal Document Identifier)이었다. 그러다가 URL(Uniform Resource Locator)이라는 이름으로 바뀌었고 최근에는 보편적이며 정보의 일관성을 반영하는 용어로 URI가 공식적으로 사용되고 있다. HTTP(Hypertext Transfer Protocol)는 하이퍼텍스트 브라우징에서 빠른 속도로 쉽게 웹 페이지를 찾을 수 있게 하는 프로토콜이다. HTTP의 입장에서 보면 웹 브라우저는 서버에 요청을 전달하는 클라이언트라 할 수 있다. 사용자가 URI를 입력하거나 하이퍼텍스트 링크를 클릭하여 파일을 요청하면 브라우저는 HTTP의 요청을 URI에 적혀 있는 IP 주소에 전달하고 지정된 서버망에서 요청된 파일을 찾아서 보내준다. HTML(Hypertext Markup Language)은 웹에서 어떤 컴퓨터라도 이해할 수 있는 하이퍼텍스트 언어로 개발되었다. HTML이 하이퍼텍스트를 표현하는 순서는 문서의 URI가 HTTP를 통해 서버에 접속한다고 클라이언트에 알리고 클라이언트가 처리 가능한 데이터의 형식을 서버에게 알려주게 된다. 그러면 서버는 해당 문서를 알맞은 데이터 형식으로 바꿔 전송해 준다. Tim Berners-Lee는 전 세계에 분산되어 있는 모든 정보자원을 연결시키는 공통된 정보 공간(information space)

으로서 웹의 궁극적 실현은 시맨틱 웹의 구현으로 가능하다고 보았다. 웹의 연결성은 하이퍼텍스트 링크를 통한 가상공간에서의 위치적 연결을 넘어선 정보 자원들 사이의 의미적 연결을 통해 이뤄질 수 있다.

<그림 1>은 웹의 계속되는 진화방향을 보여주고 있는데 궁극적으로 기계가 정보 자원의 의미를 이해하고 이를 바탕으로 논리적 추론이 가능할 수 있게 됨으로써 기계들 사이의 커뮤니케이션이 가능할 수 있는 웹으로 발전하게 될 것이다.

<그림 1> 웹의 진화방향



2. 시맨틱 웹의 개요와 구조

앞서 진술한 웹의 문제점은 검색측면에서 보면 웹 검색엔진이 주로 단어의 빈도수나 어휘 정보를 이용하여 문서의 유사도를 측정하고 랭킹을 매기기 때문에 사용자의 질의와는 관계없는 많은 문서를 결과로 나타내며 이로 인해 사용자는 불필요한 정보를 걸러내느라 시간을 낭비하게 된다. 또한 HTML로 여러 관련 문서를 확장하거나

통합, 공유하는 것도 매우 어렵다. 이런 문제가 발생하는 주된 원인은 현재의 웹이 사람을 위한 것이기 때문에 사람이 보고 잘 이해할 수 있도록 하기 위한 브라우저의 디스플레이 또는 레이아웃 기술에 초점을 맞추고 있기 때문이다. HTML 언어의 특징이 바로 디스플레이용이라는 사실이 이를 뒷받침하고 있다. HTML은 문서의 내용과 의미를 나타내는 시맨틱 정보를 표현하기가 어려우며 따라서 사람이 아닌 프로그램이나 소프트웨어 에이전트가 자동으로 문서로부터 의미를 추출하기가 힘들다.

시맨틱 웹은 메타데이터의 개념을 통하여 웹 문서에 시맨틱 정보를 덧붙이고 이를 이용하여 software agent가 이 의미정보를 자동으로 추출할 수 있는 패러다임을 조성하는 것이다. 부수적으로 의미정보의 자동 추출뿐 아니라 정보의 확장이나 공유도 가능하게 될 것이다.

팀 버너스리는 시맨틱 웹이 기존의 웹과 완전히 구별되는 개념이 아니라 현재 웹을 확장하여 웹에 올라오는 정보에 잘 정의된 의미를 부여하고 이를 통해 컴퓨터와 사람이 협동적으로 작업을 수행할 수 있도록 하는 패러다임이라고 그 역할을 정의하였다(Berners-Lee, T 의 2001).

시맨틱 웹을 한마디로 정의하면 '컴퓨터가 정보의 의미를 이해하고 의미를 조작할 수 있는 웹'이라고 말할 수 있다. 검색엔진과 에이전트 프로그램 등이 데이터 처리수준을 넘어 웹에 있는 정보의 의미를 이해한다는 것은 정보 리소스들 사이의 연관성을 표현해 주는 것을 말한다. 시맨틱 웹에서 문서의 각 부분을 컴퓨터가 이해할 수 있는 형식으로 기술(description)할 수만 있다면 복잡하게 얽혀져 있는 정보 리소스들 사이의 의미적 연관성으로 인해 웹을 통해 다양한 정보를 보다 효과적으로 활용할 수 있게 된다. 시맨틱 웹은 웹상의 정보에 잘 정의된 의미 (semantic)를 부여함으로써 사람뿐만 아니라 컴퓨터도 쉽게 문서의 의미를 해석할 수 있도록 하여 컴퓨터를 이용한 정보의 검색 및 해석, 통합 등의 업무를 자동화하기 위한 목적으로 제안되었다. 이러한 '잘 정의된 의미'를 다루고자 하는 것이 시맨틱 웹 온톨로지 언어의 역할이다 (Asuncion Gomez-Perez and Oscar Corcho 2002).

현재의 웹 환경은 컴퓨터를 매개로 하여 사람과 사람들 간의 정보 교환을 위해 발달되어 모든 정보의 추출과 이해를 사람에게 의존하기 때문에 사용자가 원하는 유용한 정보를 추출하는 일은 그리 간단하지 않다. 이런 환경을 개선하기 위해 에이전트, 검색엔진, 자연어 처리 등과 같은 기술이 웹 환경에서 적용되기 위하여 연구, 개발되

고 있으나 이러한 기술 역시 현재의 웹 환경이 지니고 있는 문제점으로 인해 한계를 나타내고 있다.

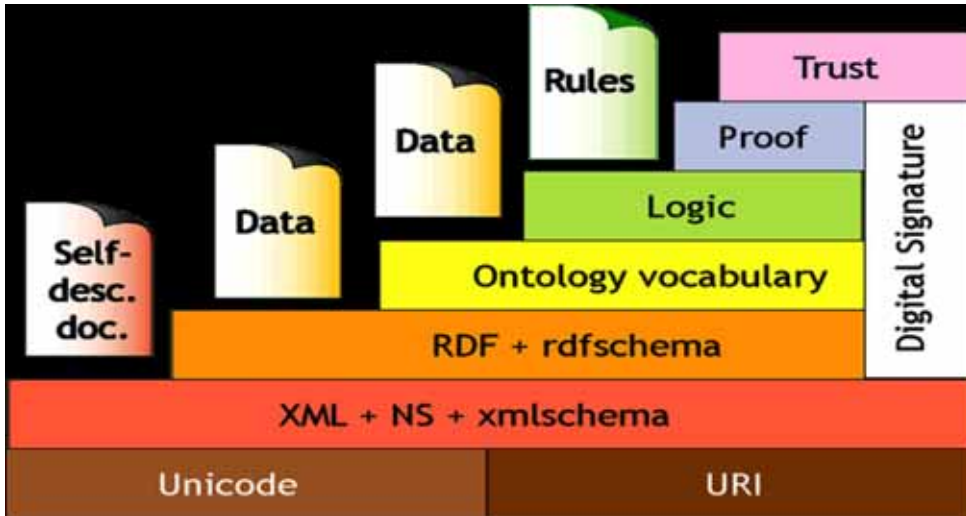
시맨틱 웹의 궁극적인 목적은 웹에 있는 정보를 컴퓨터가 좀 더 이해할 수 있도록 도와주는 표준과 기술을 개발하여 시맨틱 검색, 데이터 통합, 네비게이션, 타스크의 자동화 등을 지원하는 것인데 이를 세부적으로 기술하면 다음과 같은 작업을 가능하게 하는데 있다(최중민 2003).

- 정보를 검색할 때 더욱 정확한 결과를 가져온다.
- 서로 다른 이형질 소스의 정보를 통합하고 비교한다.
- 어떤 리소스에 대해서도 의미적이고 기술적인 정보를 연관시킨다.
- 웹 서비스의 자동화를 위해 웹에 세부 정보를 추가시킨다.

현재의 웹은 HTML 언어를 기반으로 확산되었으며 HTML로 이루어진 웹 문서들은 정보의 표현을 위해서 정보와 표현을 위한 태그와 복잡하게 섞여 있다. 이런 이유로 인하여 정보를 추출하기 어려울 뿐만 아니라 부울 연산에 의한 단어 검색을 위주로 하는 현재의 검색에서는 유의어, 동음이의어 등과 같은 문제를 해결하지 못한다. 이런 문제점을 개선하기 위하여 XML이 개발되었으나 XML도 정보의 정확한 표현은 가능하지만 의미 표현이 불가능하고 다양한 형태의 표현 가능성으로 인해 정보의 애매성을 증가시키는 단점을 지니고 있다. 시맨틱 웹은 분산처리 환경 하에서 서로 다른 데이터베이스 및 다른 기종 컴퓨터마다 상이한 의미를 갖는 데이터에게 각각의 컴퓨터 및 응용 프로그램들에서도 자유롭게 접근이 가능하도록 한다. 웹이 어떠한 형식의 데이터라도 처리할 수 있는 강력한 매체가 되기 위해서는 어떠한 데이터라도 표현할 수 있는 표현능력과 구문적 상호 운용성을 지지할 수 있는 능력을 갖추어야 할 뿐만 아니라 나아가 여러 응용프로그램에서 사용되는 다른 용어간의 의미적 관계를 매핑할 수 있어야 한다. 즉 앞으로 웹의 발전방향에서 의미적 상호운용성이 차지하는 비중이 커진 것이다(김이란 2001).

HTML을 기반으로 한 현재의 웹을 개선하는 기본 취지에서 보면 시맨틱 웹을 달성하기 위해 웹 프로토콜과 같은 하위레벨을 정의하고 이 하위레벨을 이용하여 다음 레벨의 개념을 정의하는 계층구조(layered structure)를 설정하는 것이 일반적인 연구 방향이다. 시맨틱 웹의 계층구조는 아래 그림과 같다.

<그림 2> 시맨틱 웹의 계층적 구조



시맨틱 웹을 구현하는데 있어서는 하위 계층의 기술들은 상위 계층의 기술들이 구현될 수 있는 기반이 된다. 또한 기술 발전의 시간적 순서는 하위계층에서부터 상위 계층으로 올라가고 있다. 이 계층구조에서 보면 가장 하위 레벨에서 웹 프로토콜에서 자원을 지칭하기 위한 주소지정 (addressing) 방법인 URI가 밑받침되고 이를 기반으로 XML과 Name space, RDF와 RDF 스키마, 온톨로지의 순으로 연구가 진행되고 있으며 그 위의 계층인 Logic에 대해서는 인공지능의 추론연구를 밑받침으로 일부 연구가 시작되었다. 또한 보다 더 상위 계층인 Proof와 Trust는 시맨틱 웹 정보의 신뢰성과 보안에 관한 내용으로서 아직 개념 정도만 얘기되고 있으며 차후 연구과제로 제시되고 있다.

HTML 언어와 같은 마크업 언어는 전자출판에서는 문서의 구조나 의미를 적절하게 바꿔 프린터나 웹 브라우저 등의 특정매체에서 출력이 가능할 수 있도록 해주는 기능을 가지고 있다. 이런 의미에서 마크업 언어는 일종의 메타언어라 할 수 있다.

웹에서 문서의 포맷을 정해주고 URI를 사용해 하이퍼텍스트 링크를 표시해 줄 수 있는 HTML 언어에서 태그는 출력 형식에만 상관하고 텍스트의 데이터 구조나 의미

와는 무관하다. 반면 1986년 Goldfarb에 의해 개발된 SGML(Standard Generalized Markup Language)은 문서의 출력 형식이 아니라 문서의 구조를 기술하는 언어라 할 수 있다. 의미부여를 위해서는 데이터뿐만 아니라 데이터에 대한 추론을 규정하는 규칙(rule)을 표현할 수 있는 언어가 필요한데 시맨틱 웹을 위한 의미를 부여하는 언어로 사용중인 XML(eXtensible Markup Language)과 RDF(Resource Description Frame work)가 사용되고 있다.

3. 온톨로지의 개념과 정의

‘온톨로지(ontology)’는 원래 철학 용어인데 우리말로 ‘존재론’으로 번역된다. 온톨로지, 곧 존재론이란 형이상학(metaphysics)의 한 분야로, 세계를 구성하는 모든 종류의 대상들의 존재에 관한 탐구이다.

즉, 이것은 ‘이 세계에는 어떤 종류의 존재자들이 존재하는가, 그들의 본성은 무엇인가, 그들 존재자들 사이에는 어떤 관계가 있는가, 그들 존재자들로부터 어떻게 세계가 구성 될 수 있는가’ 하는 문제들을 다룬다. 온톨로지는 용어 사이의 관계를 정의하고 있는 일종의 사전과 같은 것이라 할 수 있다. 온톨로지는 어휘의 정의를 다른 어휘와의 논리적 관계뿐만 아니라 가장 기본적(primitive) 어휘부터 파악해 나가는(bottom-out) 구조를 통해 나타낸다. 이는 지식 표현이 해당 분야의 전문가들이 동의하거나 공통적으로 사용하는 가장 기본적인 어휘에서 출발한다는 점에서 해당 분야에 특정적(domain specific)이라 할 수 있다. 온톨로지는 해당 분야의 전문가들의 관심, 목적, 관점을 반영한다. 따라서 온톨로지의 역할은 해당 분야에서 사용되는 어휘의 해석과 의미의 한계를 정하는데 있다. 이러한 의미에서 온톨로지는 사용되는 어휘 사이의 논리적 형식보다는 어휘의 내용을 중심으로 창조된다고 할 수 있다.

전산학 관련 문헌에서 ‘온톨로지’란 용어가 처음 등장한 것은 1960년대 후반이다.

그리고 인공지능 분야에서는 1980년대 초반부터 조금씩 사용되기 시작했다. 온톨로지에 대한 정의는 여러 가지가 있지만 Gruber는 온톨로지를 “공유된 개념화(shared conceptualization)에 대한 정형화되고 명시적인 명세(formal and explicit specification)”라고 정의하였다(T. Gruber 1993).

다시 말해서 “온톨로지는 공유(shared)하기 위한 개념들의 개념화(conceptualization)를 형식적(formal)이고 명백(explicit)하게 설명해 놓은 명세서(specification)”라고 할 수 있다. 이 정의를 세부적으로 살펴보면 다음과 같은 다섯 가지 용어가 복합되어 있음을 알 수 있다.

우선 공유(shared)라는 것은 개념이 해당 그룹 구성원뿐만 아니라 컴퓨터 간에 합의된 지식에 바탕을 두고 있다는 것으로 어느 개인에게만 국한되는 것이 아니라 그룹 구성원이 모두 동의하는 것을 의미하며 공유된 개념은 개별적이거나 사적인 용도의 온톨로지만이 아니라 그룹의 의견 수렴을 통해 그룹이 동의한 일치된 지식을 담고 있는 것을 말한다.

개념화(conceptualization)는 어떤 현상에 대해서 관련 있는 개념들을 식별하여 그 현상을 추상화한 모델로 설명하는 것으로 사람들이 사물에 대해 생각하는 바를 추상화한 모델을 말하며 대개는 특정한 분야에 국한시켜 논의된다.

명시적 명세(explicit specification)란 개념의 type 이나 사용상의 제약 조건들이 명시적으로 정의됨을 의미하며 정형화 혹은 형식화(formal)는 온톨로지는 프로그램이 이해할 수 있어야 하고 여러 단계의 정형화가 존재할 수 있음을 뜻하는 것으로 기계가 판독 가능한 형태이어야 한다는 것을 의미한다.

명시적(explicit)으로 표현한다는 것은 사용된 개념의 종류와 개념들이 추상화된 모델과 갖는 관계들, 그리고 그 개념들의 사용에 있어서 주어지는 제한점(constraints)들을 명백하게 정의하는 것을 의미한다.

온톨로지 공학은 온톨로지의 개발과정, 즉 다루고자 하는 도메인에서 사용되는 용어들을 정의하고 그들 사이의 관계를 규정하여 그 과정을 구현하는 것을 의미한다.

그 과정에 포함되는 것으로 개념들의 클래스화, 그 개념들의 관계를 클래스의 계층적인 구조로 성립시키고 클래스들의 특성과 속성 슬롯화 및 그 특성과 속성에 존재하는 다양성과 제한요소의 정의, 개체들의 규정을 들 수 있다. 이를 통해 얻어진 공동의 단어들과 공유된 이해를 바탕으로 상호운용 시 의사소통의 일관성을 추구한다.

‘온톨로지’라는 용어에 대하여 N. Guarino와 P. Giaretta는 다음의 <표1>과 같이 해석이 가능하다고 정리하고 있다.

<표 1> Possible interpretations of the term 'Ontology'

- | |
|--|
| <ol style="list-style-type: none"> 1. Ontology as a philosophical discipline 2. Ontology as an informal conceptual system 3. Ontology as an formal semantic account 4. Ontology as a specification of a "conceptualization" 5. Ontology as a representation of a conceptual system via a logical theory <ol style="list-style-type: none"> 5.1 characterized by specific formal properties 5.2 characterized only by its specific purpose 6. Ontology as a the vocabulary used by a logical theory 7. Ontology as a (meta-level) specification of a logical theory |
|--|

위의 해석 중 가장 일반적으로 사용되는 것은 4와 5의 해석이며 대부분의 연구자들이 “an ontology is an explicit specification of a (shared) conceptualization”라는 정의를 이용하고 있으며 덧붙여 시맨틱 웹에서는 특정 목적과 영역의 중요성을 강조하고 있다. 이러한 온톨로지의 해석들을 통해 기존의 시소러스(thesaurus)나 어휘 데이터베이스(lexical database) 의미망 등과 같이 다양한 형태의 어휘 집합(lexical set)들을 온톨로지의 일종으로 판단하여 온톨로지를 구축하고 표현하는 경우도 있다. 즉 온톨로지가 가지고 있는 일반적인 구성 요소인 개념(concept), 관계(relation), 속성(property) 등이 표현되어 있는 어휘 집합(lexical set) 들을 온톨로지로 변환 또는 구축하는 사례를 살펴보면 <표2>와 같이 비슷한 형태로 다양하게 구축되고 있음을 알 수 있다(옥철영 2004).

