

# 관광산업에 있어서 온톨로지를 이용한 지식체계 적용

강한섭\*, 김민철\*\*  
제주대학교 경영정보학과  
{qadhaifi, mck1292}@cheju.ac.kr

## The application of knowledge system using Ontology at Tour Industry

Han-Seob Kang\*, Min-Cheol Kim\*\*  
Dept. of Management Information System, Cheju National University

### 1. 서 론

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

최근 많은 사람들에 의해 연구되고 있는 시멘틱 웹은 일반적인 웹과는 다르다. 일반적인 웹은 인터넷상에서 사용자에게 컨텐츠를 제공하는 정보 시스템을 말하는 반면에 시멘틱 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 정보를 접근하거나 게시할 수 있게 되었고 결과적으로 폭발적인 정보의 증가를 가져왔다[1].

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[2].

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 시멘틱 정보를 표현하기가 어렵고, 사람이 아닌 프로그램 또는 소프트웨어 에이전트가 자동으로 문서로부터 의미를 추출하기가 어렵다는 문제점을 해결하기 위해 1999년에 W3C(World Wide Web Consortium)에서 시멘틱 웹(Semantic Web)을 제안하였다[3]. Tim Berners-Lee에 의해 제안된 시멘틱 웹은 컴퓨터(Intelligent Software Agents)가 웹 상의 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[4].

본 논문에서는 Delphi 기법을 이용하여 제주도내 관광지를 지식체계화 하고 이를 통해 온톨로지를 구축하여 확장된 시스템으로 발전시키는데 그 목적이 있다.

본 논문의 구성은 다음과 같다. 2절에서는 연구의 내용과 방법을 알아보고 3절에서 지식표현을 위한 온톨로지의 구축에 대한 이론적 배경을 기 발표된 논문을 통해서 소개하였으며 4절에서는 두 차례에 걸친 설문결과를 분석한 후에 protege-2000을 이용하여 실제 온톨로지를 구축해 보고 5절에서 말을 맺는다.

## 2. 지식표현을 위한 온톨로지 구축

### 2.1 지식표현의 정의[6]

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 불린다. 곧, 지식 표현은 곧 정보 표현(information representation)이다.

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 '지식 표현'이란 용어를 고전적 인공지능(기호주의, symbolic AI)의 기호 처리 시스템 가설 아래에서 이해하는 것이 일반적이다. 즉, 모든 지능적인 시스템은 내부의 기호들(symbols)을 써서 세계에 관한 특정 지식들을 표상 혹은 인코딩한 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).  
(A KR is a Set of Ontological Commitments.)
- 지식 표현은 존재론적 개입들의 집합이다.  
(A KR is a Fragmentary Theory of Intelligent Reasoning.)
- 지식 표현은 지능적 추리의 단편적 이론이다.  
(A KR is a Medium for Efficient Computation.)
- 지식 표현은 사람들의 표현의 매개체이다.  
(A KR is a Medium of Human Expression.)

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

#### 첫째, 논리기반 지식표현

지금까지 고안된 다양한 논리를 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 추론에

도 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 정리 증명기(theorem prover)이다.

#### 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 대상 영역내 구체적 대상들을 나타내는 노드들이다. 또한 Semantic Network에서는 간선(edge)을 보통 '아크(arc)'라 하는데, 이것은 부분집합 아크들 혹은 isa 연결들(links), 집합구성원 아크들 혹은 사례(instance) 연결들, 함수 아크들로 구분되어 진다.

#### 셋째, Frame 기반 지식 표현

프레임은 프레임 이름과 속성-속성값 순서쌍들의 집합의 쌍으로 볼 수 있다. 의미망의 관점에서 볼 때, 프레임 이름은 의미망의 노드에, 속성은 이 노드와 연결되는 아크에, 그리고 속성값은 이 아크의 다른 끝 쪽에 있는 노드에 대응한다. 이러한 속성-속성값의 순서쌍을 보통 슬롯(slot)이라고 하고, 속성을 슬롯 이름(slot name), 속성값을 슬롯값 혹은 슬롯 채우기(slot filler)라고 한다.

