

## 국가 IT 온톨로지 구축

김재호<sup>o</sup> 신지애\* 최기선  
한국과학기술원 전산학과/시맨틱웹첨단연구센터,  
한국정보통신대학교\*

jjaeh<sup>o</sup>@world.kaist.ac.kr, jjae@icu.ac.kr, kschoi@world.kaist.ac.kr

### National IT Ontology Construction

Jae-Ho Kim<sup>o</sup>, Ji-Ae Shin\* Key-Sun Choi  
KAIST CS Dept. / SWRC  
Information and Communications University\*

#### 요 약

본 논문은 2006년부터 시작된 "국가 IT 온톨로지 인프라 기술개발" 과제의 온톨로지 구축 부분을 소개한다. 이 과제는 2006년부터 2011년까지의 5년 과제로 산학연이 참여하여, 국가 IT 분야에 범용적으로 활용이 가능한 IT 온톨로지를 구축하고 인터넷, 인트라넷, 유비쿼터스 환경에서 제공되는 각종 IT 서비스에 적용하여 seamless 서비스를 제공하는 것을 목표로 하고 있다. 이 온톨로지는 한-영 2개의 언어로 제작되며, 국제표준 언어인 OWL을 사용하여 국내외적으로 널리 사용될 수 있는 대용량 IT 온톨로지를 목표로 한다.

#### 1. 서론

온톨로지는 특정 분야에서 사람들에 의해 공유되는 개념의 형식적 명세라고 일반적으로 정의<sup>1</sup>된다. 온톨로지는 의미의 기본 단위의 개념과 개념들 사이의 관계들로 표현되는데, 이렇게 표준화된 의미 단위와 표현방법을 사용함으로써 사람과 컴퓨터, 컴퓨터와 컴퓨터가 의미의 애매성 없이 정보를 해석, 공유할 수 있다.

온톨로지의 대표적 활용 분야로는 시맨틱웹[2]이 있다. 현재의 웹은 많은 정보가 가지고 있지만, 의미기반으로 되어 있지 않기 때문에 기계가 읽고 해석할 수 없어 의미 수준의 정보 추출은 사람이 할 수밖에 없다. 그러나 현재의 웹 정보를 의미에 기초하여 표현하면 이형질의 다양한 정보들을 기계가 처리, 추론하여 사람의 개입 없이도 정보의 의미적 통합이 가능해진다. 그러므로 온톨로지는 시맨틱웹을 실현하는 데 필수 요소이다.

그 외에도 온톨로지는 자연 언어 처리, 정보 검색 및 추출, 정보 시스템, 지식 관리 시스템, 기업 모델링, e-비즈니스 등 다양한 분야에서 활용되고 있다.

Mills-Davis의 Project10X 2006년 보고서에 의하면 온톨로지를 활용하는 시맨틱 시장은 2006년부터 2010년 사이에 10배의 성장하고, 2010년부터 2015년까지도 10배가 성장하여 500조원 규모가 될 것으로 예측하고 있다[3]. 이러한 다양한 응용의 범위와 시장성을 감안하여, 본 연구

팀과 산학연 공동연구기관<sup>2</sup>에서는 IT 분야에 범용적으로 활용이 가능한 국가 IT 온톨로지를 구축하고 인터넷, 인트라넷, 유비쿼터스 환경에서 제공되는 각종 IT 서비스에 적용하여 seamless 서비스를 제공하는 "국가 IT 온톨로지 인프라 기술 개발" 과제<sup>3</sup>를 2006년부터 시작하게 되었다. 이 과제는 국내 IT 분야뿐만 아니라 국제적으로도 표준적으로 활용할 수 있는 온톨로지 구축을 목표로 한다.

본 논문에서는 이 과제에서 핵심인 국가 IT 온톨로지 구축 부분을 주로 소개하고자 한다. 본 과제에서 온톨로지를 구축하기 위해 사용하는 접근방법은 다음과 같다. 서비스 요구 사항에 적합한 실용화 가능한 온톨로지를 설계하고, 공유/재사용이 가능한 구조를 정립한다. 기 구축된 부분적 온톨로지를 융합하는 방식과 텍스트마이닝 기술을 이용하여 온톨로지를 확장하는 방법을 혼합하여 온톨로지를 구축한다. 그리하여 세계적으로 전례가 없는 대용량 IT 분야 온톨로지를 구축하고, 대용량 말뭉치를 활용하여 개념, 개념간의 관계, 속성 등을 자동으로 추출하는 기술을 개발하고자 한다. 표준화를 위하여 온톨로지는 표준화된 온톨로지 표현 언어인 OWL로 기술한다.