<표 2> Ontology Example

<p>Ontology Example (in practice)</p>	<ul style="list-style-type: none"> • Simple concept hierarchies • Semantic network • Thesaurus • Frame system • Logical models • Lexical field • Category • Taxonomies on the Web • Catalogs for on-line shopping • Domain-specific standard terminology • Etc.
---	--

시맨틱 웹에서의 웹 온톨로지(web ontology)의 등장으로 인해 기존의 시소러스나 의미망, 어휘 데이터베이스, 표준분류체계(WordNet, UMLS, UNSPSC, RosettaNet, etc.) 등을 RDFs, DAML+OIL, OWL과 같은 웹 온톨로지 언어를 이용하여 온톨로지 로 표현하는 연구개발이 활발하게 진행됨과 동시에 기존의 계층적 구조로 표현된 어휘집합들을 재사용하고 구축시간을 단축하기 위한 구축사례가 다양한 형태로 표현되었다고 할 수 있다. 물론 <표2>와 같은 사례들을 통해 온톨로지는 클래스(class)와 서브클래스(subclass)의 관계 즉 개념간의 계층적 구조를 형성하고 있다는 점과 어떠한 의미적 관계(semantic relation)를 가지고 있다는 점에서 위 사례들이 공통적일 수 있으며 웹 온톨로지 언어의 사용여부와 활용적인 측면에서는 다르다고 할 수 있는 것이다. 앞에서 온톨로지에 대한 해석들과 형태적인 사례를 통해 온톨로지가 어떠한 형식으로 표현되고 있는지를 간략하게 살펴보았다. 그렇다면 온톨로지에 대한 해석만 가지고는 온톨로지를 파악하기가 쉽지 않을 뿐만 아니라 실질적으로 온톨로지를 구축하는 측면에서도 기존의 시소러스 구축원리나 표준 분류 체계 등의 구축원리를 이용한다면 문제가 될 수 있다. 이러한 문제가 발생하는 것은 ‘개념’과 ‘관계’, 그리고 ‘개념화’에 대한 이해 때문이라고 할 수 있다. ‘개념’은 사전적으로 여러 관념 속에서 공통된 요소를 뽑아내어 종합하여 얻은 하나의 관념으로 정의될 수 있으며 ‘개념’은 일반적으로 언어로 표현된다. ‘관계’는 어휘의미론(Lexical Semantics)에서의 의미 관계(semantic relation)와 폭넓은 의미에서의 개념 관계(conceptual relation)가 있다. 특히 시소러스나 WordNet에서 일반적으로 사용 중인 의미 관계는 상하관계, 동의/유의 관계, 부분/전체 관계 등이 보편적으로 사용되며 개념관계는 조금 넓은 의미에서 구성원 관계(has-member), 위치 관계(has-location) 등으로 세분화하여 개념간의 관계를 설정할 수 있다. ‘개념화’는 사물이나 표상을 어떤 성질, 공통성, 본질에 착안하여 그것을 추출하여 파악하는 과정으로서 그 형태적 표현은 분류적 구조(classified structure)와 계층적 구조(hierarchical structure)로 나타날 수 있다(옥철영 2004).

이러한 개념화 양상은 기존의 시소러스와 의미망 등과 같은 계층적 어휘 집합과 온톨로지를 구분하기 어려운 것 중의 하나이다. 온톨로지는 단순히 용어(term)의 체계적 구조화가 아니라 개념들을 특정한 영역의 개념에 대한 정의와 관계, 그리고 개념이 가지는 특수한 속성들로 이루어진 집합체임을 알 수 있다. 또한 온톨로지는 사람과 사람 사이의 원활한 커뮤니케이션처럼 기계와 기계 사이에도 형식적 모델을 통

해 원활하게 커뮤니케이션이 가능하도록 하는 의미적인 구조를 가져야 한다. 이러한 온톨로지의 일반적인 개념 속에 추가로 고려해야 하는 것이 추론과 통합기능이다. 온톨로지가 단순히 계층적 구조와 의미 관계 등으로 구성되어 있다면 온톨로지의 유용성은 시소러스와 의미망, 분류체계, 어휘 데이터베이스 등과 다를 바 없을 것이다.

즉 온톨로지를 통해 사람이 가지는 사고방식과 비슷한 추론 방법으로 새로운 개념을 유추하고 관계를 설정할 수 있어야 하는 것이다. 또한 다양한 온톨로지를 어떻게 통합할 수 있는가도 중요한 문제이다. 시맨틱 웹의 목적인 의미에 따른 자동적인 실행과 추론을 하기 위해서 온톨로지는 <표 3>과 같이 구분할 수 있다(Fensel 2001).

<표 3> 온톨로지의 종류

온톨로지의 종류	설 명
영역 온톨로지 (Domain Ontology)	- 특정 영역에 적합한 지식을 포함하는 온톨로지
메타데이터 온톨로지 (Metadata Ontology)	- 더블린 코어와 같이 온라인 정보원의 내용을 기술하기 위한 어휘집을 제공하는 온톨로지
일반 온톨로지 (Generic or common sense Ontology)	- 세상의 모든 사물에 대한 일반적인 지식 및 개념 (시간, 공간, 상태, 사건 등)을 포함하는 온톨로지 - 공통적인 상위 개념을 표현한 온톨로지 - 온톨로지의 재사용에 유용한 온톨로지 - 온톨로지를 이용한 자연언어 표현의 분석과 온톨로지 생성에 중요
표현 온톨로지 (Representational Ontology)	- 특정 영역에 결부되지 않는 온톨로지으로써 무엇을 표현해야 하는가에 대한 언급 없이 개체를 제공하는 온톨로지
메소드 온톨로지 (Method Ontology)	- 특정 업무에 관한 정보를 제공하는 온톨로지

한편 전산학적인 입장에서는 온톨로지를 다음과 같이 정리할 수 있다(최호섭 외 2004).

- 일정한 체계에 의한 어휘 사전이나 용어의 확보
- 특정 영역(specified domain) 뿐만 아니라 보편 영역(generic domain)의 기본 개념에 대한 정의와 그들 간의 관계에 대한 명세화
- 개념, 관계, 속성들의 유기적인 집합
- 전산적 처리가 가능한 구조화와 구체화
- 공유와 재사용의 가능
- 논리적 추론
- 통합

4. 온톨로지의 역할과 필요성

사람들이 웹을 통해 필요한 정보를 구하고자 할 때는 신속하고 정확하며 유용한 정보검출을 기대한다. 이를 위해서는 이질적인 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 형태로 구성한다. 이러한 협력적 구성은 도메인 지식의 재사용을 가능하게 한다. 온톨로지를 통해 도메인 내의 가설을 명시하여 차후 수정이 용이하도록 하고 도메인 지식을 분석, 이해하기 위해서도 온톨로지는 필요하다. 이러한 역할을 수행하기 위해 제작되는 바람직한 온톨로지는 첫째, 특정 도메인의 전문적인 용어들의 관계를 나타내는 전문가가 사용 가능한 메카니즘이어야 하고 둘째, 전문적인 지식이 없이도 빠르게 정보를 찾을 수 있도록 용어들을 모아놓은 메카니즘이어야 하며 셋째, 컴퓨터 프로그램을 통해 사용자가 원하는 정보를 같은 도메인의 다른 데이터 집합들의 일관성 있는 내용과 정보를 찾아 추출할 수 있도록 자동화된 추론이 가능한 메카니즘과 새로운 도메인과의 합병이 가능한 메카니즘을 포함하는 것이다. 온톨로지를 이용한 지식 모델링이 가져다주는 기술적인 장점은 지식베이스의 일관성을 자동화 하여 검사한다는 것, 특정하게 명시되지 않은 데이터를 유추할 수 있는 추론 메카니즘의 사용 가능성, 사용자의 요구를 광범위하게 수용하고 검색을 용이하게 해주는 것 등을 들 수 있다. 더욱이 지식 베이스를 디자인

함으로써 얻을 수 있는 즉각적인 혜택은 필요로 하는 지식에 대해 설명을 하고 그에 맞추어 지식을 구성하여 구조에 맞추는 과정을 통한 처리과정 자체를 모델링 한다는 데 있다. 온톨로지는 다음과 같은 응용프로그램에서 사용될 수 있다.

- ① 커뮤니티 참조(community reference) : 중립적 온톨로지 작성
- ② 데이터베이스 스키마 정의나 온톨로지 명세화 같은 데이터베이스 주석(annotation)을 위한 공통의 사전(vocabulary) 정의
- ③ 정보 이용을 위한 공통적인 방법의 제공
- ④ 데이터베이스 주석과 기술적인 문서의 이해

첫째, 커뮤니티 참조는 동일한 언어로 작성해야 하고 서로 다른 기종의 환경 하에서 사용하기 위해 다른 언어로의 전환이 가능해야 한다. 즉, 지식을 표현하는데 있어서는 동일한 언어로 표현해야 하며 다른 시스템에 지식을 이식할 때 다른 시스템에 적합한 언어로 번역을 해야 한다.

둘째, 데이터베이스 스키마의 정의나 혹은 온톨로지의 명세화 같은 데이터베이스 주석을 위한 공통의 사전 정의는 지식에 대한 서술과 공유를 위해 동일한 표준 용어으로써 기술한다. 공통의 사전 어휘로써 온톨로지는 지식의 작성, 관리, 공유, 재사용 및 신뢰성에서 장점을 지닌다.

셋째, 정보에 대한 동일한 방법의 접근은 서로 상이한 언어로 작성된 지식을 공유할 수 있게 하는 것이다. 즉, 온톨로지는 공통적으로 공유되는 지식의 용어를 제공함으로써 서로 다른 언어로 작성이 되어 있어도 지식을 제공할 수 있다.

마지막으로 데이터베이스의 주석과 기술적 문서의 이해이다. 온톨로지는 자연어 처리를 제공함으로써 도메인 내의 지식의 연결뿐만 아니라 지식이 문법과 어휘와 같은 언어적인 구조의 연관관계를 이해할 수 있도록 설계되어 있다.

이런 네 가지 형태의 응용프로그램에서 온톨로지의 사용을 필요로 하고 있다.

또한 온톨로지의 사용은 기존의 데이터베이스 기반의 응용프로그램의 기능뿐만 아니라 자연언어 처리와 같은 방법을 사용하여 의미를 파악하여 단순한 키워드 매칭 방법이 아닌 용어들 간의 연관관계를 파악하여 질적인 정보와 지식을 제공할 수 있다. 온톨로지는 간단히 표현하면 단어와 관계들로 구성된 사전으로서 어느 특정 도메인에 관련된 단어들을 계층적 구조로 표현하고 추가적으로 이를 확장할 수 있는 추

론규칙을 포함한다. 온톨로지의 역할 중 하나는 서로 다른 데이터베이스가 같은 개념에 대해 서로 다른 단어나 식별자를 사용할 경우에 이를 해결해 주는 데에 있다. 예를 들어 주소를 포함하는 두 데이터베이스에서 postal code와 zip code는 같은 것을 의미한다. 이 두 데이터베이스의 정보를 비교하거나 통합하려는 프로그램이 있다면 이 두 단어가 같은 것을 지칭한다는 사실을 알아야 하며 이것이 바로 온톨로지를 통해서 이루어진다. 온톨로지는 웹 기반의 지식처리나 응용프로그램 사이의 지식공유, 재사용들을 가능하게 하는 아주 중요한 요소로 자리 잡고 있다. 온톨로지에는 계층분류(taxonomy)와 추론규칙(inference rule)에 대한 정의가 포함된다. 계층분류는 객체의 클래스(class)와 서브클래스(subclass), 그들 간의 관계(relationship)를 정의한다.

예를 들어 주소를 뜻하는 address는 위치를 뜻하는 location의 서브 타입이므로 address는 location의 서브클래스로 정의될 수 있고 city codes는 location에만 적용될 수 있으므로 city codes의 대상은 반드시 location 클래스의 객체여야 한다는 제약조건이 관계로 정의될 수 있다. 추론규칙은 프로그램이 새로운 사실을 자동으로 추출하거나 제약조건에 맞지 않는 오류를 찾아내는데 이용된다. 온톨로지를 표현하기 위해 스키마와 구문구조 등을 정의한 언어가 온톨로지 언어이며 현재 DAML+OIL, OWL, Ontolingua 같은 온톨로지 언어가 정의되었다.

5. 시맨틱 웹에서 제공하는 온톨로지의 특징

첫째, 온톨로지의 작성 주체로 국가나 대규모 단체와 같은 집단뿐만 아니라 개인이나 그룹 같은 소규모의 다수 집단을 지향한다. 현재의 웹도 대규모 단체뿐만 아니라 개인을 포함한 다수의 소규모 집단에 의해서 자발적으로 수많은 문서가 작성되고 있다. 시맨틱 웹의 온톨로지도 이와 같이 다수의 소속 집단에 의한 자발적 온톨로지 작성을 전망하고 있다(Fensel, D. 2001).

둘째, 온톨로지의 확장과 수정을 전제로 작성된다.

Yahoo의 경우를 보면 사용자는 이미 정해진 카테고리를 확장하거나 수정하는 것을 전제로 만들어진 것이 아니다. 시맨틱 웹에서는 이미 정의된 다른 온톨로지를 검색하여 이를 수정하고 확장하여 개별적 목적에 맞게 진화시키는 기능을 추구한다. 또한 하나 이상의 온톨로지를 결합하여 새로운 온톨로지를 만들어 낼 수도 있다. 이렇게

수정되거나 만들어진 온톨로지는 다시 웹을 통하여 다른 사용자가 검색하고 내려받아 사용할 수 있게 된다(James Hendler 2001).

셋째 사용자에게 맞게 특화되거나 전문분야에 적합하게 세분화되고 경우에 따라서는 상충되는 온톨로지의 모임을 허용한다. 가장 보편적인 주제와 분류를 위해 만들어진 Yahoo와 같은 독점적이고 일관성 있는 카테고리와는 달리 전문 분야 또는 극히 개인적인 시각을 반영하는 다수의, 경우에 따라서는 서로 상충되는 온톨로지들이 자연스럽게 공존하며 산재하게 될 것이다. 시맨틱 웹 이전의 CYC나 WordNet 등이 모순이 없는 일체의 온톨로지를 추구하는 것과는 대조적이다.

넷째 시맨틱 웹의 목적과 부합되도록 기계에 의한 처리를 전제로 만들어 진다. 다른 온톨로지도 물론 컴퓨터에 의한 처리를 전제로 하겠지만 시맨틱 웹의 온톨로지는 지능화된 에이전트 간의 주요한 의사소통 수단으로 이용된다. 이를 위해 웹의 표준 언어인 XML을 근간으로 개발이 이루어지고 있다. XML은 기본적으로 계층구조를 순차적으로 표현하는 문법을 제공한다. 이를 바탕으로 에이전트에 의한 의미처리가 수월하도록 데이터 모델과 문법을 추가한 RDF가 시맨틱 웹 온톨로지의 바탕이 된다.

다섯째, 온톨로지의 작성이 자발적, 무의식적으로 이루어진다. 현재의 웹 문서도 초기에는 HTML 태그를 직접 이용하는 전문가에 의해 제한적으로 만들어졌으나 점차 이러한 작업을 지원 하는 소프트웨어의 개발로 인하여 단순한 문서 작업만으로도 손쉽게 작성할 수 있게 되었다. 사용자는 HTML 구문에 대해 전혀 이해할 필요가 없고 또한 이미 만들어진 웹 문서를 수정하여 새로운 문서를 작성할 수 있게 되었다. 시맨틱 웹의 온톨로지도 같은 과정을 거칠 것으로 전망된다. 사용자는 온톨로지에 대한 이해가 없어도 일반 문서 작업과 유사한 작업을 하면서 자연스럽게 이미 만들어진 온톨로지를 이용하고 수정하고 확장하는 작업을 하게 될 것이다. 사용자는 문서 작업 소프트웨어의 일부로 편집된 기능을 써서 온톨로지를 사용하게 될 것이다.

마지막으로 온톨로지의 추론 기능이 제공된다. 온톨로지의 중요한 역할중의 하나는 개념과 개념 간의 관계를 정의하는 것이다. 대표적으로는 상속과 부분-전체 관계를 들 수 있다. 이러한 관계와 description logic을 이용하면 개념간의 관계를 추론을 통하여 제시하거나 이에 대한 질의에 답할 수 있게 된다. 시맨틱 웹에서는 자연어 위주의 기존 웹 문서와 달리 컴퓨터가 해석하기 쉽도록 의미를 부여한 계층을 가지고 있기 때문에 자동화된 에이전트나 정교한 검색엔진들이 부여된 의미를 이용하여 고수

준의 자동화와 지능화를 이룰 수 있게 된다. 온톨로지를 핵심으로 의미를 부여하고자 하는 시맨틱 웹은 현재의 ‘링크의 그물’인 웹을 ‘의미의 그물’로 변환시킬 것이다 (Fensel, D. 2003).

6. 온톨로지 언어

1) 메타데이터 스키마

일반적인 의미로 메타데이터는 ‘데이터에 관한 데이터’로 정의할 수 있는데 데이터 형식에 관한 내용을 기술한다. 메타데이터는 그 자체가 또한 데이터이다. 웹에서 메타데이터의 목적은 정보 검색의 처리 과정을 줄여주고 사용자가 원하지 않는 데이터를 미리 걸러주며 관련성이 많은 정보의 발견 가능성을 높여줌으로써 정보 검색을 향상시키기 위한 것이다. HTML 페이지 내의 태그도 문서를 설명해 주는 메타데이터이다. 메타데이터에는 웹 페이지의 작성자, 버전, 내용에 대한 목록 정보를 표시할 수 있으며 웹 페이지 상호 간의 관련성을 나타낼 수 있고 또한 배포권과 프라이버시 코드 등과 같은 사회적 정보도 표시할 수 있다.

정보의 공유 및 교환, 접근 등에 있어 각기 다른 개념과 용어 사용으로 인한 의미적 다양성 문제를 해결하는 한 가지 방법은 관련된 시스템 간에 특정 영역에 적용되는 메타데이터를 공유하는 것이다. 메타데이터 스키마는 정보를 기술하는데 필요한 요소 이름의 집합(vocabulary)과 정보를 구조화하는 의미(semantics)의 집합으로 구성된다. 다시 말해서 메타데이터 스키마는 여러 메타데이터 형식에서 규정하고 있는 기술요소(속성)들의 의미, 특성, 속성간의 관계, 속성 값의 제한사항, 스키마 간의 연계, 정보자원의 유형 등을 정의한다(김세정 1999).

메타데이터는 관계성에 관한 것이다. 관계성이란 정보 리소스에 대해 기술하는 것을 의미한다. 메타데이터로 데이터를 기술할 때 정해진 방식에 제한 받지 않는다.

시맨틱 웹에서 메타데이터 사용의 다양성은 중앙집중방식을 지양하고 웹의 분산성을 증진하려는 웹의 기본정신과 관련이 있다. 앞서 시맨틱 웹에서의 의미는 정보 리소스들 사이의 관계성에 있다고 언급했는데 관계성을 다른 말로 표현하면 연결성이라 할 수 있다. 웹의 시작은 이러한 연결성을 통해 구현됐고 이로 인해 웹은 분산적 공간이 됐던 것이다. 전통적 의미의 월드와이드웹(WWW)에서의 연결성은 하이퍼텍

스트 링크를 통한 가상공간에서의 위치적 연결을 의미하지만 시맨틱 웹에서는 정보 자원 사이의 의미적 연결을 더 강조하고 있다. 사실 URL은 링크를 위한 리소스의 위치적 정보를 표시하는데 비해 좀 더 보편적 개념인 URI는 의미적 정보의 인식을 위한 가장 기본적 수단인 지칭(referring)과 정체성(identity)과 관련된 단어인 인식자(identifier)를 사용한다. 정보 리소스 사이의 관계를 컴퓨터가 처리할 수 있는 방식과 언어로 표현하는 것을 지식표현(knowledge representation)방법이라고 한다. 이 방법은 인공지능 분야에서 다양한 목적의 전문가 시스템(Expert System) 개발에 널리 이용되었다.

시맨틱 웹은 웹 페이지의 문서별로 서로 다른 의미체계를 구축할 수 있지만 서로 다른 웹 페이지들에서 정보 리소스들 사이에 의미적 연관성을 지을 수도 있으며 또한 이미 구축돼 있는 지식베이스에 있는 정보를 메타데이터를 이용해 가져올 수도 있다. 이것을 시맨틱 웹에서의 상호 운용성(interoperability)이라고 한다. 시맨틱 웹은 향상된 수준의 상호 운용성이 확립될 때 성립될 수 있다. 문서의 구문형식뿐 아니라 문서의 의미적 내용에 대한 표준이 정의되어야 한다. 이러한 의미적 상호 운용성은 XML(S) 및 RDF(S)를 표준화하려는 웹 컨소시엄(W3C) 등에 의해 연구되고 있다.

웹이 어떠한 형식의 데이터라도 처리할 수 있는 강력한 매체가 되기 위해서는 어떠한 데이터라도 표현할 수 있는 표현능력과 구문적 상호 운용성을 지지할 수 있는 능력을 갖추어야 할 뿐만 아니라 나아가 여러 응용프로그램에서 사용되는 다른 용어간에 의미적 관계를 매핑할 수 있어야 한다. 즉, 앞으로 웹의 발전 방향에서 의미적 상호 운용성이 차지하는 비중이 커진 것이다. 달리 말하면 시맨틱 웹의 궁극적 목표는 의미사용에 대한 분산성을 증진시키는 것인데 정보를 사용하는 사람들이 사물에 대한 동일한 확정적 의미를 가질 필요가 없다는 뜻이다.