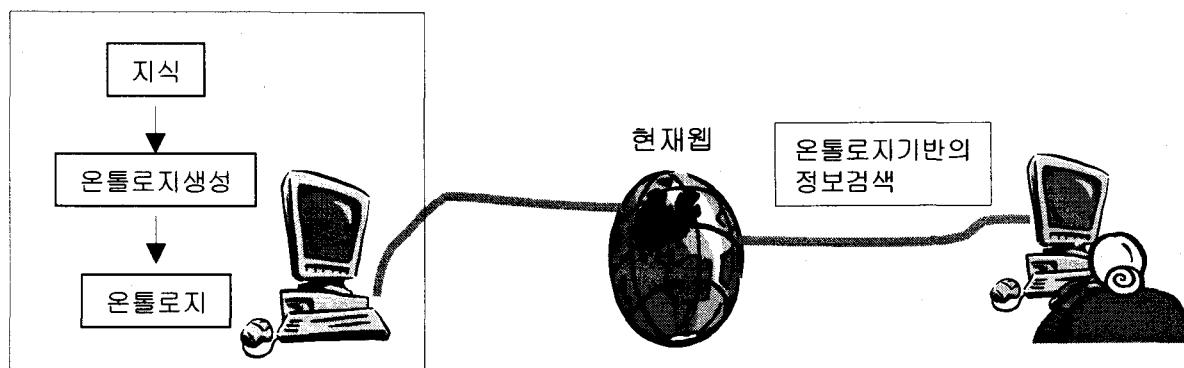
프레임 체계에서는 서로 관계있는 지식들이 그들 사이의 계층 구조에 따라 조직화되므로, 일단 표현된 지식들은 사람이 보기에 자연스러우며 따라서 그 처리가 용이하다. 반면, 실세계의 복잡한 대규모 지식들을 프레임 구조로 실제로 조직화하는 작업은 매우 힘들고 많은 노력과 비용을 요구한다

### 2-2 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 번거러움을 준다. 이는 HTML이 의미적인 검색이 이루어지지 않기 때문이다. 따라서 시멘틱 웹 환경에서 온톨로지를 기반으로 의미적 검색을 한다면 이런 불편을 줄일 수 있을 것이다. <그림 4>에 온톨로지가 존재하는 웹 서버의 시스템 구조를 그림으로 표시하였다. <그림 4>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[7].

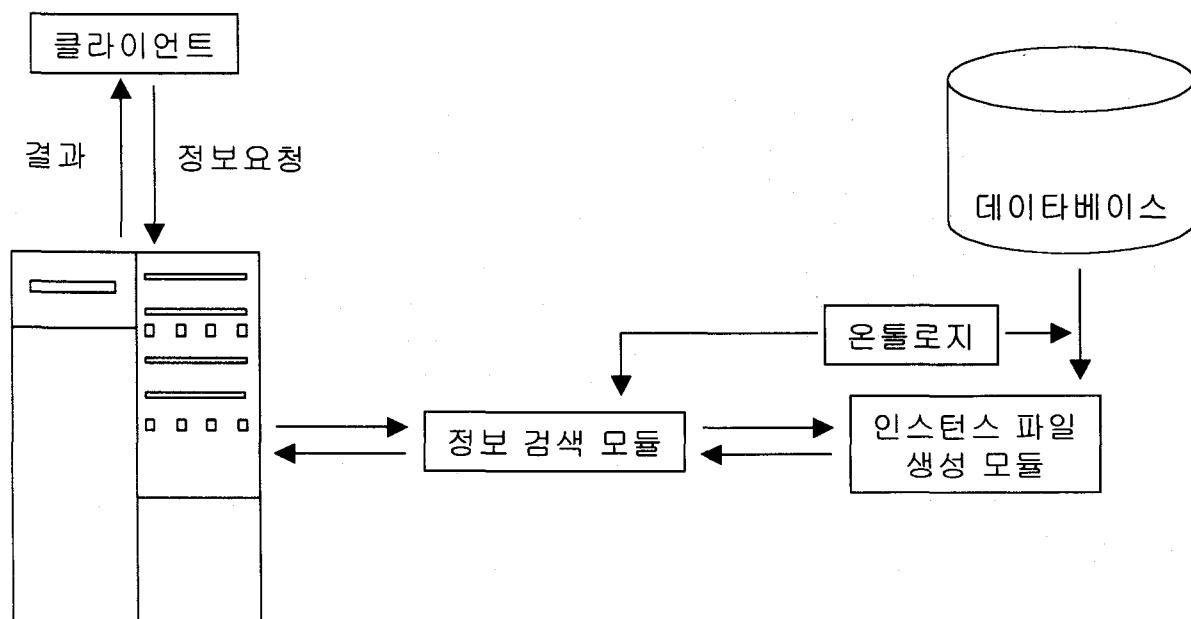
<그림 1> 시스템 구조

온톨로지가 존재하는 웹 서버



<그림 2>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 2> 시스템 설계



### 2.3 온톨로지 설계 방법

온톨로지 공학은 원칙적인 설계, 수정, 응용, 평가 단계를 거쳐 개발된다. 이러한 온톨로지 설계 방법은 다음과 같이 다섯 분류로 나누어 볼 수 있으며 이들 간의 혼합적인 방법도 가능하다[8].