#### 2. 관련연구

온톨로지의 중요성이 커지면서 세계적으로 온톨로지를 구축하는 사례가 늘어나고 있다. 특히, 모든 온톨로지의 기

<sup>1</sup> Gruber의 정의[1]

<sup>2</sup> KAIST, 경북대학교, 숭실대학교, ETRI, ㈜솔트룩스, ㈜코리아와이즈넷

<sup>3</sup> <http://coreonto.kaist.ac.kr>

반이 될 수 있는 상위 온톨로지에 대한 연구는 많이 진행되어 있다. 대표적인 상위 온톨로지로는 IEEE Standard Upper Ontology Working Group (SUO WG<sup>4</sup>) 에서 창조하고 있는 SUMO (Suggested Upper Merged Ontology) [4], DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [5], Cyc Upper Ontology [6]가 있다. SUMO는 Teknowledge 사<sup>5</sup>에서 개발 중인 상위 온톨로지. 추상적이고 광범위한 개념들을 포함한 685개의 클래스로 구성되어 있다. 데이터 상호연동, 정보 탐색 및 검색, 자동 추론, 자연언어 처리 등 광범위한 목적으로 사용되고 있으며, GNU 라이선스 하에서 자유롭게 사용할 수 있다.

유비쿼터스 환경에 적합한 표준 온톨로지로는 상황인지 온톨로지인 SOUPA (Standard Ontology for Ubiquitous and Pervasive Ontology) [7], CoBra (Context Broker Agent Ontology) [8] 등이 개발되어 프로토타입 수준의 지능화된 개인화 서비스로의 활용 연구가 활발히 진행되고 있다. SOUPA는 유비쿼터스 환경에서의 응용을 위해 개발된 OWL 온톨로지. 코아 온톨로지를 기본으로 특정 응용을 위한 확장 온톨로지 구성되어 있다. CoBra는 SOUPA의 온톨로지 상에서의 확장으로 Privacy 표현도 포함하고 있다. 또한 코아 온톨로지로서 FOAF (Friend of A Friend) [9], DAML-Time (DARPA Agent Markup Language-Time) [10]이 개발되어 활발히 이용되고 있다.

온톨로지 구축에 대한 국내 기술은 시소러스 수준에 머물러 있는 실정에서 다소 벗어나, 학계와 관련 산업체 일부에서 구축을 시도하고 있으나, 소규모의 특정 도메인의 온톨로지를 수동으로 개발하는 단계에 머물고 있는 실정이다.

KISTI에서는 2005년에 최상위 7개 클래스를 포함한 89개 클래스로 구성된 과학기술 분야 R&D 관련 기반 정보의 모델링을 위한 OWL 기반 온톨로지를 구축하였으며, 2006년부터는 이 온톨로지와 연계된 학술정보 온톨로지도 구축하고 있다 [11]. 울산대에서는 기구축된 면역학 시소러스를 기반으로 개념 노드 4,703개의 면역학 도메인 온톨로지를 개발하였고 [12], KT에서는 영화에 관련된 시맨틱 웹 서비스를 제공하기 위하여 영화정보 온톨로지를 구축하고 있다 [13].

### 3. “국가 IT 온톨로지 인프라 기술 개발” 과제의 구성

본 과제는 국가 IT 분야에 범용적으로 활용이 가능한 코아 온톨로지와 이를 기반으로 확장된 분야별 온톨로지를 정의하고, 효과적으로 국가 IT 온톨로지를 구축하기 위한 핵심 인프라 기술 및 IT 온톨로지 기반 응용 시스템을 개발하는 것이 목표이다. 2006년부터 2011년까지의 5년 과제로, 단계별로 기반, IT 코아, IT 디바이스, IT 서비스, IT 라이프 온톨로지를 구축하며, 이 온톨로지는 온톨로지 포털, 시맨틱 서비스 포털, 유비쿼터스 컴퓨팅 서비스 등의 지능형 서비스에 활용된다 (그림 1).