2) 시맨틱 웹에서의 XML의 역할

메타언어로서 SGML은 문서와 데이터를 기술하기 위한 문법 규칙의 집합으로서 HTML과 같은 언어의 태그들의 의미와 사용 규칙을 정의 내려주는 언어이다. 웹의 활용범위가 다양해지고 복잡해지면서 단순히 텍스트나 이미지를 보기 위한 목적을 넘어서는 새로운 형태의 마크업 언어가 필요하게 됐다. 이런 의미에서 SGML은 하나의 대안일 수 있었으나 기능이 너무 복잡하기 때문에 SGML의 축소 형태인 XML

(eXtensible Markup Language)이 개발되었다. XML은 SGML의 subset으로 구성된 마크업 언어로서 시맨틱 웹이라는 개념과는 별개로 HTML의 비구조성을 극복하기 위해 제안되었다. HTML에 비해서 XML은 잘 정의된 구조화 문서(well-structured documents)를 작성할 수 있도록 해준다. 즉, 요소(element)라고 불리는 시작 태그와 종료 태그가 반드시 쌍으로 존재해야 한다는 것과 중첩구조가 반드시 지켜져야 한다는 등의 제약조건이 만족되어야 한다. XML은 컴퓨터 프로그램에 의한 관독이 용이한 문서작성 규정이다. 마크업이란 문서내용의 역할을 지정하는 추가 정보를 지정한다는 것을 뜻하며 문서에 논리적 구조를 부여하고 정보 자체를 기술하는 역할을 한다. 시맨틱 웹과 관련된 XML의 역할은 이러한 구조화된 문서의 생성을 이끌어내는 것도 있지만 태그의 이름을 사용자가 자유롭게 정의할 수 있기 때문에 의미정보를 나타낼 수 있는 태그이름을 사용할 수 있다는 것이 더 큰 비중을 차지한다(Decker, S. 외 2000).

HTML이 웹 문서를 화면에 보여주기(displaying) 위한 언어라면 XML은 웹에서 데이터를 기술(describing)하기 위한 언어이다. XML에서 정의 내려지는 태그의 의미는 사용자가 이해하기 쉽게 결정할 수 있다. 그러나 XML은 문서 요소들 간의 구조적 정의만을 제공해 주기 때문에 정보 리소스들 사이의 의미적 관계를 정의 내려 주지는 못한다. 때문에 이러한 XML 표현 방법이 시맨틱 웹을 달성하기엔 부족한 점이 있다. 예를 들어 메모의 작성을 표현하기 위한 아래의 간단한 XML 문서의 태그를 보면 메모를 보낸 사람(from), 받는 사람(to), 제목(heading), 메모 내용(body) 등의 의미를 파악할 수 있다.

```

<note>
<date>10/10/2004</date>
<to>Kim</to>
<from>Kang</from>
<heading>Reminder</heading>
<body>Happy birthday to you!</body>
</note>

```

위에서 메모의 제목을 <heading>으로 하였지만 같은 의미의 <subject>라는 다른

태그를 이용할 수도 있다. 따라서 상호 운용성(interoperability)을 위해서는 이 두 태그 이름이 같은 의미를 가진다는 것을 따로 표현해야 한다. 이것이 바로 스키마 또는 온톨로지의 필요성이다. 온톨로지가 잘 정의되지 않은 경우에는 문서의 공유나 확장이 어려울 수 있다. 또한 위의 메모 내용에서 날짜를 연월일로 구분하여 표현하면 아래처럼 되어 이 두 가지 문서 내용이 같은 것이지만 이것을 에이전트 프로그램이 파악하기는 매우 어렵다.

```

<note>
<date>
  <day>10</day>
  <month>10</month>
  <year>20004</year>
</date>
<to>Kim</to>
<from>Kang</from>
<heading>Reminder</heading>
<body>Happy birthday to you!</body>
</note>

```

XML은 문법을 정의하기 위한 형식이어서 문법적으로 정의될 수 있다면 어떠한 것이라도 표현할 수 있으며 파서를 통해 스키마에서 사용되는 구문들을 이해할 수 있게 된다. 그러나 XML은 의미적 상호 운용성을 만족시키는데 문제가 있는데, XML이 문서의 구조와 형식을 정의하는 반면 문서에 포함된 데이터의 내용은 해석하지 않기 때문이다.

XML은 임의의 문서 구조를 지정할 수 있는 반면에 문서 구조가 갖는 의미에 대해서는 정의하지 않으며 따라서 표현한 문서의 의미를 해석하는 역할은 하지 않는다. RDF는 이러한 문서의 의미를 지정하는 기본적인 역할을 하며 상호운용의 측면에서 RDF의 역할이 중요하게 된다(Hjelm, J. 2001).

3)시맨틱 웹에서의 RDF의 역할

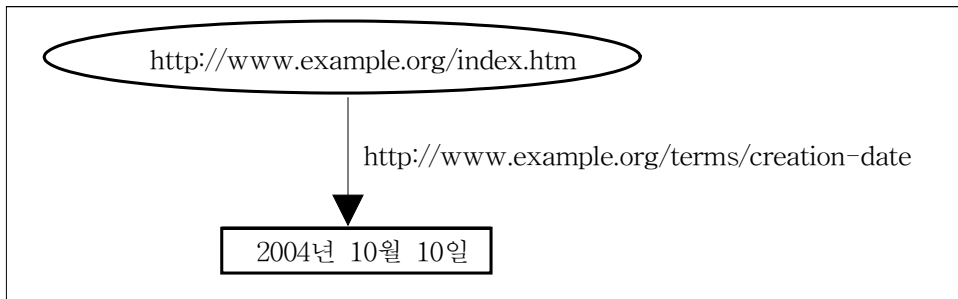
RDF는 W3C의 가장 기본적인 시맨틱 웹 언어로서 웹에 있는 자원(resource)에 관한 메타정보를 표현하기 위한 언어이다. 특히 웹 자원을 표현하는데 기본이 되는 제목, 저자, 최종 수정일, 저작권 같은 웹 문서에 관한 메타데이터를 표현할 목적으로 개발되었으나 웹 자원의 개념을 웹상에서 다른 것과 구별하여 식별할 수 있는 대상으로 일반화 하면 다른 목적으로도 RDF를 활용할 수 있다. RDF는 XML의 문제점을 해결하고 시맨틱에 초점을 맞추기 위해 제시된 기반구조로써 RDF의 근본을 이루는 개념은 메타데이터이다(Lassila, O. 1998).

전술한 바와 같이 메타데이터는 데이터에 대한 데이터, 즉 어떤 객체나 리소스에 대한 서술적인 정보를 말한다. 웹 문서에 대한 메타데이터라고 한다면 그 문서의 주제, 요약, 저자, 작성일과 같이 해당 문서의 외적 요소들을 망라한다고 볼 수 있다. RDF는 구조화된 메타데이터의 생성, 교환, 재사용 등을 가능하게 해주는 기반구조로 되어 있으며 보다 정확한 검색 엔진의 성능을 제공하는 리소스의 발견(resource discovery), 특정 웹 페이지나 디지털 라이브러리의 내용과 관계를 기술하는 문서분류(cataloging), 지식 공유와 교환이 가능한 지능형 소프트웨어 에이전트(intelligent software agent), 문서내용 등급표시(content rating)나 사용자의 개인 선호도 표현 등의 용도로 사용된다. RDF 모델은 리소스(Resource), 특성(Property), 서술문(Statement)의 개념으로 구성된다. 웹 페이지나 웹 사이트 등의 모든 사물(thing)은 리소스로 표현되고 각 리소스의 특성이나 다른 리소스와의 관계 등을 특성으로 나타낸다(Fensel, D. 외 2000).

어떤 리소스의 한 특성에 대한 값을 나타내는 것이 서술문이며 이것이 RDF 문의 기본 단위가 된다. RDF의 서술문은 그래프 모델로 나타낼 수도 있고 XML로 표현할 수도 있다. RDF를 XML로 표현한 것을 Serialization이라고 한다. RDF는 기본적으로 세 개의 정보를 지닌 쌍들을 정의한다. 이 세 개의 정보는 일반 문장의 주어, 동사(서술어), 목적어에 해당하는 것으로 사람이나 웹 문서 등 특정 대상(object)이 특정 속성(attribute)에 대하여 특정 값(value)을 가지고 있는 상태를 표현한다. RDF는 XML의 메타언어 기능을 사용하여 표현된다. 주어에 해당하는 대상과 목적어에 해당하는 값, 그리고 동사나 서술어에 해당하는 속성들을 모두 URI로 지정할 수 있다. 따라서 사용자는 새로운 개념이나 동사를 URI를 써서 손쉽게 정의할 수 있는 것이다.

RDF Primer에 나오는 예로 “<http://www.example.org/index.htm>이 2004년 10 월

10일에 만들어 졌다”라는 문장을 RDF로 표현하는 경우를 보면 다음과 같다. “http://www.example.org/index.htm은 creation-date라는 속성을 가지고 있으며 그 속성의 값은 2004년 10월 10일이다”라고 표현될 수 있는데 이를 RDF의 도식화 규칙에 의해 다음과 같이 나타낼 수 있다.



위의 문장 구성요소를 분석하면 다음과 같다.

주어(자원)	http://www.example.org
서술어(속성)	http://www.example.org/terms/creation-date
목적어(값)	“2004년 10월10일”

이를 RDF를 써서 다음과 같이 표현할 수 있다.

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:exterm="http://www.example.org/terms/">
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <exterm:creation-date>2004년10월10일</creation-date>
  </rdf:Description>
</rdf:RDF>
  
```

위의 RDF 문서는 앞서 표현한 도식을 XML로 표현한 것이다. RDF는 이러한 표현에 필요한 잘 정의된 구문을 제공한다. 위 문서의 첫 번째 줄은 이 문서의 내용이 XML 되어 있다는 것과 XML 버전을 선언한다. 다음의 rdf:RDF 엘리먼트는 </rdf:RDF>로 끝나는 부분까지가 RDF 표현식이란 것을 나타내며 동시에 xmlns를 써서 XML의 name space를 지정한다.

다시말해 rdf:로 시작하는 태그들은 URI <http://www.w3.org/1999/02/22-rdf-syntax-ns#>로 지정된 name space에, exterms:로 시작하는 태그들은 <http://www.example.org/terms/>에 속한 것들이라는 것을 나타낸다. rdf:Description 태그는 자원의 서술이 시작됨을 알리고 서술 대상이 about에 지정한 주어임을 나타낸다. 다음 줄에서는 이 주어의 속성인 exterms:creation-date가 목적으로 표현된 “2004년10월10일”이란 값을 가짐을 나타낸다. RDF는 이렇게 자원-속성-값의 표현을 반복하거나 중첩하여 자원에 대한 서술을 하게 된다. 아래의 RDF 문장은 <http://www.w3.org>라는 리소스의 책임기관(Publisher), 제목(Title), 작성일(Date)의 세 가지 특성에 대한 정보를 표현하고 있다.