<표 1> 온톨로지 설계 방법

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

첫째, 영감적 방법은 개발자가 개인적 관점에서 관심 분야의 독특한 상상이나 독창성을 이용해 혁신적인 온톨로지를 개발할 수 있으나 이론적 바탕이 부족하고 비현실적인 단점이 있다.

둘째, 귀납적 방법은 관심분야의 특정한 case를 관찰, 조사, 분석하여 얻어진 특징들을 그 분야의 다른 case에 일반화하여 적용시키는 방법이지만 일반성을 결여하기 쉬운 단점이 있다.

셋째, 연역적 방법은 특정 case를 도출하기 위해 선택할 수 있는 일반성을 지닌 스키마가 존재한다는 가정하에 일반적인 원칙들로 이루어진 온톨로지를 특정한 case에 맞춰 점차적으로 수정해 가는 방법이다.

넷째, 통합적 방법은 이미 존재하는 온톨로지들로부터 부분적으로 관심분야를 나타내는 부분들을 합성하여 하나의 단일화된 온톨로지를 구성하는 방법으로 위의 세가지 방법을 통합적으로 적용하지만 개발자 개인의 해석이나 통합 기술에 의존적이다.

다섯째, 협력적 방법은 초기 온톨로지를 기준으로 하여 여러 사람이 함께 협력하여 개발하는 방법으로 다수의 관점을 결합하여 지속적인 평가와 제안을 통해 온톨로지를 개발하지만 개발자들의 계속적인 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.

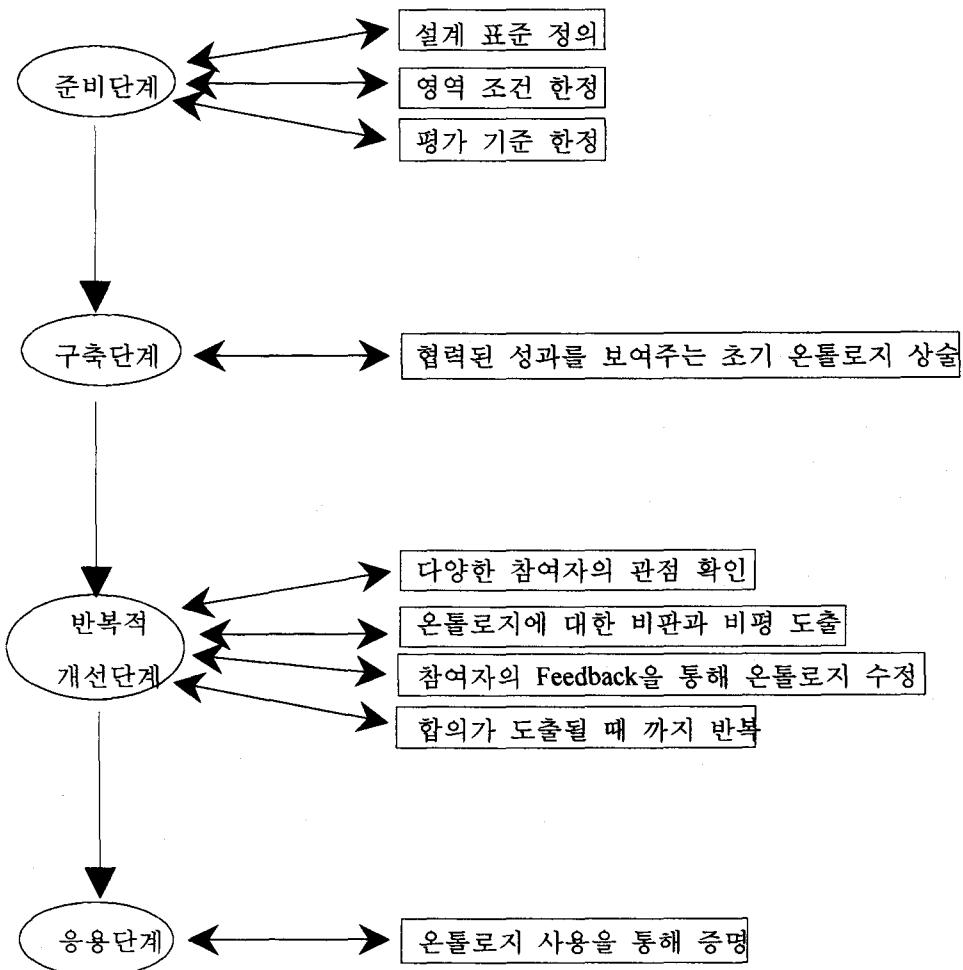
<그림 3>는 조직 내에서 지식관리 행위를 위한 온톨로지 개발을 위해 사용되는 협력적인 Delphi 지향적 구조를 나타내는 온톨로지 설계단계를 예시하고 있다.