```

<rdf:RDF>
  <rdf:Description about="http://www.w3.org">
    <s:Publisher>World Wide Web Consortium</s:Publisher>
    <s:Title>W3C Home Page</s:Title>
    <s>Date>1998-10-03T02:27</s>Date>
  </rdf:Description>
</rdf:RDF>

```



RDF 모델은 XML이 가지고 있던 문제점을 다음과 같이 해결하고 있다. 즉 의미가 리소스와 그 특성 값으로 표현되므로 같은 내용(의미)에 대해서는 해석(interpretation)이 하나로만 귀결된다는 것이다. 달리 표현하면 XML에서와 같이 서로 다른 구조를 가진 여러 가지 표현방법이 존재하지 않기 때문에 문서의 내용에 대한 이해가 쉽다. 하지만 RDF에서도 XML의 문제점 중 하나였던 태그 이름의 중첩성과 모호성은 여전히 존재한다. 즉 서로 다른 태그이지만 실제로는 같은 의미일 수 있고 반대로 같은 태그이지만 사용자에게 따라서 다른 의미로 쓰일 수도 있다. 이 문제는 XML에서와 같이 온톨로지 개념으로 해결해야 한다. RDF에서는 온톨로지와 유사한 RDF 스키마가 존재한다. RDF 스키마는 특성에 대한 정의나 사용상의 제약사항을 기술한 것이다. 따라서 RDF의 의미는 이 스키마를 통해서 표현된다고 보면 된다. 스키마는 사전과 비슷한 개념인데 RDF 문을 구성하는 단어(term)를 정의하고 그 단어들에 대한 세부적인 의미를 기술하고 있다. 온톨로지는 RDF 스키마와 유사하지만 좀 더 일반적이고 확장된 개념이다. 앞의 <그림 2> ‘시멘틱 웹의 계층적 구조’에서 보듯이 RDF

문서의 문법은 XML에 기초한다. RDF는 객체지향적 접근의 지식표현 방식을 취하며 세 개의 요소로 이루어 졌다. 즉, 객체(object), 속성(attribute), 값(value)의 구조를 갖고 있으며 A(O,V)로 표현된다. 그러나 일반적인 객체지향적 혹은 프레임 방식과는 달리 RDF는 객체중심이 아니라 속성중심 (property-centered)의 구조를 가지고 있다.

RDF 스키마는 XML 문서의 요소들 사이의 구조를 정의해 주는 XML 스키마와는 달리 RDF 문장에서 쓰인 어휘들의 정의를 위한 메타언어로서의 역할을 한다. RDF 스키마에는 미리 정해진 몇 개의 어휘들이 있어서 RDF 문장에서 쓰이는 어휘 사이의 관계를 의미적으로 정의 내리는데 사용한다. 예를 들면 RDF 문장에서 각각의 객체는 type 속성을 갖는 하나 이상의 클래스의 인스턴스로 정의 내려진다. 또한 subClassOf 속성은 클래스 사이의 계층적 구조를 나타내는데 사용된다.

4)DAML+OIL

DAML+OIL은 웹 온톨로지 언어로서 DAML(DARPA Agent Markup Language) 프로그램의 DAML-ONT(<http://www.daml.org/2000/10/daml-ont.html>)와 유럽에서 개발된 OIL(Ontology Inference Layer)의 결합을 통하여 만들어 졌다. DAML은 웹 페이지에 존재하는 정보를 컴퓨터가 읽고 이해할 수 있도록 하기 위해서 XML 기술을 기반으로 고안된 의미론적 언어로서 문맥정의언어(context definition language)라고 정의된다(Dieter Fensel 외 2001).

DAML은 서로 관련 있는 웹 페이지가 동일한 의미의 태그를 상이하게 표현하거나 또는 한 태그가 서로 다른 의미로 사용되어 발생하는 의미론적 장벽(semantic barrier)을 해결하기 위한 수단으로 온톨로지에 의한 연결(ontological connection)을 제공한다. 점에서 평범한 메타 태그와 구별된다.

DAML 이전에 이미 OIL이라고 하는 온톨로지를 정의하기 위한 언어가 개발되었으나 DARPA(Defense Advanced Research Project Agency)에 속한 많은 연구원들은 OIL 만으로는 온톨로지를 표현하기에 부족하며 이를 보완할 새로운 언어가 필요하다는 사실을 인식하였다. 이로 인해 OIL이 RDF나 RDFS와 가지고 있었던 관계보다 의미론적으로 더욱 밀접하게 결합된 DAML+OIL이라는 언어가 탄생하게 되었다. DAML+OIL은 컴퓨터가 쉽게 읽을 수 있고, 쉽게 이해할 수 있도록 정보를 표시할 수 있으며 온톨로지를 생성할 수 있다. 의미론적 웹(Web)을 만들고 이를 프로그램에

자유롭게 사용하기 위해서는 DAML+OIL은 그 핵심 기술이라고 할 수 있다.

DAML+OIL의 목적은 웹 페이지, 데이터베이스, 프로그램들 간의 의미론적 상호 운영을 가능하게 하는데 있다. DAML+OIL은 다른 메타 태그와는 다르게 서로 관련이 있는 웹 페이지들 간의 온톨로지를 제공함으로써 기존에 장애였던 의미론적 장벽을 극복할 수 있도록 하고 있다. 그 결과로 단순히 웹 페이지를 가져와 디스플레이 하던 기존의 방식과는 다르게 컴퓨터에 의해 읽혀지고 처리될 수 있는 웹 페이지를 가질 수 있다. 온톨로지를 표현하기 위해 스키마와 구문구조 등을 정의한 언어가 온톨로지 언어(ontology language)인데 현재 DAML+OIL, OWL, Ontolingua 같은 온톨로지 언어가 정의되었다. 이 중에서 W3C에서 표준안으로 제시한 DAML+OIL은 웹 리소스에 대한 시맨틱 마크업 언어이며 W3C의 RDF와 RDF 스키마 표준에 기반을 두고 이들을 확장한 프레임 기반의 온톨로지 표현 언어이다. 기본적으로 DAML+OIL로 표현된 온톨로지는 크게 클래스 요소(class element)와 특성 요소(property element)로 구성된다. DAML+OIL 웹 온톨로지 언어는 현재 W3C에 의해서 OWL 웹 온톨로지 언어로 계승 발전되고 있다. OWL은 DAML+OIL의 네임스페이스와 속성, 클래스, 이름 등을 변경하고 RDF 및 RDF Schema의 변화를 수용하였다. 온톨로지 언어로서 DAML+OIL은 관심영역(domain)의 구조를 서술하기 위한 목적을 갖는다. 이러한 구조는 객체 지향적인 방법으로 클래스(class)와 속성(property)을 써서 표현된다. 온톨로지는 클래스와 속성의 성격을 서술한 공리(axiom)의 집합으로 구성된다. 자원(resource)이 특정 DAML+OIL 클래스의 인스턴스(instance) 이거나 특정 속성을 갖는 것을 표현할 때는 이러한 목적에 적합한 RDF를 사용한다.

5)OWL(Web Ontology Language)

OWL은 DAML+OIL에 기반을 둔 온톨로지 구축 경험을 토대로 개념의 일관성을 확보하여 class와 property의 개념 및 그들 사이의 관계가 보다 명료하게 정의되도록 정리한 온톨로지 언어로써 차세대 온톨로지 언어로 광범위하게 활용될 것으로 보인다. DAML+OIL과 같은 언어는 표현력에서 우수하지만 처음 접근하는 이용자, 개발자 및 개발도구 지원 등에 용이하지 않은 측면이 있어 이를 개선하여 보편적인 이용자를 확보하기 위해 좀더 간결하고 사용하기 쉬운 언어가 필요하게 되었다.

OWL 웹 온톨로지 언어는 단지 사람에게 정보를 표시하는데 그치지 않고 정보의

내용을 직접 처리할 수 있는 애플리케이션을 구현하는데 활용될 수 있도록 설계된 언어이다. OWL은 풍부한 어휘(vocabulary)와 형식적 의미론(formal semantics)을 포함하고 있기 때문에 기계 해석이 가능한 웹 콘텐츠를 제작하는데 있어 XML, RDF 및 RDF 스키마 보다 뛰어나다. OWL은 표현력이 서로 다른 세 개의 하위 언어인 OWL Lite, OWL DL, OWL Full로 구성되어 있으며 후자로 갈수록 표현력이 더 크다(<http://www.w3c.org/TR/2004/REC-owl-features-20040210/>).

OWL의 서브 셋으로 제안된 OWL Lite는 OWL과 DAML+OIL의 공통적이고 유용한 부분들을 간추려 만들어 졌다. 기능적인 면에서 볼 때 웹 응용프로그램을 지원하기 위해 간단하면서도 RDFS에 비해 풍부한 표현력을 가지는 언어라고 할 수 있다 (McGuinness, D 외 2003).

OWL Lite는 시소러스에 접근이 용이하고 단순한 면을 강조하여 웹 응용에 강점을 가지고 있으며 OWL DL(Description Logic)은 Lite보다 좀더 논리적인 표현을 위한 온톨로지이다.

마지막으로 OWL Full은 표현력에 있어서 가장 풍부하며 RDF의 자유로운 구문을 모두 허용하고 있다. 즉 DL과 Lite의 모든 기능을 포함하는 관계이며 유효성 및 호환성에 있어서는 가장 완벽하다. OWL은 문서에 포함된 정보를 애플리케이션을 이용하여 자동 처리하고자 할 때 활용하는 언어이다. OWL을 이용하면 임의의 어휘를 구성하는 용어(term)의 의미와 용어들 간의 관계를 명시적으로 표현할 수 있다.

이와 같이 용어와 용어들 간의 관계를 표현한 것을 온톨로지라 한다. OWL은 XML, RDF, RDF-S 보다 더 많은 의미 표현 수단을 제공하므로, 웹상에서 기계가 해석할 수 있는 콘텐츠를 작성하는데 있어 이들 언어보다 뛰어나다. OWL 온톨로지는 클래스와 속성, 그리고 제약사항(constraints)의 집합으로 다음과 같은 요소들을 포함한다(Smith, M. K. 외 2003).

- 클래스들 사이의 분류체계
- 클래스의 요소가 되는 속성의 값
- 클래스 요소간의 관계에 관한 기술
- 클래스와 속성들의 인스턴스(사례)

시맨틱 웹을 위한 온톨로지 언어의 가장 최근 동향으로 제시되고 있는 OWL은 DAML의 취약점으로 지적되어 온 용어 의미의 혼동을 보완하여 이전 버전의 온톨로지인 DAML과는 거의 완벽한 호환성을 유지한다. OWL을 정립하는 과정에서 발생한 주요 변동사항으로는 DAML 네임스페이스가 OWL 네임스페이스로 변환되었고 일관성을 위해 DAML의 여러 클래스와 속성명들이 OWL 명칭으로 바뀌었으며 DAML에 결여된 대칭적 속성(Symmetric Property)을 추가하고 RDF(S)와 동일한 의미를 갖고 있는 요소(daml:domain, daml:range)들을 삭제하고 DAML의 disjoint With 등 몇 가지 요소를 새롭게 대체한 점 등이다(오삼균 2002).

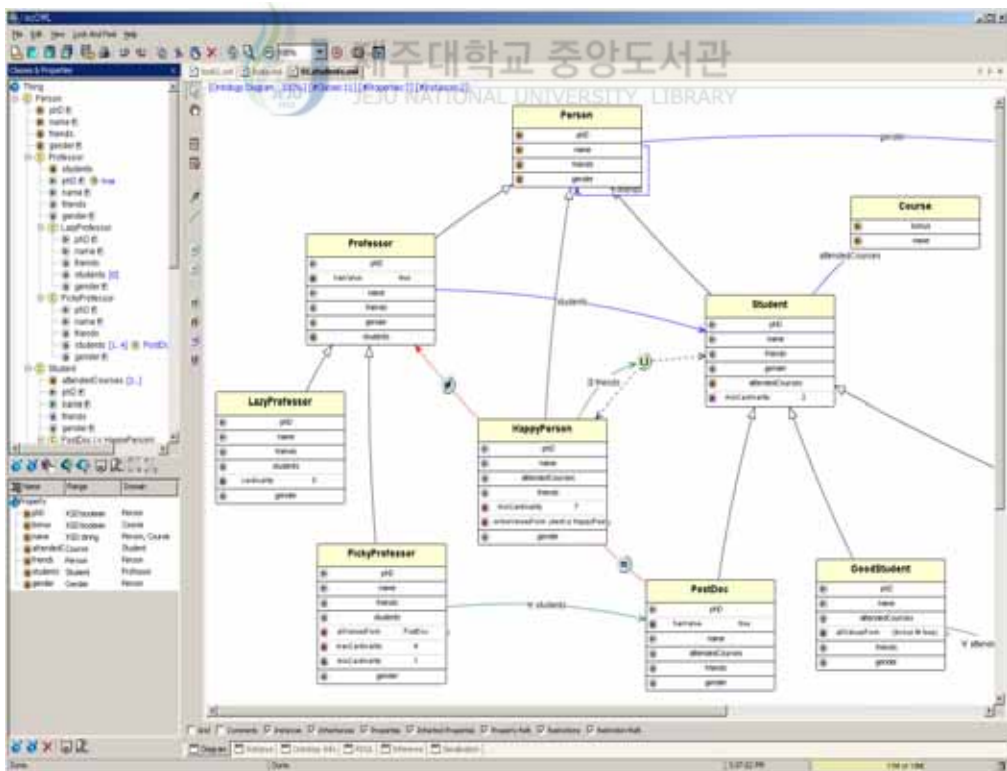
6)ezOWL

컴퓨터가 인터넷상의 웹 문서의 내용을 이해하고 자동으로 처리하는 차세대 지능형 웹 기술인 시맨틱 웹(Semantic Web) 기술 개발에 있어 가장 핵심이 되는 온톨로지(용어 사이의 관계를 집대성한 DB) 저작 툴인 ‘이지아울(ezOWL)’이 한국전자통신연구원(ETRI, www.etri.re.kr)에 의해 2003년 9월에 개발되었다. 이지아울은 컴퓨터가 웹 문서의 용어들을 이해할 수 있게 함으로써 이러한 작업까지 컴퓨터가 직접 수행하도록 하는 기술이 바로 지능형 웹(시맨틱웹, Semantic Web) 기술이다. 예를 들어 지능형 웹에 연결된 컴퓨터에 환자의 증상을 입력하면 정확한 병명과 처방까지 알려주는 식의 완전 자동화가 가능해진다. 이를 위해서 컴퓨터가 웹 문서의 의미를 인식할 수 있도록 각종 용어와 동의어, 반의어, 포함관계 등 용어들 간의 관계를 집대성한 데이터베이스를 구축해야 하는데 이를 온톨로지(Ontology)라고 전문한 바 있다. ‘온톨로지’는 컴퓨터를 지능화시킬 수 있어 시맨틱 웹뿐만 아니라 지능형 로봇 시스템을 위한 지능형서비스, 지능형 e-비즈니스, 의료정보화, 바이오 인포매틱스 등의 다양한 분야에서 활용될 수 있다. 수십만, 수백만 개의 용어들을 일일이 타이핑하는 수작업에 의해 이루어져 오던 온톨로지 구축작업의 효율을 높이기 위한 S/W들이 소수 개발되었으며 이 중 스탠포드 대학의 의료정보 기관인 SMI (Stanford Medical Informatics)에서 개발한 ‘protege-2000’이라는 온톨로지 구축 S/W가 가장 널리 쓰이고 있었다. 그러나 이 S/W는 온톨로지에 대한 정확한 이해와 지식 없이는 사용이 불가능하였다. ETRI가 개발한 ‘이지아울(ezOWL)’은 복잡한 온톨로지를 그래픽 다이어그램으로 생성 및 편집할 수 있어 온톨로지 형식에 대한 지식 없이도 그 내용을 쉽

게 파악하고 제작할 수 있다. 'ezOWL'은 독자적인 S/W이지만 현재 가장 널리 보급되어 있는 protege-2000과 플러그인 (Plug-in)하여 사용할 수도 있게 함으로써 기존의 온톨로지 개발자들이 쉽게 접할 수 있게 배려하였다. ezOWL은 W3C(국제 웹 표준화기구)에서 표준화 작업 중인 차세대 웹 온톨로지 언어 OWL (Web Ontology Language)을 사용하고 있으며 기존 언어인 DAML+OIL, RDF /RDFS도 지원하고 있어 호환성도 뛰어나다.

컴퓨터는 상상력이 없기 때문에 '아름답다', '성공했다', '망했다' 등 감정이나 상상력을 수반하는 단어를 직접 가르치는 것은 불가능하다. 대신 컴퓨터는 상반(\leftrightarrow), 유사(\approx), 포함 관계(\in 또는 \subset) 등의 논리 관계를 인식할 수 있다는 점에 착안하여 고안된 것이 '온톨로지(Ontology)'라는 것으로서 단어들 간의 관계를 입력함으로써 컴퓨터에게 말을 가르친 효과를 낼 수 있게 하는 것이다.

<그림 3> ETRI의 'ezOWL'



이상을 정리해 보면 XML은 문서를 구조적으로 기술할 수 있는 문법을 제공한다. 그러나 문서의 의미를 제한할 수 있는 수단을 제공하지는 않는다. XML 스키마는 XML 문서의 구조를 기술할 수 있는 언어이다. 또한 XML 스키마는 XML에 데이터 타입을 추가한다. RDF는 객체(자원) 및 객체 사이의 관계를 표현하는 데이터 모델이다. RDF는 이 데이터 모델에 대한 간단한 의미론(semantics)을 제공한다. 또한 이 데이터 모델은 XML 구문으로 표현될 수 있다. RDF 스키마는 RDF 자원의 속성(property)과 클래스(class)를 표현할 수 있는 어휘로써 속성과 클래스의 일반화 계층 구조(generalization hierarchies)에 대한 의미론을 제공한다. OWL은 속성과 클래스에 대하여 기술할 수 있는 더 많은 어휘를 제공한다. 예를 들면 클래스 간의 관계, 관계 차수, 동치성(equality), 풍부한 속성 타입, 속성의 특성, 열거형 클래스 등을 기술할 수 있다(<http://www.w3c.org/TR/2004/REC-owl-features-20040210/>).



Ⅲ. 지식표현을 위한 은톨로지 구축

1. 지식표현의 정의와 방법론

인식론에서 ‘지식(knowledge)’은 정당화된 참인 믿음(justified true belief)으로 정의되나 전산학에서는 ‘지식’이 꼭 이와 같이 엄격한 의미로 쓰이는 것은 아니다. 이것 외에 ‘정보(information)’ 나 ‘데이터(data)’ 또한 함께 사용되는데 이들은 비슷한 의미를 가진 용어들로 간주되기도 하고 혹은 때때로 서로 구분되기도 한다. 만약 구분될 경우 지식은 합목적적인 구조화 혹은 체계화란 측면에서 정보나 데이터와는 다르다. 그러나 지식 표현(knowledge representation, KR)에 있어서는 어떤 종류의 정보라도 그것이 전산 시스템 상에서 표현될 수만 있다면 흔히 ‘지식’으로 불리기 때문에 지식 표현은 곧 정보 표현(information representation)이라 할 수 있다.

지식 표현은 인공지능의 주요한 하위 분야 중 하나이다. 추상적인 지식이 어떤 방식으로든 전산 시스템 상에서 실제로 구현된다면 원칙적으로 그것은 지식 표현의 사례라 할 수 있다. 예컨대 지식은 Neural Networks에서 노드들 사이의 연결 가중치(connection weight)의 벡터로, 그리고 DNA 컴퓨팅에서는 DNA 가닥의 염기 서열로 각각 다르게 나타난다. 그러나 대개는 ‘지식표현’이란 용어를 고전적 인공지능(기호주의, symbolic AI)의 기호처리시스템 가설 아래에서 이해하는 것이 일반적이다.

즉, 모든 지능적인 시스템은 내부의 기호들(symbols)을 써서 세계에 관한 특정 지식들을 표상 혹은 인 코딩한 후, 이 기호들을 조작함으로써 그것에게 기대되는 지능적인 작업을 수행하게 된다. 이와 같이 기호는 의미론적 성질과 통사론적 성질을 동시에 갖기 때문에 지능적인 시스템은 지식 베이스와 추론 엔진이라는 두 개의 분리된 모듈의 결합으로 간주될 수 있다.

R. Davis 등은 “지식 표현이란 무엇인가(1993)”란 글에서 지식 표현이 갖는 역할을 다음과 같이 5 가지로 요약하여 제시한 바 있다

- 지식 표현은 대용물이다(A KR is a Surrogate).
- 지식 표현은 존재론적 개입들의 집합이다(Set of Ontological Commitments).
- 지식 표현은 지능적 추리의 단편적 이론이다

(Fragmentary Theory of Intelligent Reasoning).

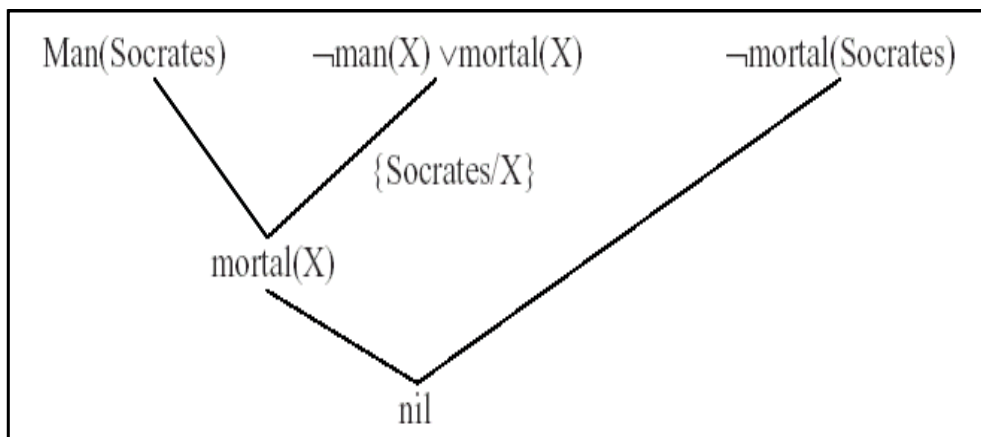
- 지식 표현은 효율적인 계산을 위한 매개체이다(Medium for Efficient Computation).
- 지식 표현은 사람들의 표현의 매개체이다(Medium of Human Expression.).

주요한 지식 표현 방법들 중 가장 대표적인 몇 가지를 살펴보면 다음과 같다.

첫째, 논리기반 지식표현

지금까지 고안된 다양한 논리들 중 지식 표현 분야에서 가장 많은 사람들의 관심의 대상이 되었던 것은 일차 술어 논리(first-order predicate logic)이다. 왜냐하면 이것은 비교적 적은 수의 primitives만을 포함하고 있으면서도 매우 풍부한 표현력(expressiveness)을 자랑하고 잘 정의된 model theory 및 proof theory를 갖고 있기 때문이다. 무엇보다도 다른 논리에 비해 사람들이 이해하기 용이하다. 일차 술어 논리를 이용한 지식 표현은 형식화된 연역적 이론 T를 구성하고 T의 비논리 상황들에 의미론적 해석을 부여함으로써 이루어진다. 이것은 지식 표현 외에 지식 추론에도 또한 이용된다. 논리 기반의 기계적 추론, 즉 논리 프로그래밍은 인공 지능의 또 다른 주요한 분야이다. 추론을 보다 용이하게 하기 위해 표현력에 일정한 제한을 가하는 경우가 많고(예컨대 horn clause 논리 등) 연역 장치로는 resolution refutation 방법이 흔히 사용된다. 예컨대 논리 프로그래밍 언어인 Prolog 자체도 이 장치에 기반한 정리증명기(theorem prover)이다.

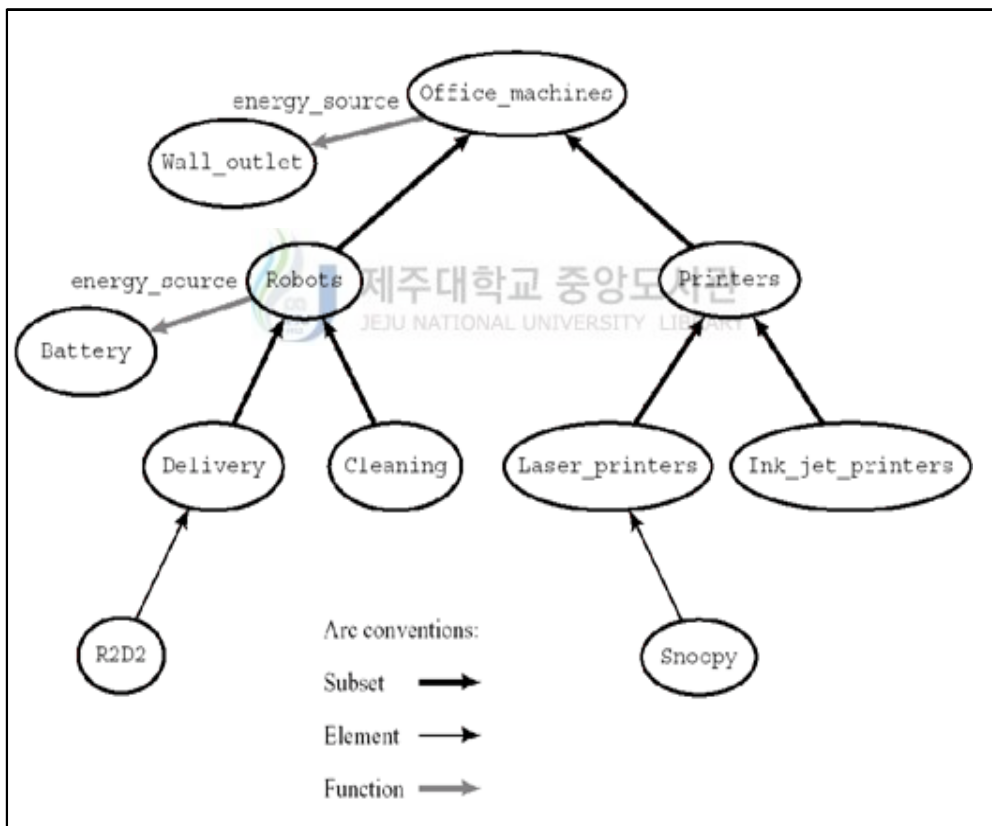
<그림 4> resolution refutation을 이용한 정리 증명의 예



둘째, Semantic Network 기반 지식 표현

Semantic Network은 대상들과 그것들의 속성들에 관한 분류학적 지식을 나타내는 graph이다. 이 경우 보통 정점(vertex)을 ‘노드(node)’라 하며 이것은 두 종류로 구별된다. 하나는 분류학적 범주나 속성을 나타내는 노드들이고 다른 하나는 대상영역내의 구체적 대상들을 나타내는 노드들이다. 또한 Semantic Network에서는 간선(edge)을 보통 ‘아크(arc)’라 하는데 이것은 부분 집합 아크들 혹은 isa 연결들(links), 집합 구성원 아크들 혹은 사례(instance) 연결들, 함수 아크들로 구분되어 진다.

<그림 5> 의미망의 예 : 사무실 내 기계들(N. Nilsson 1998)



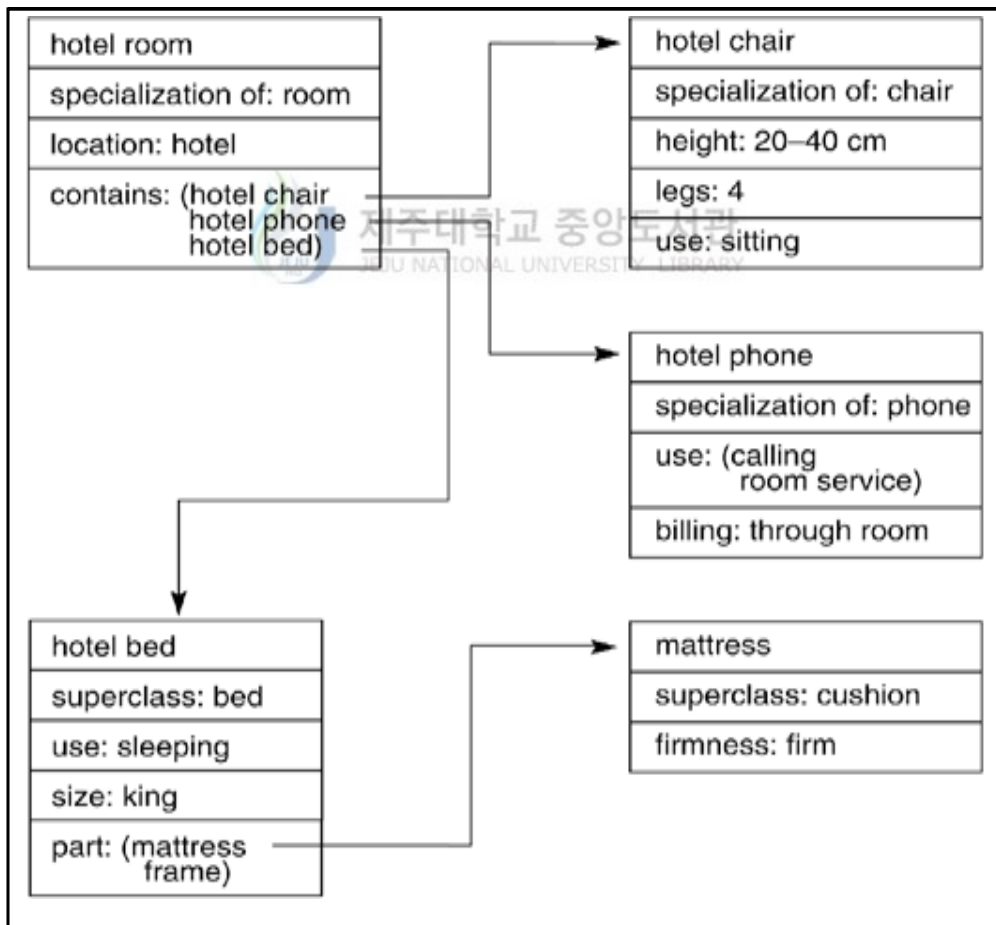
셋째, Frame 기반 지식 표현

프레임은 프레임 이름과 속성-속성 값 순서쌍들의 집합의 쌍으로 볼 수 있다. 의미망의 관점에서 볼 때 프레임 이름은 의미망의 노드에, 속성은 이 노드와 연결되는 아

크에, 그리고 속성 값은 이 아크의 다른 끝 쪽에 있는 노드에 대응한다. 이러한 속성-속성 값의 순서쌍을 슬롯(slot)이라고 하고, 속성을 슬롯 이름(slot name), 속성 값을 슬롯 값 혹은 슬롯 채우기(slot filler)라고 한다.

아래 그림은 호텔방 및 그 안의 비품들을 프레임 구조를 이용하여 표현한 예이다. 프레임 체계에서는 서로 관계있는 지식들이 그들 사이의 계층 구조에 따라 조직화되므로 일단 표현된 지식들은 사람이 보기에 자연스러우며 따라서 그 처리가 용이하다. 반면 실세계의 복잡한 대규모 지식들을 프레임 구조로 실제로 조직화하는 작업은 매우 힘들고 많은 노력과 비용을 요구 한다.

<그림 6> Frame의 예: 호텔방 내 비품들(임해창 2003)



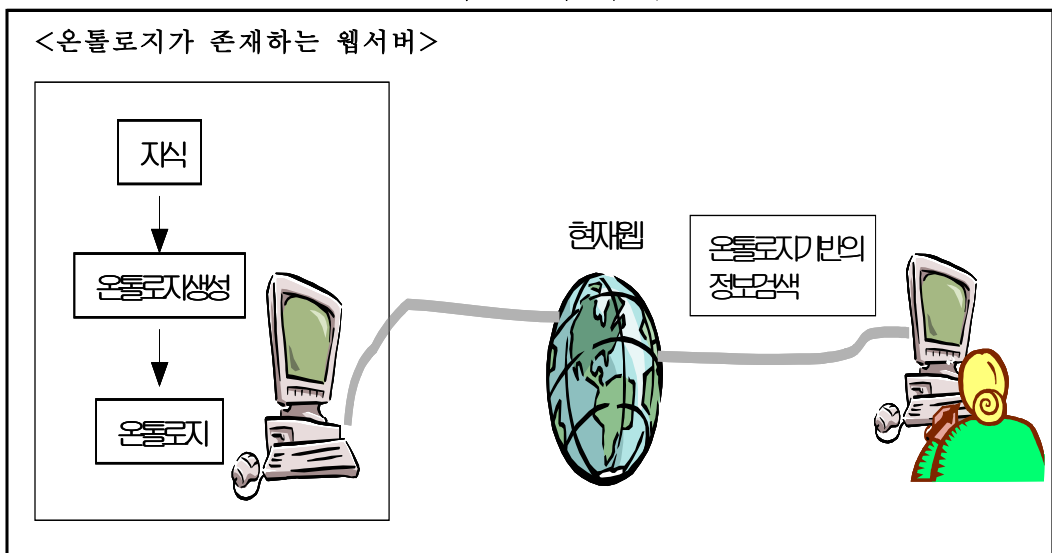
2. 시스템 설계

기존의 웹은 분산된 환경 하에서 자료를 공유한다는 목적으로 만들어져 웹 개발자들에 의해 많은 양의 정보가 제공되었다. 그러나 폭주된 정보량은 사용자로 하여금 필요한 정보만을 찾기 위해 불필요한 정보들을 걸러내야 하는 번거로움을 준다. 이는 HTML이 의미적인 검색이 이루어지지 않기 때문이다. 따라서 시맨틱 웹 환경에서 온톨로지를 기반으로 의미적 검색을 한다면 이런 불편을 줄일 수 있을 것이다.

<그림 7>에서 보면 웹 서버는 온톨로지를 생성하고 그 온톨로지를 기반으로 데이터베이스에서 정보를 검색하여 인스턴스 파일을 생성하여 저장하고 있다. 기존 웹에서 웹 사이트를 제공하는 웹 서버는 특정 분야의 지식을 기반으로 의미 구조를 정리하는 온톨로지를 생성하여 자신의 온톨로지를 갖고 있다. 물론 온톨로지는 재사용이 가능하여 기존의 온톨로지를 상속받아 사용할 수도 있다. 기존 웹에 분산되어 있는 온톨로지는 온톨로지 mapping에 의해 연결되어 지능적인 에이전트들에 의해 서로 간에 의사소통하며 더 나아가 스케줄링이나 가격협상과 같은 웹 서비스를 제공할 수도 있다.

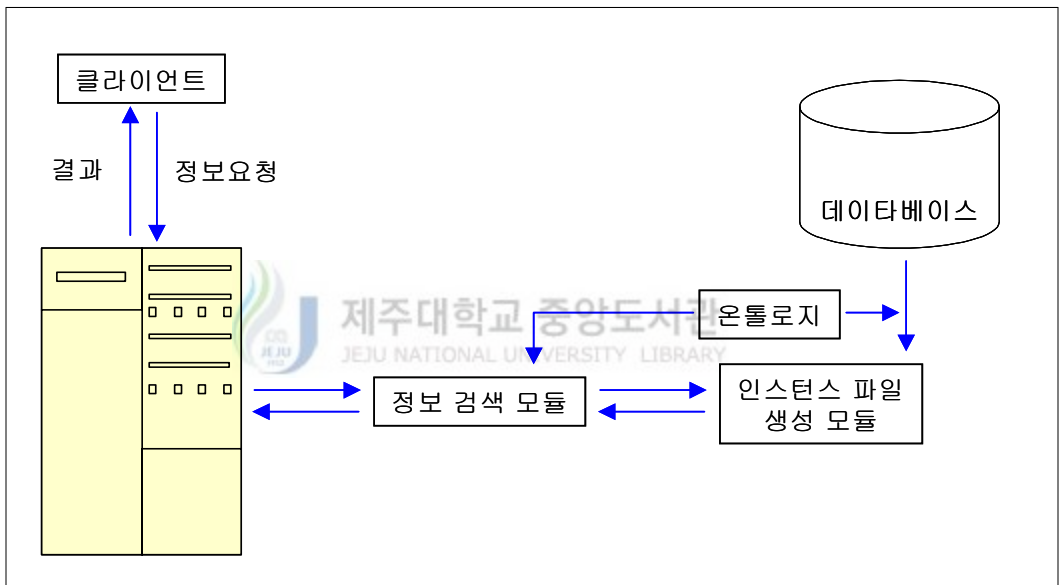


<그림 7> 시스템 구조



<그림 8>에서 보면 인터넷 망에 연결된 클라이언트는 알고 싶은 정보를 자주 사용하는 키워드를 이용하여 입력한다. 입력받은 정보를 이용하여 서버내의 정보검색 모듈에서는 온톨로지를 이용하여 키워드의 의미를 파악하고 그에 해당하는 정보를 검색한다. 그리고 그 결과를 클라이언트에게 출력한다.

<그림 8> 시스템 설계



3. 온톨로지의 개발과정과 라이프사이클

소프트웨어 공학 분야에서 지식과 소프트웨어의 재사용(reuse)은 프로젝트의 개발 기간을 단축하고 소요되는 자원을 절약할 수 있기 때문에 다양한 방법론이 존재하고 있다. 특히 지식기반 시스템에서 지식의 생성이나 획득은 시스템 개발의 핵심 과정이기 때문에 지식의 재사용은 무엇보다 중요하다. 그러나 재사용은 여러 가지 요소를 함께 고려해야 하는 문제가 있다. 일반적으로 지식의 재사용과 관련되어 언급되는 것에는 언어, 표현 형식, 개발 도구의 異種性, 어휘적 (lexical) /의미적(semantic) 문제

와 같은 것들이 있다. 온톨로지를 개발하는 과정에서도 이와 같은 문제들은 여전히 존재하며 분산되어 있는 지식 베이스에서 지식을 관리하기 위해 일관성 유지와 같은 관리적 측면도 함께 고려되어야 한다. 온톨로지의 개발은 각각의 팀이 유기적으로 협업하여 진행하는 활동이라 할 수 있다. 즉 개발에 참여하는 팀들은 공유할 수 있는 개발 지침, 디자인 기준과 같은 원칙에 기준하여 개발에 참여하게 된다. 구조적인 가이드라인이나 방법이 존재하지 않을 경우 팀 내에서 발생하는 이견뿐만 아니라 팀 사이의 문제, 온톨로지 사이의 문제에 대한 일관성을 유지하기 힘들기 때문에 결과적으로 완결성 높은 산출물을 기대하기 힘들다. 이러한 문제의 근본적 원인은 온톨로지 개발에 있어 명시적으로 문서화된 개념 모델이 존재하지 않기 때문이다.

온톨로지의 개발은 기존에 사용되었던 소프트웨어 공학 기술과 지식 기반 시스템의 개발에 적용되었던 방법론에서 많은 부분을 응용할 수 있다. 그러나 현실적으로 이러한 방법론을 직접 온톨로지의 개발에 적용하는데 문제가 있기 때문에 온톨로지 개발자들은 온톨로지의 개발 목표와 단계에 따른 적절한 방법론을 적용하고 보완해 갈 수 있는 과정이 부가적으로 요구된다.

온톨로지를 만들기 위한 과정은 Noy & McGuinness의 Ontology Development 101에 의하면 다음과 같은 과정이 요구된다.

- step 1 : 온톨로지 도메인과 범위 결정
- step 2 : 온톨로지 재사용 여부 결정
- step 3 : 중요한 용어 목록 결정
- step 4 : 클래스와 클래스의 계층 구조 결정
- step 5 : 클래스-슬롯과 속성 결정
- step 6 : 슬롯의 패싯(facets) 결정
- step 7 : 인스턴스 생성

온톨로지의 개발 과정은 크게 명세화(specification), 개념화(conceptualization), 형식화(formalization), 구현(implementation), 관리 및 유지(maintenance)로 구분할 수 있다. 명세화는 온톨로지를 실제 개발하기 전에 수행되는 과정으로 온톨로지의 범위와 사용자의 범위 등을 규정하게 된다. 온톨로지의 개발 단계에는 개념적 모델과 정형화

된 모델로 구성하고 컴퓨터 가 처리 가능한 형태로 만드는 과정이 포함되게 된다.

위에 기술된 온톨로지 개발 과정은 일반적으로 온톨로지를 개발할 때의 절차를 보여주고 있고 개발 목적이나 도메인에 따라 다양하게 적용된 방법론들이 존재한다. 개발적인 온톨로지의 라이프사이클은 아래와 같다(http://protege.stanford.edu/ontology_development/ontology/101.html).

- step 1 : 초기 온톨로지 디자인
- step 2 : 개념적인 개선
- step 3 : 온톨로지 간의 연결
- step 4 : 평가
- step 5 : 진화

제 1단계 : 초기 온톨로지 디자인

도메인의 특성과 온톨로지 사용용도를 고려하여 온톨로지 구성영역을 정하는 단계이며 온톨로지 영역이라 함은 온톨로지가 응답해야 하는 질의들을 나열하여 온톨로지 내에 포함되어야 할 내용의 세부사항들을 정하고 기존의 온톨로지를 경제적으로 활용하기 위해 그리고 특정 온톨로지를 이미 사용하고 있는 애플리케이션과 상호적인 처리를 위해 기존 온톨로지의 활용은 중요하다. 온톨로지 학습은 언어학적인 분석과 기계 학습관계를 이해하여 멀티미디어와 같은 다양한 자료에서 자동적으로 학습될 수 있는 부분을 고려하여 온톨로지 개발 목적에 따라서 메타데이터의 주석이나 찾는 대상의 구조, 개인별 특징도 고려한다.

제 2단계 : 개념적 개선