준비단계는 온톨로지를 위해 설계 기준들을 정의하고 영역 조건들을 열거하며 평가 기준들을 결정하는 단계이다. 지식관리 온톨로지를 위해 선택한 설계 기준들은 포괄적이고, 정확하고, 간결하고, 명쾌하고 유용하다. 이 기준들은 온톨로지 개발을 안내하고 온톨로지의 성공 정도를 평가하는데 중요하다. 비즈니스 영역, 記述的 영역, 세부 영역의 세가지 영역 조건에 설계기준을 적용시킨다. 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. 설계 기준들이 평가 기준에 만족할 때 수정을 반복한다. 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.

<그림 3> 협력적 온톨로지 설계방법[9]



## 2.4 온톨로지 개발과정

소프트웨어 공학 분야에서 지식과 소프트웨어의 재사용(reuse)은 프로젝트의 개발기간을 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 이종성, 어휘적(lexical)/의미적(semantic) 문제와 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 기준에 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(ontologists 혹은 ontology developers) 온톨로지의 개발 목표와 단계에 따른 적절한 방법론을 적용하고 보완해갈 수 있는 과정이 부가적으로 요구된다. 온톨로지를 만들기 위한 과정은 Noy & McGuinness의 Ontology Development 101([http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html))에 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

#### 4. 온톨로지 구축

#### 5. 결론

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 시멘틱 웹이다.[9]

시멘틱 웹 환경에서는 사용자가 필요로 하는 'What'에 대한 내용을 컴퓨터에 알려주고, 컴퓨터는 입력된 조건에 맞는 작업을 수행한 후 사용자에게 제공하는 것이다. 이때 사용자는 'How'(어떻게 작업해야 하는지)는 알 필요가 없으며 이 'How'의 문제는 컴퓨터가 담당하게 된다.

시멘틱 웹의 궁극적인 목적은 웹에 있는 정보를 컴퓨터가 좀 더 이해할 수 있도록 도와주는 표준과 기술을 개발하여 시멘틱 검색, 데이터 통합, 네비게이션, 타스크의 자동화 등을 지원하는 것이다. 시멘틱 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.

시멘틱 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 있

도록 만든다는 매카니즘 때문이다.

본 논문에서는 Delphi 기법을 이용하여 관광지를 지식체계화 하여 표현하였고 protege-2000을 사용하여 온톨로지를 구축해 보았다. 이를 제주의 관광산업, 예를 들어 전 관광지, 숙박업소, 렌터카, 식당 등을 대상으로 확대하는 연구들이 계속된다면 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 제주 관광에 대한 정보검색이나 이해 등이 지능화될 뿐만 아니라 e-비즈니스 분야에도 인간의 사고방식과 같은 추론기능까지 갖출 수 있는 시스템이 개발될 수 있을 것이다.

#### [참 고 문 헌]

- [1] 최중민, “시멘틱 웹의 개요와 연구동향”, 정보과학회지 제21권 제3호, pp.4-9, (2003)
- [2] 정인정, 박기식, 민재홍, 김성한, 이강찬, “시멘틱웹 기반의 검색 시스템 구조”, 주간기술동향 통권 1094호, (2003) <<http://www.itfinder.or.kr>>
- [3] S. A. McIlraith, T. C. Son, and H. Zeng. 2001, "Semantic Web Services", IEEE Intelligent Systems, vol.16, no.2, pp.46-53, (2001)
- [4] 박재홍, 임유전, 김도완, 박찬규, 조현규, “Semantic Web 환경에서의 자원발견”, 정보처리학회 추계학술발표대회 논문집 제9권 제2호, pp.1-4, (2002)
- [5] 이종성, “델파이 방법과 고등교육 연구”, 연세교육연구. 88-2, 연세대 교육연구소, (1988)
- [6] Dent, C.L. and Cobble, M.H.: 4th ICTEC, 7, (1982)
- [7] 양정진, “시멘틱 웹에서의 온톨로지 공학”, 정보과학회지 제21권 3호, (2003)
- [8] Clyde W. Holsapple, K. D.. Joshi, "A Collaborative Approach to Ontology Design", Communications of the ACM, vol.45, no.2, pp.42-47, (2002)
- [9] 권혁철, “시멘틱웹의 가능성과 한계”, 지식정보인프라 통권15호, pp.15-19, (2004)