개념적 구조에서 시작하여 개념들을 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. 실제 적용대상인 콘텐츠에 맞추어 기존의 온톨로지를 수정, 재사용한다. 기존에 사용했던 개념의 개요나 유의어 사전과 같은 것을 수정하여 활용하기도 한다. 개념들에 대한 일반적인 의견이나 다른 관점을 수용할 수 있는 틀과 방법들도 동시 개발하며 최적의 모듈화를 고려하고 모

들간의 관계를 나타낸다. 특정 태스크 수행에 있어서 필요한 위임, 위탁, 제어, 최소한의 여론, 일치성을 검증하기 위해 필요한 관련 에이전트들 간의 상호작용 방법을 활용하며 점진적인 수정 개발 작업을 가능케 할 추론을 고려한다.

제 3단계 : 온톨로지 간의 연결

정렬, 제휴 방법과 통합, 병합 방법이 있다.

제 4단계 : 평가

온톨로지가 요구조건에 맞는지 확인하고 다른 애플리케이션에 사용하여 실험적인 결과를 분석 하며 온톨로지의 일관성을 확인하며 표현 능력, 정확도, 인지학적 적절성 등 이론적인 방법을 통하여 온톨로지를 비교한다.

제 5단계 : 진화

변해가는 즉, 버전이 다른 온톨로지 간의 관계 성립과 함께 온톨로지 사용과 새로운 정보를 통해 온톨로지 변화의 필요성과 아울러 온톨로지 변화가 메타데이터 주석에 미치는 영향을 고려한다.



4. 온톨로지 설계 방법

온톨로지 공학은 원칙적인 설계, 수정, 응용, 평가 단계를 거쳐 개발된다. 이러한 온톨로지 설계방법은 다음과 같이 다섯 분류로 나누어 볼 수 있으며 이들 간의 혼합적인 방법도 가능하다(Clyde W. Holsapple, K. D. Joshi 2002).

<표 4 > 온톨로지 설계 방법

방 법	기 본 원 리
영감적 방법	관심분야에 대한 개인적 관점
귀납적 방법	관심분야의 특정 case
연연적 방법	관심분야에 대한 일반적 원칙
통합적 방법	관심분야에 부분적 설명을 제공하는 기존 온톨로지들의 합성
협력적 방법	초기 온톨로지를 기준으로 다수 개인들의 관점을 결합

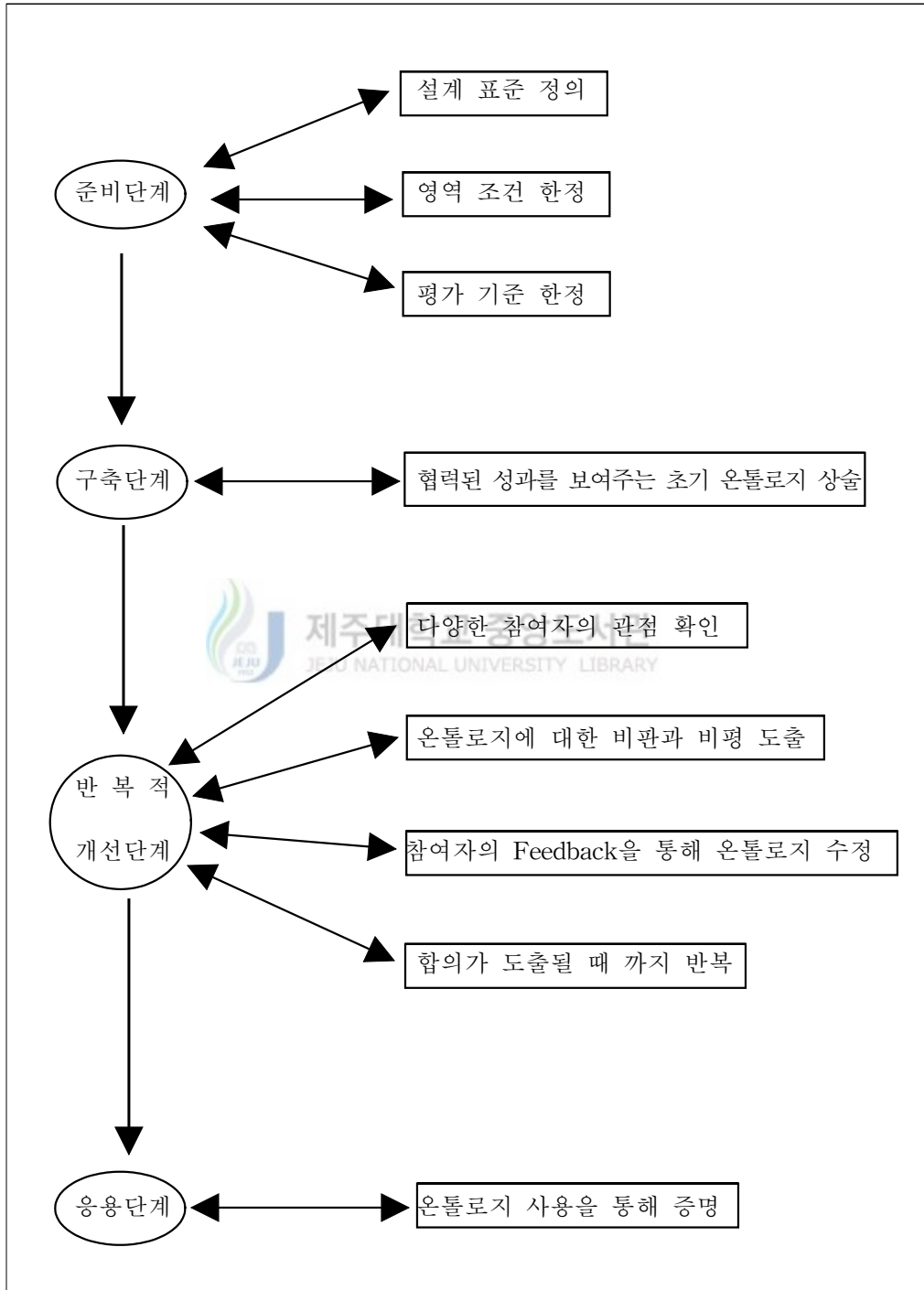
영감적 방법은 개발자가 개인적 관점에서 관심 분야의 독특한 상상이나 독창성을 이용해 혁신적인 온톨로지를 개발할 수 있으나 이론적 바탕이 부족하고 비현실적인 단점이 있다. 귀납적 방법은 관심분야의 특정한 case를 관찰, 조사, 분석하여 얻어진 특징들을 그 분야의 다른 case에 일반화하여 적용시키는 방법이지만 일반성을 결여하기 쉬운 단점이 있다. 연역적 방법은 특정 case를 도출하기 위해 선택할 수 있는 일반성을 지닌 스키마가 존재한다는 가정 하에 일반적인 원칙들로 이루어진 온톨로지를 특정한 case에 맞춰 점차적으로 수정해 가는 방법이다. 통합적 방법은 이미 존재하는 온톨로지들로 부터 부분적으로 관심분야를 나타내는 부분들을 합성하여 하나의 단일화된 온톨로지를 구성하는 방법으로 위의 세 가지 방법을 통합적으로 적용하지만 개발자 개인의 해석이나 통합 기술에 의존적이다. 협력적 방법은 초기 온톨로지를 기준으로 하여 여러 사람이 함께 협력하여 개발하는 방법으로 다수의 관점을 결합하여 지속적인 평가와 제안을 통해 온톨로지를 개발하지만 개발자들의 지속적인 참여가 요구되고 그들의 개발기술에 영향을 받는다. 일치된 정보를 공유하고 재사용하는 차원에서 협력적 방법이 주로 시도되고 있으며 이 방법은 표현되어진 개념과 그런 개념과 적절하게 연관지어진 정보들이 공동 집단의 의견과 일치되도록 계속 변화할 수 있다는 장점을 지니고 있다.

<그림 9>는 조직 내에서 지식관리 행위를 위한 온톨로지 개발을 위해 사용되는 협력적인 Delphi 지향적 구조를 나타내는 온톨로지 설계단계를 예시하고 있다.

준비단계는 온톨로지를 위해 설계 기준들을 정의하고 영역 조건들을 열거하며 평가 기준들을 결정하는 단계이다. 지식관리 온톨로지를 위해 선택한 설계 기준들은 포괄적이고 정확하고 간결 하고 명쾌하고 유용하다. 이 기준들은 온톨로지 개발을 안내하고 온톨로지의 성공 정도를 평가하는데 중요하다. 비즈니스 영역, 記述的 영역, 세부 영역의 세 가지 영역 조건에 설계기준을 적용 시킨다. 비즈니스 조직에서 지식관리의 초점은 지식관리의 현상들을 묘사하는 것 보다 오히려 방법론들을 규정하는데 있다.

구축단계는 통합된 언어와 개념, 평가기준에서 상호연관성 등을 종합하는 과정으로 몇 번의 반복과정을 통해 온톨로지를 발전시키는 단계이다. 설계기준들이 평가기준에 만족할 때 까지 수정을 반복한다.

<그림 9> 협력적 온톨로지 설계방법



반복적 개선단계는 수정을 통해 구축된 온톨로지를 Delphi 기법을 적용한 협력적 방법으로 온톨로지를 설계하는 단계이다. Delphi 기법은 어떤 주제에 대해 다양한 사람들의 견해를 통합하고 모아주는 형식적인 기법으로 반복적 개선단계에서 온톨로지의 토대를 마련해 주는 견해와 기준들을 모으는데 규칙적인 방법을 제공하며 온톨로지 설계 기준에 관하여 독립된 판단을 모으는데 편리하다. 30명 정도의 참여자를 선정하여 그들의 관점을 확인하고 온톨로지에 대한 비평을 이끌어내고 Feedback을 통해 온톨로지를 수정하여 합의가 도출될 때 까지 이런 과정을 반복하는 단계이다.

마지막으로 응용단계는 합의 도출된 온톨로지를 다양한 방법을 적용하여 유용성을 증명하는 단계이다(Clyde W. Holsapple, K. D. Joshi. 2002).

5. 델파이 적용

‘Delphi’라는 이름은 고대 희랍신화 중에서 아폴로 신이 미래를 통찰하고 신탁을 하였다는 신전 델파이를 따른 것이다. 델파이 방법은 예측하려는 문제에 대하여 전문가들의 견해를 유도하고 종합하여 집단적 판단으로 정리하는 기법이다.

이 방법은 미국의 랜드연구소(Rand Corporation)에서 대면토의의 제한점을 제거하고 당시 긴급한 문제였던 소련의 핵무기 보유량 문제에 관하여 전문가들의 합의를 얻는데 최초로 사용하였다. 이 방법은 추정하려는 문제에 대한 정확한 정보가 없을 때에는 “두 사람의 의견이 한 사람의 의견보다 정확하다”는 통계적 객관의 원리와 “다수의 판단이 소수의 판단보다 타당하다”는 민주적 의사결정의 원리에 논리적 근거를 두고 있다(이종성 1988).

연구과정을 좀더 구체적으로 소개하면 미천굴 관광지구의 홈페이지와 관련 자료들을 기초로 하여 초기의 델파이 설문지를 개발하였다. 이 1차 설문지의 내용은 세 가지, 즉 미천굴 관광지구를 자랑할 수 있는 특징적인 것 5가지를 중요한 순서대로 적어달라는 것과 제주도내 다른 관광지와 비교하여 차별화되었다고 생각하는 것을 적어달라는 것, 마지막으로 미천굴 관광지구를 하나의 그림으로 나타냈을 경우 입장, 관람, 체험, 편의시설, 조직으로 구분하였는데 응답자의 추가 견해를 묻는 항목으로 구성하였다. 전체 직원 37명 중 46%인 17명이 1차 설문에 응답했는데 회수율은 저조했지만 개방형 질문에 답해준 내용은 매우 풍부했다. 1차 설문지 반응결과를 기초로

2차 설문지를 작성하여 동일한 집단에게 조사하였다. 2차 설문지는 1차 설문 시 반응이 애매하거나 응답률이 저조했던 부분은 연구자가 일부 수정하여 작성하였다. 응답자별 반응 빈도는 미친굴 관광지구 직원 전체로 하되 부서 및 매장별로 인원을 배분하여 산출하였는데 이는 델파이 설문은 집단 추정치에 기초하여 특정 결과에 대하여 예측하는데 그 목적이 있기 때문이다. 1차 설문은 전체 직원 37명 중 35명에게 개방형 설문을 직접 전달하여 그 중 17부를 회수 하였다. 일주일 후에 2차 설문을 동일한 집단을 대상으로 실시하였는데 2차 설문의 내용은 온톨로지의 CLASS와 CLASS의 계층구조를 결정하기 위한 합의를 도출하기 위한 문항으로 구성하였다. 미친굴 관광지구의 지식체계를 적용하기 위한 분석내용을 좀더 구체적으로 살펴보면 우선 단순 관람과 체험 관광으로 양분되어 있는 목적별 구분을 수용하고 그 외에 중요하게 고려되어야 할 요소들의 순위를 정하도록 하는 것은 자칫 응답자들에게서 성의 없는 답변이 나올 수도 있기 때문에 응답자가 생각하기에 중요한 순서대로 5가지만을 선택하도록 범위를 축소시켰고 선택되지 않은 요소들에 대해서는 중앙 순위 값을 일률적으로 부여하여 상위 3개 요소는 2차 설문문에 포함시켜서 2차 설문지의 1번 문항은 8개 요소 중에서 선택하도록 조치하였다(설문지 참조).

6. 온톨로지 구축시 고려사항

온톨로지 구축단계를 간략히 설명하면 첫째, 특정한 목적과 영역을 고려한 다음 개념을 자동, 반자동으로 추출하거나 어휘사전을 확보하여 개념들을 정의하고 조직화해야 한다. 조직화한다는 것은 개념들 간의 관계를 설정함과 동시에 개념이 가지는 특수한 속성을 추출하여 체계화시키는 것을 의미하는 것으로서 이론적 체계와 더불어 실질적인 구축원리를 마련해야 한다.

둘째로 온톨로지를 표현할 웹 온톨로지 언어나 기계가독형 표현언어(machine readable representation language)를 설정하여 형식화하고 실질적으로 구현해야 하며 구축중인 온톨로지와 다른 온톨로지와의 통합 문제와 기존에 존재하는 많은 자원(resources)을 어떻게 활용할 것인가를 모색해야 한다. 마지막으로 구축된 온톨로지를 대상 애플리케이션(target application)에서 실험하거나 사용패턴(usage patterns)을 분석하여 평가해야 한다. 또한 이러한 평가 결과를 바탕으로 유지보수를 해야 하며

조금 더 발전적인 온톨로지로 개선해야 한다. 이러한 일련의 구축단계에서 가장 어려운 문제는 바로 온톨로지를 실질적으로 구축하는 이론적 체계와 원리가 아직까지 마련되지 않았다는 것이다. 기존의 구축사례들을 보면 기존의 시소러스나 의미망, 분류체계 등을 이용한 온톨로지 구축이나 기 구축된 온톨로지를 이용한 애플리케이션 개발이 대부분이다. 이것은 온톨로지 구축에 있어서 국내외적으로 가지는 공통적 문제라 할 수 있다. 이를 위해서는 WordNet, UMLS와 같이 관련 학문에 대한 이론 습득과 더불어 자연어 처리기법의 활용을 통한 이해처리 등과 같은 부수적인 연구가 뒤따라야 할 것이다. 이외에 온톨로지 구축시 고려사항을 정리하면 다음과 같다.

- 온톨로지 구축 대상자 선정
- 온톨로지 구축 방법(자동/반자동/수동)
- 응용분야별 온톨로지 구축 원리 파악
- 기구축된 시소러스, 분류체계 활용방법
- 범용적/국제적 수준의 온톨로지 구축
- 온톨로지 제작도구 개발
- 온톨로지 구축관련 표준방식 연구
- 시맨틱 마크업 언어 표준안 개발
- 기타

7. 온톨로지 구축

본 온톨로지는 protege-2000에 의해 구축하였다.

STEP 1 : 온톨로지 도메인과 범위 결정

온톨로지 도메인은 미천굴 관광지구로 한정한다.

STEP 2 : 온톨로지 재사용 여부 결정

DAML+OIL에 관련된 온톨로지는 IMPORT 받아서 사용하고 미천굴 관광지구 에 해당하는 온톨로지는 새로이 작성한다.

STEP 3 : 중요한 용어 목록 결정

Property	동일의미
Admission-Fee	입장료, 관람료, 入場料
Adult	成人入場料, 어른입장료, 일반입장료, 성인입장료
Youth	青少年入場料, 중고생입장료, 청소년입장료
Child	小人入場料, 어린이입장료, 소인입장료
Soldier	軍人入場料, 警察入場料, 군경입장료
Elder	敬老入場料, 老人入場料, 경로입장료, 노인입장료
Name	이름, 名稱, 명칭
Tel	전화, 電話番號, Telephone, Phone, Tel
Price	가격, 價格, 요금, 料金

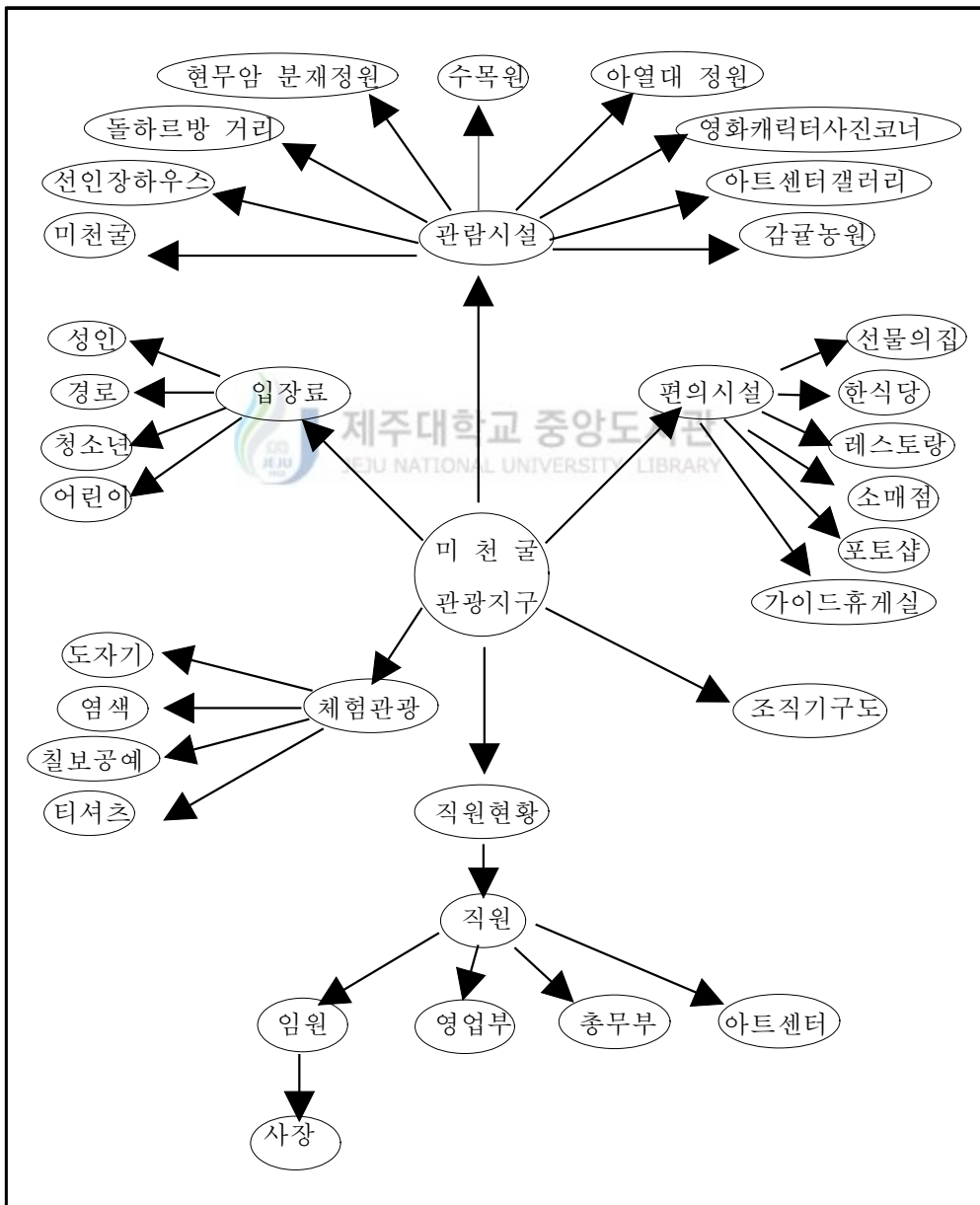
STEP 4 : 클래스와 클래스의 계층 구조 결정

두 차례에 걸친 델파이 설문지를 통하여 조사한 내용을 <부록 2>와 <부록 3>에 정리하였는데 이를 분석한 결과를 토대로 클래스의 계층구조를 결정하였다. 먼저 클래스로 입장료, 관람시설, 체험관광, 편의시설, 조직기구도, 직원현황으로 구분하였고 입장료의 Sub Class로 성인, 청소년, 어린이, 경로의 항목으로 구성하였다. 관람시설의 Sub Class로는 미천굴, 선인장하우스, 아열대정원, 현무암 분재정원, 수목원, 감귤농장, 돌하르방 거리, 영화 캐릭터 사진코너, 아트센터 갤러리로 구성하였으며 체험관광의 Sub Class로는 도자기, 염색, 칠보공예, 티셔츠를 설정하였으며 편의시설의 Sub Class로 선물의 집, 식당, 소매점, 포토샵, 안내자 휴게실로 구성하였다. 직원현황의

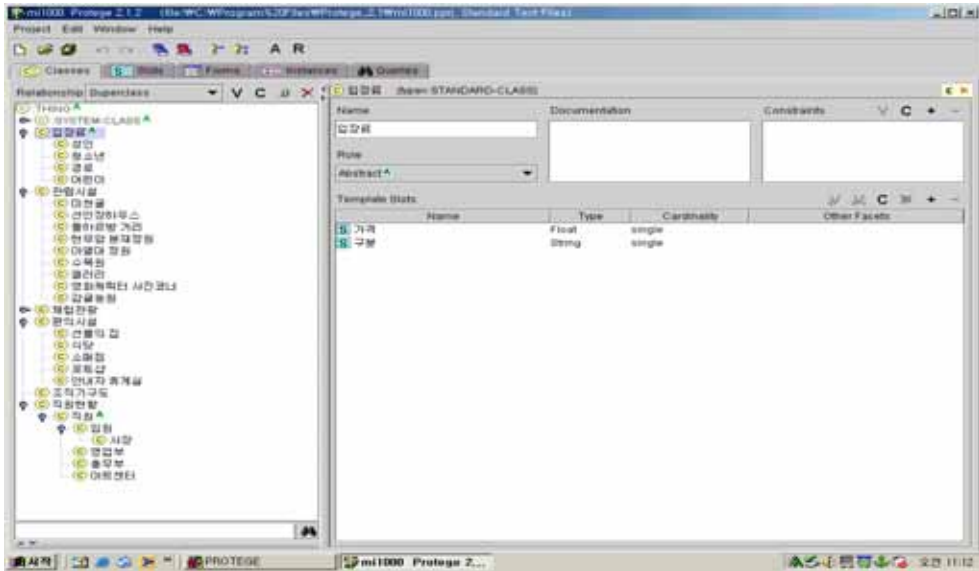
Sub Class로 직원을, 직원의 Sub Class로는 임원과 영업부, 총무부, 아트센터를 두었으며 임원의 Sub Class로 사장의 계층 구조를 결정하였다.

<그림 10>은 클래스와 클래스의 계층구조도를, <그림 11>은 <그림 10>을 기초로 하여 작성한 protege-2000에서 온톨로지를 구축한 결과를 제시하였다.

<그림 10> 미천굴 관광지구의 클래스 계층구조도



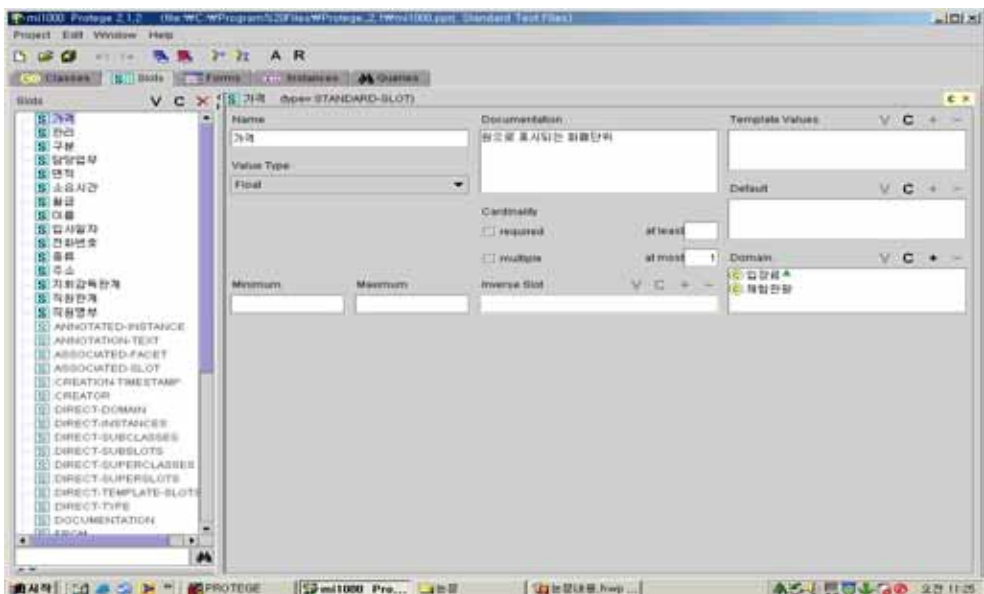
<그림 11> protege-2000에서의 클래스 계층 구조



STEP 5 : 클래스-슬롯과 속성 결정

각 클래스의 서브클래스와 슬롯은 <그림 12>에 나타난 바와 같다.

<그림 12> protege-2000에서의 클래스-슬롯과 속성결정



STEP 6 : 슬롯의 패시(facets) 결정

STEP 3에서 정리한 property 중 전화번호에 대한 DAML+OIL로 작성된 온톨로지 부분이다.

```
<daml:DatatypeProperty rdf:ID='전화번호'>
  <daml:range
    rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
  <daml:equivalentTo rdf:resource='#TEL'/>
</daml:DatatypeProperty>
<daml:DatatypeProperty rdf:ID='PHONE'>
  <daml:range
    rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
  <daml:equivalentTo rdf:resource='#TEL'/>
</daml:DatatypeProperty>
<daml:DatatypeProperty rdf:ID='TELEPHONE'>
  <daml:range
    rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
  <daml:equivalentTo rdf:resource='#TEL'/>
</daml:DatatypeProperty>
<daml:DatatypeProperty rdf:ID='電話番號'>
  <daml:range
    rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
  <daml:equivalentTo rdf:resource='#TEL'/>
</daml:DatatypeProperty>
```

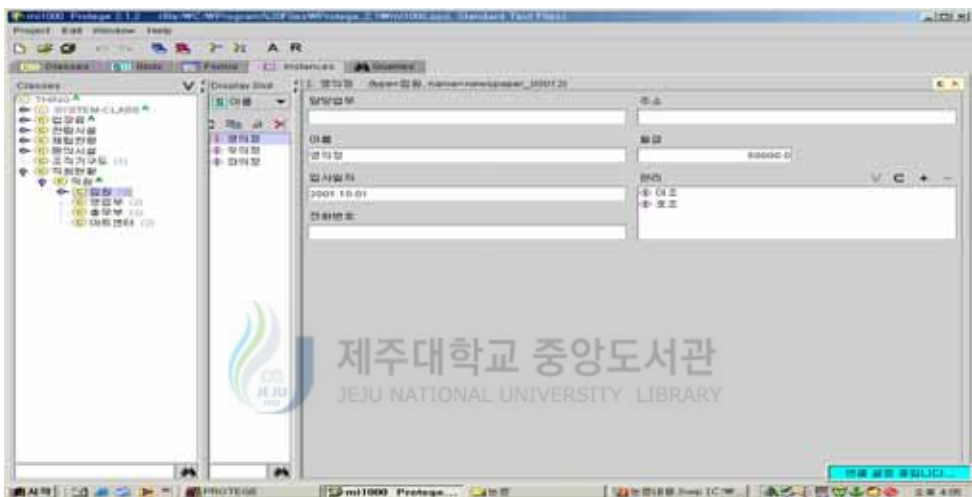
이처럼 '전화번호', 'TELEPHONE', '電話番號', 'TEL', 'PHONE'을 DAML+OIL의 equivalentTo를 이용하여 동일 의미를 나타낼 수 있다. STEP 3에서 보았듯이 각각의 property들은 하나의 property만 존재하는 것이 아니라 같은 의미를 나타내는 property들의 관계를 온톨로지에 작성한다. 따라서 전화번호를 알고 싶으면 영어권의 관광객은 TEL, PHONE, TELEPHONE 등의 단어를 입력할 것이고 한자권의 관광객

은 電話番號를 입력할 것이다. 이렇게 특별한 언어번역처리 없이도 의미관계에 의해 같은 의미를 나타내는 단어들을 지원할 수 있다.

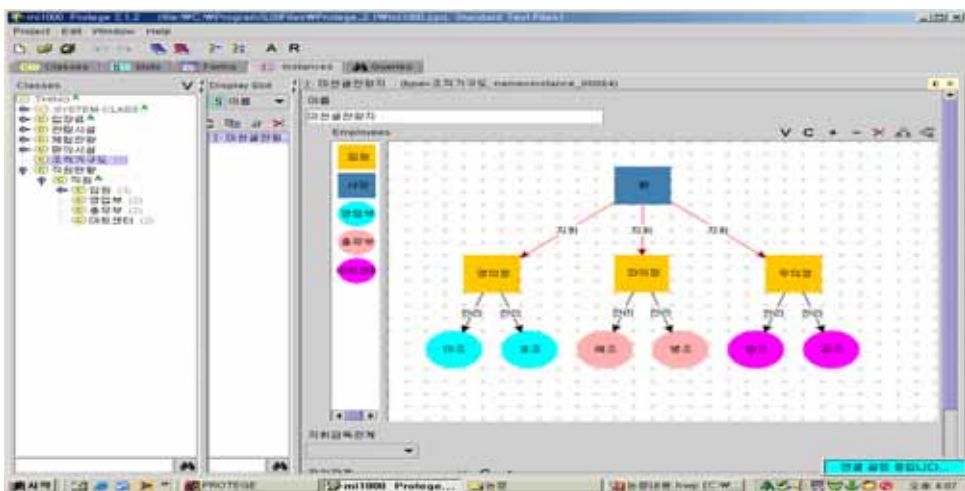
STEP 7 : 인스턴스 생성

6단계까지 작성된 온톨로지 파일을 이용하여 자동 생성모듈에 의해 인스턴스를 생성한다. <그림 13>은 자동 생성된 인스턴스를 보여주고 있고 <그림 14>에는 조직기구조를 Form에서 Chart로 작성한 그림이다.

<그림 13> 자동생성된 인스턴스

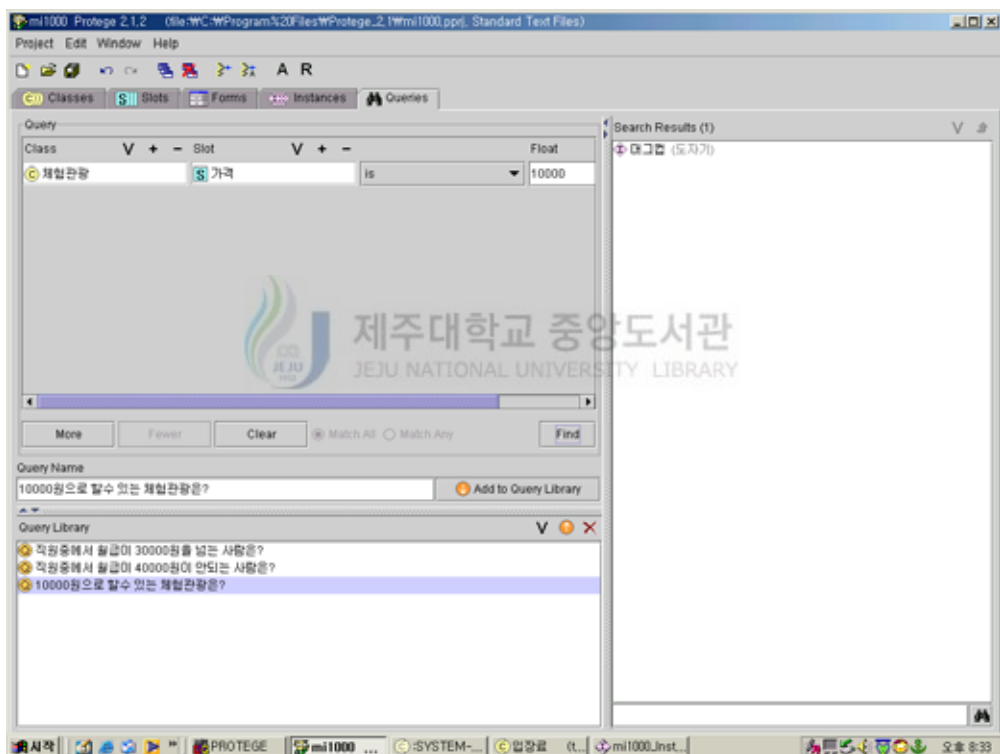


<그림 14> Chart로 표시된 조직기구조



이상과 같은 단계를 거쳐 미천굴 관광지구에 대한 온톨로지를 protege-2000을 이용하여 구축해 보았으며 <그림 15>에서 구문형으로 작성된 세 가지 질의사항에 대하여 보다 정확한 검색 결과를 얻을 수 있었고 그 중에서 10,000원으로 할 수 있는 체험관광의 프로그램을 질의하여 검색된 결과로 도자기 체험 중에서 ‘머그컵 만들기’ 항목의 결과를 확인할 수 있다.

<그림 15> protege-2000에서의 구문형 Queries에 대해 검색된 결과



IV. 온톨로지의 응용과 발전방향

1. 응용 사례

시맨틱 웹의 응용은 에이전트 기반의 웹 서비스 제공과 Annotation 이나 authoring 등과 같은 유용한 응용 프로그램의 개발로 요약된다(McIlraith, S. 외 2001). 현재 연구되거나 개발된 몇 가지 시맨틱 웹 응용사례를 살펴보면 다음과 같다.

Annotation은 시맨틱 웹을 가장 쉽게 응용할 수 있는 메카니즘이다. Annotation은 이미 존재 하는 웹 페이지에 추가적인 설명을 덧붙여서 다시 웹에 publish하는 것으로 주로 정보 검색의 정확도를 높이는데 크게 기여할 수 있다.

이러한 annotation을 가능하게 해 주는 tool로서는 OntoMat-Annotizer, SHOE, Annotea, Annozilla, COHSE Annotator 등이 있다. MusicBrainz는 응용 프로그램으로서 사용자가 자신의 데이터베이스로 음악 메타데이터를 post 방법을 이용하여 저장하고 이 데이터를 다른 사용자가 get 방법을 이용하여 검색할 수 있도록 해준다.

음악 데이터에 대한 메타데이터라 하면 앨범명, 아티스트명, 제작사, 트랙 번호, 연주시간 등의 데이터를 말한다. 이를 위해 RDF 문을 사용하며 이러한 기능들이 FreeAmp라는 MP3 플레이어에 내장되어 있다. FreeAmp를 수행시켜 음악 CD를 열게 되면 MusicBrainz 서버에 트랙 이름과 아티스트에 대한 메타데이터를 요청해서 정보를 얻게 되고 이 정보에 따라 트랙을 선택하거나 기타 원하는 다른 작업을 할 수 있다. ITTalks는 DAML을 이용하여 IT 분야와 관련된 세미나 또는 초청 강연들에 대한 데이터베이스를 운영하고 이를 이용해 웹을 통해 세미나 내용을 검색할 수 있는 응용 서비스이다. ITTalks의 데이터베이스는 세미나 관련 정보에 대한 웹 페이지와 DAML specification을 자동으로 생성하는데 사용되며 또한 세미나와 연관된 에이전트 기반 서비스의 중심 역할을 수행한다. 세미나에 대한 메타데이터를 DAML로 표현하기 위해 ITTalks에서는 calendar, person, place, profile, talk, topic 등 여러 가지 종류의 온톨로지를 정의하고 이용한다. 또한 세미나의 주제와 사용자 관심도 등이 온톨로지를 이용해 자동으로 분류하거나 DAML을 소프트웨어 에이전트간의 통신

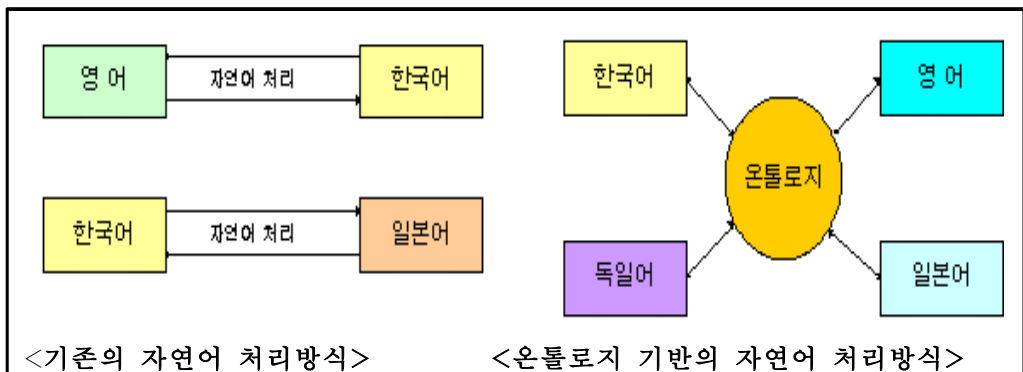
언어로 사용하는 등 고수준의 기능도 갖추고 있다.

IT분야에서 온톨로지는 에이전트와 컴퓨터 사이에서 이루어지는 커뮤니케이션을 형성하기 위한 수단으로 사용될 수 있다. 즉 컴퓨터가 온톨로지를 통해 정보의 의미를 보다 더 정확하게 이해하고 이를 조작할 수 있다면 온톨로지는 전자도서관, 정보 통합, HCI(Human Computer Interfaces), 비즈니스 프로세스 모델링, 의료정보, 바이오 인포메틱스 등 IT 분야의 다양한 영역에서 적용될 수 있으며 본 논문에서는 전술했던 자연어 이해처리와 지식관리, 그리고 e-비즈니스 적용가능성에 대해 기 발표된 자료들을 정리해 보았다.

2. 자연어 이해(Natural Language Understanding)

자연어 처리(Natural Language Processing)에서 가장 핵심이 되는 것은 자연어 이해이며 이때 이해 여부 및 이해 정도는 자연어 번역(Natural Language Translation)을 통해 확인 가능하다. 온톨로지는 개념(concept) 중심의 구조이며 여기에 적절한 indexing 알고리즘이 주어질 경우 이것은 동일한 개념을 표현하고 있는 서로 상이한 자연어의 어휘(lexicon)들을 매개시켜 주는 역할을 할 수 있다. 따라서 온톨로지는 자연어 문서에 대해 구문론 수준이 아니라 의미론 수준에서의 접근을 지원하며 이를 통해 보다 지능적인 방식의 자연어 번역 및 나아가 생성(generation)이 가능하다.

<그림 16> 기존 방식과 온톨로지 기반의 자연어 처리방식



3. 지식관리(Knowledge Management)

지식관리는 조직 안에서 필요한 지식을 어떻게 효과적으로 찾아서 가져오고 유지할 수 있으며 접근할 수 있는가를 다룬다. 시맨틱 웹 기반에서의 지식관리는 문서 중심의 관리체계를 문서 내에 존재하는 지식의 조각들로 그 관점을 옮겨 놓았다.

문서 내에 존재하는 지식들을 의미적으로 상호 연결시켜 놓음으로써 통합적인 지식관리와 비즈니스 프로세스의 처리가 가능하도록 하는 데에 온톨로지는 핵심적인 수단을 제공한다.

4. e-Business

최근 기업간 거래에 사용되는 용어에 대한 표준화가 이루어지고 있다. 특정 산업이나 기술 분야에 대한 많은 문서의 타입을 정의하고 있지만 아직은 구문적 정의 수준을 벗어나지는 못한 실정이다. B2B 전자상거래에서 트랜잭션 자동화는 상호 교환되는 문서의 형식이 사람이 이해할 수 있는 구문적 기술을 넘어서 기계가 이해하고 처리할 수 있는 형태로 기술된 형식을 필요로 한다.

비즈니스 간에 서로 다르게 사용되는 용어들을 공통의 의미로 해석하기 위해서는 의미적 상호 운용이 가능하고 지능적으로 정보의 통합이 가능한 온톨로지 형식으로 구성되어져야 한다.

정부가 142억원을 들여 2007년까지 시맨틱 웹과 지식처리엔진 등 지능형 e-비즈니스 플랫폼 기술을 개발한다고 밝혔고 이 지능형 e-비즈니스 플랫폼 기술이 지금의 전자상거래 시스템을 지능화, 자동화한 차세대 기술로 ERP, e-Market place, SCM 등 기존 e-비즈니스 시스템에 적용할 경우 생산성을 향상시키고 거래비용을 획기적으로 절감해줄 수 있을 것으로 기대하고 있다.

5. Ubiquitous Computing Business

유비쿼터스 컴퓨팅의 핵심은 임베디드(Embedded) 센서와 무선인식(RFID) 시스템 및 네트워크로 연결되어 상호 협력하는 스마트 기기이다.

각각의 유비쿼터스 기기들은 유무선 네트워크로 연결되어 있고 위치와 상황에 따른 인식 및 판단을 하고 정보의 효율적 전달을 위하여 동적인 네트워킹을 하는 ‘스마트 텔레컴퓨터’가 된다. 네트워킹된 스마트 기기는 물리적 공간에 존재하는 모든 물체에 컴퓨팅과 통신능력을 부여하여 디지털 공간과 융합되어진 ‘유비쿼터스 환경’을 창출한다. 그리하여 사람 중심의 P2P(person to person), P2M (person to machine)에서 사물 중심의 M2M 통신으로 패러다임이 변화하게 된다. 유비컴 기술은 실제와 가상의 세계 사이의 중개자 역할을 수행함으로써 정보시스템에서 물리적인 자원과 동작을 정보화하여 효율을 증가시킨다. 물리적인 자원은 인간의 개입 없이도 내부와 외부의 컴퓨터 네트워크 간에 통신할 수 있고 실시간 정보를 기반으로 지속적으로 처리과정을 제어하는 것이 허용된다. 예를 들어 물품 수송에 있어 경로를 놓치거나 잘못 배달된 물품에 대한 조사는 위치추적 시스템이 없다면 많은 시간을 소모한다.

실례로 XEROX는 전 세계에 공급하는 복사기를 적재한 컨테이너에 수동형 RFID 태그를 부착하여 다른 트럭에 잘못 적재되는 경우에 경고를 함으로써 추가적 물류비용의 발생을 줄이고 있으며 FORD에서는 엔진 조립 공정의 정보가 실시간으로 생산 계획 시스템에 전송되어 효율적인 작업지시를 내릴 수 있게 되었으며 차량운송에서 컨테이너의 운송위치 등에 대한 즉각적인 정보 제공으로 비용과 시간을 절약하고 있다(이근호 2003).

인터넷 냉장고, 스마트 폰으로 대표되는 디지털 컨버전스, 유무선 통합 서비스, 방송과 통신 융합형 서비스 등의 정보통신 컨버전스, IT와 금융산업의 결합에 의한 모바일 뱅킹, IT와 자동차 산업의 결합에 의한 텔레매틱스 등의 산업 컨버전스가 유비쿼터스 컴퓨팅의 커다란 축으로 자리하고 있다.

V. 결 론

시맨틱 웹의 궁극적인 목적은 웹에 있는 정보를 컴퓨터가 좀 더 이해할 수 있도록 도와주는 표준과 기술을 개발하여 시맨틱 검색, 데이터 통합, 네비게이션, 타스크의 자동화 등을 지원하는 것이다. 시맨틱 웹에서 이러한 기능을 지원하기 위해서는 컴퓨터의 지능적인 정보처리가 가능하도록 웹 문서 내에 지식표현을 위한 온톨로지를 삽입하고 지식간의 관계를 설정하며 추론규칙을 포함시켜야 한다. 이를 통해서 운용 면에서 다른 기종 간에 호환성을 보장하고 사용자가 원하는 웹 서비스의 발견, 자동적인 웹 서비스의 실행과 동시에 웹 서비스들의 통합과 상호작용을 통하여 사용자가 원하는 정보를 찾고 더 나아가 추론이 가능하도록 해야 한다.

시맨틱 웹에서의 온톨로지 활용기술은 웹 서비스 관련 분야뿐만 아니라 생물 정보학, 자연언어 처리, 데이터베이스, 인공지능, 정보검색, 기계번역, 분산 시스템 등 다양한 분야에서 연구될 수 있을 것이다. 이처럼 온톨로지가 차세대 정보처리의 핵심기술로 부각된 이유는 먼저 웹이라는 거대한 지식 정보의 처리 방법과 현재 우리가 처한 수많은 지식 정보의 체계화 방법이 어느 정도 일맥상통 하며 사람의 사고방식과 동일한 의미정보를 컴퓨터에 담아 컴퓨터가 의미를 이해할 수 있도록 만든다는 매카니즘 때문이다.

본 논문에서는 Delphi 기법을 이용하여 현직 관광지 종사원들을 대상으로 2차에 걸친 설문을 통해 관광지를 지식체계화 하여 표현하였고 protege-2000을 사용하여 온톨로지를 구축해 보았다. 소규모 관광지에 Delphi 기법을 적용하는 것이 큰 의미는 없다 하더라도 Local Domain에서부터 합의를 도출하여 지식을 체계화하여 표현하고자 하는 취지에서 시도되었으며 그 결과 어느 정도 해당관광지 조직구성원 간에 합의된 지식표현을 얻어낼 수 있었고 그 결과를 토대로 구축된 온톨로지를 통해 단순 검색이 아닌 좀 더 지능적인 구문형 질의를 통해 현재의 해당관광지 홈페이지에서 얻을 수 있는 단순 정보 보다 좀 더 발전된 정보를 검색할 수 있었다.

이번 연구를 통해 지식표현을 위한 지식간의 관계를 설정하고 추론이 가능한 시스템을 설계한다면 시맨틱 웹이 현재의 웹 보다 자연어 처리에 더욱 효과적이며 유용하다는 결론을 얻을 수 있었다. 앞으로 제주 관광산업을 대상으로 한 시맨틱 웹 기반

의 온톨로지에 대한 연구가 계속 확대되어 발전된다면 텔레메틱스 시범도시로 지정된 제주도는 유비쿼터스 컴퓨팅의 초석이 될 수 있을 것이며 더 나아가 수많은 지식 정보의 저장소로 생각할 수 있는 웹을 중심으로 웹상의 지식 정보에 인간의 사고방식을 반영할 수 있는 잘 정의된 의미(semantic)를 지식체계화 한다면 현재 컴퓨터를 이용한 정보검색이 지능화될 뿐만 아니라 e-비즈니스 분야에도 인간의 사고방식과 같은 추론기능까지 갖출 수 있는 시스템이 개발될 수 있을 것이다.



參 考 文 獻

- 권혁철. 2004. “시멘틱웹의 가능성과 한계”. 지식정보인프라 통권15호. pp.15-19.
- 김세정. 1999. “RDF를 이용한 정보자원의 기술 및 KORMARC 스키마 개발에 관한 연구”. 석사학위논문. 연세대 대학원 문헌정보학과.
- 김이란. 2001. “온톨로지 의미정보를 이용한 RDF 스키마 생성에 관한 연구”. 석사학위논문. 연세대 대학원 문헌정보학과.
- 김흥기. 2002. “시멘틱 웹의 개념과 이해”. 마이크로소프트웨어. 2002. 4월호.
- 박재홍, 임유전, 김도완, 박찬규, 조현규. 2002. “Semantic Web 환경에서의 자원발견”. 정보처리학회 추계학술발표대회 논문집 제9권 제2호. pp.1-4.
- 양정진, 2003. “시멘틱 웹에서의 온톨로지 공학”. 정보과학회지 제21권 3호. pp.28-32.
- 오삼균. 2002. “시멘틱 웹 기술과 활용방안”. 정보관리학회지 제19권 4호. pp.297-319.
- 이근호. 2003. “유비쿼터스 비즈니스 모델개발과 기회창출전략”. ETRI 기업기술교육자료.
- 이종성. 1988. “텔파이 방법과 고등교육 연구”. 연세교육연구. 88-2, 연세대 교육연구소
- 정도현, 2003. “시소러스를 기반으로 한 온톨로지 시스템 구현에 관한 연구”. 석사학위논문. 연세대 대학원 문헌정보학과.
- 정인정, 박기식, 민재홍, 김성한, 이강찬. 2003. “시멘틱웹 기반의 검색 시스템 구조”. 주간기술동향 통권 1094호. pp.15-17
- 최중민. 2003. “시멘틱 웹의 개요와 연구동향”. 정보과학회지 제21권 제3호. pp.4-9.
- 최호섭, 옥철영, 2004. “정보검색 시스템과 온톨로지”, 정보과학회지 제22권 제4호. pp.62-63
- 제민일보. 2004. 6. 28
- 제민일보. 2004. 7. 20
- 제주관광신문. 2004. 6. 18
- 제주관광신문. 2004. 8. 27
- <http://www.cseric.or.kr>
- <http://www.ontology.or.kr>
- <http://www.kiss.or.kr>
- <http://www.itfinder.or.kr>

- Asuncion Gomez-Perez and Oscar Corcho. 2002. "Ontology Languages for the Semantic Web". IEEE Intelligent Systems. vol.17. no.1. pp.54-60.
- Berners-Lee, T., Hendler, J., Lassila, O., 2001. "The Semantic Web". Scientific American. May 2001
- Clyde W. Holsapple, K. D. Joshi. 2002. "A Collaborative Approach to Ontology Design". Communications of the ACM. vol.45. no.2 pp.42-47.
- Fensel, D., F. van Harmelen, I. Horrocks, D. L. McGuinness, and P. F. Patel-Schneider. 2001. "An Ontology Infrastructure for the Semantic Web". [online]. [cited: 2003. 4. 5].
<<http://www.cs.vu.nl/~frankh/postscript/IEEE-IS01.pdf>>
- Gruber, T. R. 1993. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". [online]. [cited: 2002. 12. 20].
<<http://www.cise.ufl.edu/~jhammer/classes/6930/XML-FA02/papers/gruber93ontology.pdf>>
- Guarino, N. 1998. "Formal Ontology in Information System". [online]. [cited: 2003. 4. 20].
<<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf>>
- Hjelm, J. 2001. "Creating the Semantic Web with RDF". John Wiley & sons, Inc.
- James Hendler. 2001. "Agents and the Semantic Web". IEEE Intelligent Systems. vol.16. no.2. pp.30-37
- Lassila, O., and R. R. Swick. 1999. "Resource Description Framework(RDF) Model and Syntax Specification". W3C Recommendation 22 February 1999. [online]. [cited: 2002. 10. 20].
- McGuinness, D. L., and F. van Harmelen. 2003. "Owl Web Ontology Language Overview". W3C Working Draft 31 March 2003". [online]. [cited: 2003. 4. 10]. <<http://www.w3c.org/TR/owl-features/>>
- S. A. McIlraith, T. C. Son, and H. Zeng. 2001. "Semantic Web Services". IEEE Intelligent Systems. vol.16. no.2. pp.46-53
- Smith, M. K., C. Welty, and D. L. McGuinness. 2003. "Web Ontology Language (OWL)Guide Version 1.0". [online]. [cited: 2003. 3. 20].

<<http://www.w3.org/TR/owl-guide/>>

<http://protege.stanford.edu>

<http://www.w3c.org>



[부록 1] 웹 온톨로지 모델링 - 삼촌 관계 표현하기

웹 온톨로지 언어를 이용하여 지식을 표현하다 보면 표현하기가 매우 까다롭거나 불가능한 것들이 종종 있다. 원인은 여러 가지가 있을 수 있으나 웹 온톨로지 언어의 표현력의 한계가 원인이 되는 경우가 많다. 웹 온톨로지로 표현하기 어려운 삼촌 관계를 살펴보면 다음과 같이 기술된다.

“삼촌” 관계는 “어떤 사람의 자식이 그 사람의 형제와 갖게 되는 관계”이다. 이 관계는 이행관계(Transitive Relation)와 유사한데 구조상 특이한 점은 서로 다른 두개의 속성이 이행에 개입한다는 점이다. ‘삼촌’ 관계에는 ‘자식’ 관계와 ‘형제’ 관계가 포함된다. 또한 이 두 관계가 중간에 하나의 개체를 매개로 연결된다.

즉, 하나의 매개체를 중심으로 그 매개체와 각각 ‘자식’과 ‘형제’ 관계를 갖는 개체들 간의 관계가 ‘삼촌’ 관계인 것이다. 그러나 이런 관계를 웹 온톨로지로 표현하는 것은 원칙적으로 불가능하다. 왜냐하면 ‘삼촌’ 관계는 클래스 간의 관계로 표현될 수 없기 때문이다. 물론 어떤 A라는 개체의 자식들의 클래스와 A의 형제들의 클래스를 선언할 수는 있다. 각각 ‘부모’ 속성의 값으로 A를 갖는 제약(Restriction) R1과 ‘형제’ 속성의 값으로 A를 갖는 제약 R2로 모델링 할 수 있다. 그러면 ‘삼촌’은 정의역(domain)을 R1으로 갖고 공역을 R2로 갖는 속성으로 정의될 수 있다. 그러나 이는 개체 A를 명시하고 있으므로 A와 다른 개체에 대해 ‘삼촌’ 관계를 정의하려면 별도의 클래스 정의와 속성 정의를 필요로 하게 된다. 이를 일반화하여 다음과 같이 ‘자식(Child)’ 클래스와 ‘형제(Brother)’ 클래스를 정의할 수 있다.

:Child

rdfs:subClassOf

[rdfs:type owl:Restriction; owl:onProperty :hasParent; owl:allValuesFrom :Person]

[rdfs:type owl:Restriction; owl:onProperty :hasParent; owl:maxCardinality “2”].

:Brother rdfs:type :Male; rdfs:subClassOf

[rdfs:type owl:Restriction; owl:onProperty :hasSibling; owl:allValuesFrom

:Person];

rdfs:subClassOf

**[rdfs:type owl:Restriction; owl:onProperty :hasSibling; owl:minCardinality
“1”].**

이제 삼촌 관계는 다음과 같이 정의될 수 있다.

:uncle rdf:type owl:ObjectProperty;

rdfs:domain :Child;

rdfs:range :Brother.

이 때 만약 다음과 같은 개체 정의가 주어졌다고 가정해 보자.

:Bob a :Male.

:John a :Male; :hasSibling Bob.

:Sam :hasParent :Bob.

:Jack a :Male.

:Nil a :Male; :hasSibling :Jack.

:Jim :hasParent :Jack;

위의 정의는 다음과 같이 전개될 수 있다.

Sam --- [hasParent] --- *Bob*

John --- [hasSibling] --- *Bob*

Jim ---[hasParent] --- *Jack*

Nil ---[hasSibling] --- *Jack*

Sam과 Jim은 위에 정의한 Child 클래스에 속하고 John과 Nil은 Brother 클래스에 속한다. 그렇다면 여기서 삼촌 관계는 uncle 속성의 정의에 따라 다음과 같이 유도해 낼 수 있다.

:Sam :uncle :John.

:Jim :uncle :Jack.

:Sam :uncle :Jack.

:Jim :uncle :John.

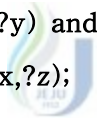
지금까지 정의한 웹 온톨로지로 보면 위의 네 문장 중에서 실제 ‘삼촌’의 의미를 따져 보면 마지막 두 문장은 참이 될 수 없다. 즉 지금까지 시도한 ‘삼촌’의 정의에 허점이 있음을 알 수 있다. 그 허점은 ‘형제’와 ‘부모’ 사이에 중간 매개 역할을 하는 매개체가 동일해야 하는데 이 제약사항이 기술되지 않았다는 점이다. 이와 같은 제약사항을 기술하려면 중간 매개체를 ‘형제’ 및 ‘부모’ 각 관계에 대하여 각각 변수로 표현하고 변수가 동일하다는 사실을 표현할 수 있어야 한다.

그러나 웹 온톨로지 언어는 변수를 기술할 수 없다. 웹 온톨로지 언어가 기반하고 있는 기술 논리(description logics)는 변수를 포함하지 않는 명제 논리(propositional logic)이기 때문이다. 그렇다면 삼촌관계를 표현할 수 있는 적절한 방법은 무엇인가?

해답은 바로 ‘웹 규칙 언어(web rules language)’이다. 삼촌관계도 이 규칙을 이용하면 다음과 같이 쉽게 기술할 수 있다.

rule Uncle is

**if :hasParent(?x,?y) and :hasSibling(?y,?z) and :Male(?z)
then :hasUncle(?x,?z);**



제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

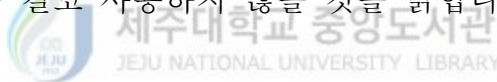
미천굴 관광지

웹 표현 체계를 위한 설문지(1차)

본 설문지는 미천굴 관광지를 웹상에서 좀더 효과적으로 표현하는 방안을 연구하여 관광객들이 인터넷에서 미천굴을 검색할 때 불필요한 정보는 걸러내고 필요한 정보만을 검출해 낼 수 있는 프로그램을 구축하고자 하는 취지에서 고안되었습니다.

실제 현장에서 근무하고 계신 여러분들이 평소 업무를 담당하시다가 느끼고 생각 한 바대로 솔직하게 작성해 주시기 바랍니다. 아울러 본 설문지에 작성한 내용과 개인정보에 관하여 현행 연구 이외에는 결코 사용하지 않을 것을 밝힙니다.

감사합니다.



2004. 10

제주대학교 경영대학원
경영정보과 강 한 섭

소 속 :

담당 업무 :

성 명 :

▣ 다음의 세가지 질문에 대해 자유롭게 답해주시기 바랍니다.

1. 미천굴 관광지를 자랑할 수 있는 특징적인 것 5가지를 중요한 순서대로 적어주십시오.

번호	특징적 요소 내용	비고
①		
②		
③		
④		
⑤		

2. 미천굴 관광지가 다른 관광지와 비교해 차별화되었다고 생각하는 것이 있다면 모두 적어주십시오.(인적, 물적자원을 모두 포함하여)

번호	차별적 요소 내용	비고
①		
②		
③		
④		
⑤		

3. 미천굴 관광지를 표로 다음과 같이 정리하여 보았습니다.
 귀하께서 다른 의견이 있다면 항목란을 추가해 주시고 각 항목별로
 포함되어야 할 세부내용을 중요한 것부터 적어 주십시오.

관광지명	항목	세부내용	비고 (설문대상부서)
미천굴	입장료		매표소
	단순관람		영업부
	체험관광		영업부, 아트센터
	편의시설		총무부
	조직구성		인사담당

☞ 설문에 응답하여 주셔서 감사합니다.

향후 1차 설문지를 취합한 결과를 정리하여 다시 2차 설문지를
 보내 드리도록 하겠습니다.

미천골 관광지

웹 표현 체계를 위한 설문지(2차)

안녕하십니까?

지난번 1차 설문조사는 직원 여러분의 적극 협조로 총 17부를 회수하였습니다. 1차 설문조사는 질문의 성격상 쉽게 대답하기가 어려웠던 항목도 있었다고 생각되지만 여러분의 정성어린 답변으로 소기의 성과를 거두게 되었습니다. 이점 깊이 감사드립니다.

이번 2차 설문조사는 1차 설문조사의 응답내용을 정리하여 제시하였습니다. 설문내용은 1차와 같습지만 제시된 1차 응답결과를 보시고 다시 한번 생각해서 답변해 주시기 바랍니다. 귀하의 생각은 얼마든지 바뀔 수 있다고 여겨집니다. 이번 조사에서는 응답의 범위를 점차 축소하여 합의를 지향하기위해 응답률이 전체 응답인원의 20%에 미달하는 항목은 생략하였습니다.

아울러 2차 설문조사의 내용과 개인정보 역시 현행 연구목적 이외에는 사용하지 않을 것을 밝힙니다.

감사합니다.

2004. 10

제주대학교 경영대학원
경영정보과 강 한 섭

소 속 :

담당 업무 :

성 명 :

▣ 다음의 세가지 질문에 대해 자유롭게 답해주시기 바랍니다.

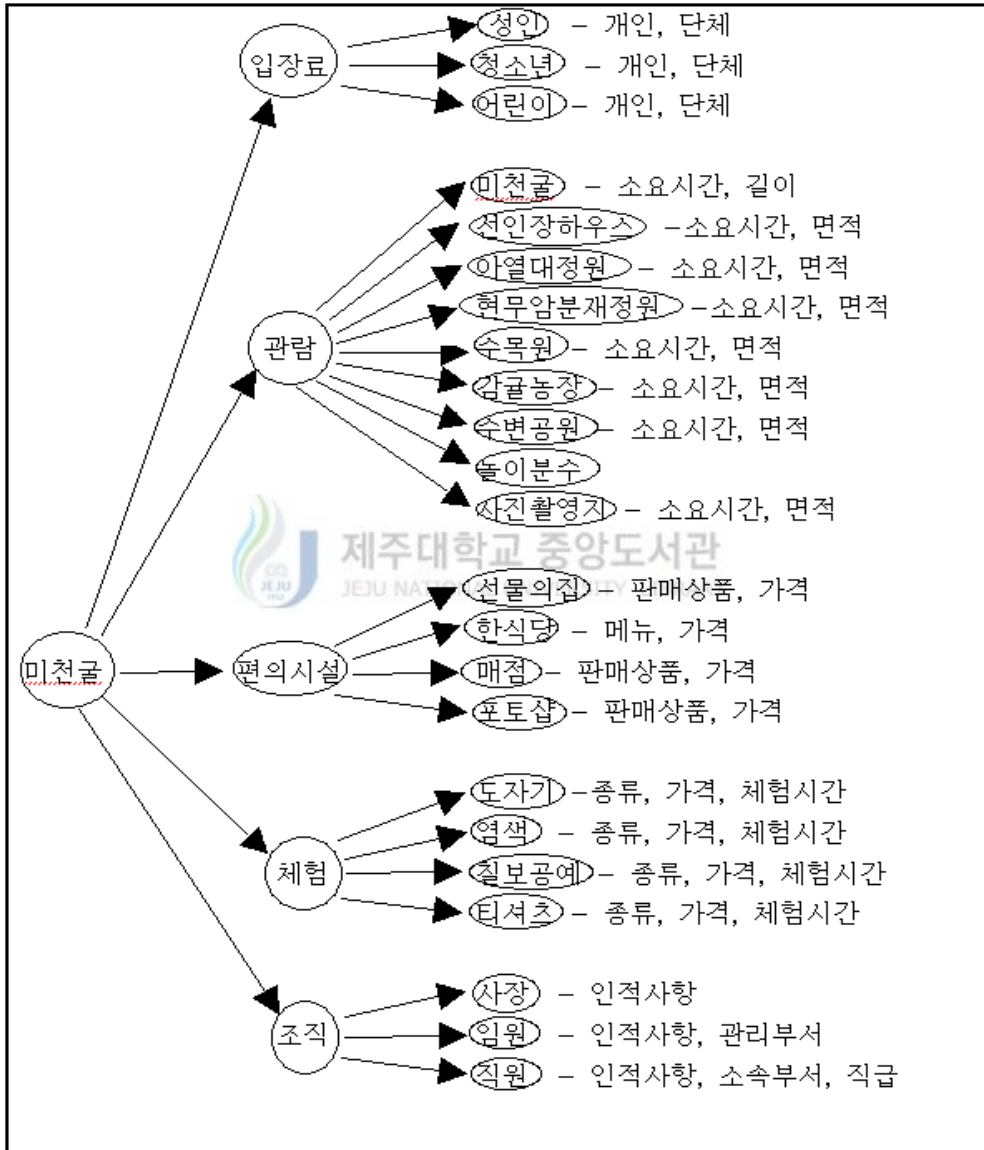
1. 아래 나열된 항목 중에서 미천굴 관광지를 자랑할 수 있는 특징적인 것들을 순위를 적어 표시해 주십시오.(1~8 까지 순위표시)

번호	특징적 요소 내용	1차응답결과(명)	2차 응답
①	미천굴	17	()
②	아트센터 체험	15	()
③	세다리 팽나무	13	()
④	선인장하우스	12	()
⑤	영화 캐릭터 사진촬영지	11	()
⑥	철쭉	10	()
⑦	바닥분수	8	()
⑧	현무암분재	5	()

2. 미천굴 관광지가 다른 관광지와 비교해 차별화되었다고 생각하는 것 중에서 귀하가 생각하는 순위를 적어 주십시오.

번호	차별적 요소 내용	1차응답결과(명)	2차 응답
①	아트센터 체험 프로그램	17	
②	직원 친절도 1위 관광지	13	
③	동양최대 손바닥 선인장(5M)	12	
④	동양최대 철쭉 식재(20만본)	10	
⑤	건물설계	6	

3. 1차 설문을 토대로 하여 미천골 관광지를 다음과 같이 그림으로 정리하여 보았습니다. 귀하께서 생각하시기에 추가해야 할 항목과 삭제해야 할 항목을 그림 안에 직접 표시하여 주십시오.



☞ 설문에 응답하여 주셔서 감사합니다.

ABSTRACT

A Study about Ontology design that use Delphi technique

Han-Seob Kang

Dept. of MIS

The Graduate School of Business Administration

Cheju National University



Continual attempts to accumulate and apply information eventually gave birth to the concept of the ‘Semantic Web’. Thus, the ‘Semantic Web’ can be defined as a product of mankind’s desire to standardize information. At the same time, the term provides ‘a method that standardizes mankind’s concept of linguistical expression’, and can be noted as an effort to combine such methods into a standard web environment that may materialize to form a catalogue. Term of knowledge is used with information or data in computer science. These are regarded and are divided sometimes each other in terminologies that have similar meaning. If is divided, knowledge is different from information and data in side that is structure anger which is purpose enemy conjunction. But, some kind of information in Knowledge Representation is called knowledge often if it can be expressed in computing system. Therefore, knowledge representation can talk soon as information representation.

The purpose of the study was systematized knowledge through knowledge representation that use Delphi technique and was designed ontology utilizing assistance editor called protege-2000 to construct a system based on the semantic web environment's ontology.

Level of interest regarding the construction and evaluation of search systems based on ontologies are set to increase. If defined well semantic that can reflect human's thinking to knowledge information on web laying stress on web that can think as storing place of great many knowledge information furthermore systematizes knowledge, search of information and comprehension about Jeju tour that use present computer may be done intelligence. As well as, systems that can equip to reasoning function such as human's thinking to e-business field may be developed.