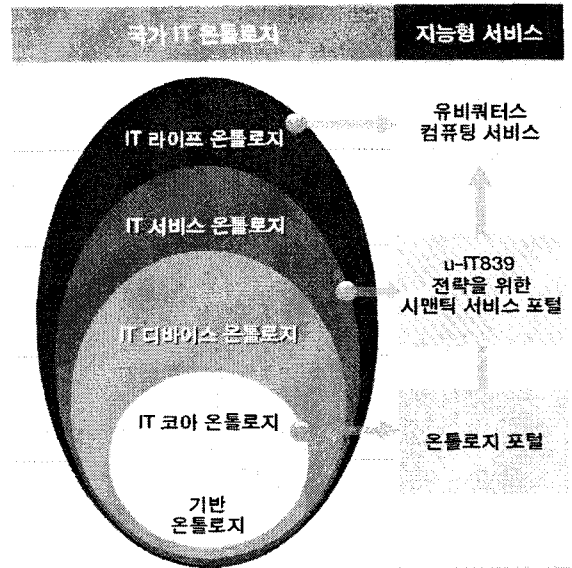


그림 1 과제의 최종 목표

이 과제는 다음과 같은 세부 과제들을 포함한다:

국가 IT 온톨로지 구축: IT839 전략 서비스<sup>6</sup>에 공통적으로 적용되는 IT 온톨로지를 구축한다. 이 온톨로지에는 IT 관련 개념 및 인스턴스에 대한 계층이 나타나 있고, 모든 개념관계 및 속성이 기술되어 있다. 인스턴스를 자동으로 찾아 해당 개념에 넣어주는 인스턴스 생성 소프트웨어 개발, 온톨로지를 이용하여 문서 내 개념 및 인스턴스에 의미 부여하는 시맨틱 태깅도 포함된다.

온톨로지 기반 추론: 구축한 국가 IT 온톨로지를 지능형 서비스에 이용하기 위한 실시간 추론 엔진을 개발한다. OWL (Web Ontology Language)로 표현한 IT 온톨로지의 개념, 개념들 간의 Rule 관계를 SWRL (Semantic Web Rule Language)을 기반으로 규칙을 표현하여 추론에 이용한다. 추론을 통해 온톨로지에 표현되지 않은 정보를 추론에 의해 만들어 이용할 수 있다.

지능형 서비스 시스템 개발: IT 온톨로지 기반으로 주제별 커뮤니티를 형성하고 주제 관련 정보가 집적/공유될 수 있는 서비스를 제공하는 지식포털 서비스 기술을 개발한다. 그리고 다양한 유비쿼터스 환경 지원, 상황인지 미들웨어들을 국가 IT 온톨로지 인프라에 기반하여 구축하고, 상호 연동/통합할 수 있는 유비쿼터스 서비스 브로커 시스템도 개발한다.

국가 IT 온톨로지는 지식의 기본 개념과 속성을 정의할 수 있는 “기반 온톨로지”, IT 분야에서 공통적이고 기반이 될 수 있는 “IT 코아(Core) 온톨로지”, IT 분야에서 사용되는 모든 디바이스에 관한 “IT 디바이스(Device) 온톨로지”.

<sup>4</sup> <http://suo.ieee.org>

<sup>5</sup> <http://ontology.tekknowledge.com>

<sup>6</sup> 차세대 이동통신, 홈네트워크, 디지털 TV/방송, 텔레매틱스, 지능형로봇, 차세대 PC, 임베디드 SW 등

IT839 전략에서 제공되는 IT 서비스에 관한 “IT 서비스 온톨로지”, 미래 유비쿼터스 환경에서 누릴 수 있는 IT 생활에 관한 “IT 라이프(Life) 온톨로지”를 포함한다. 그림 2는 국가 IT 온톨로지의 개념도를 나타낸다.

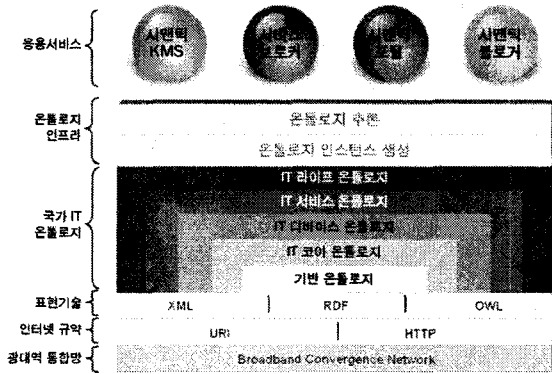


그림 2 국가 IT 온톨로지의 개념도

유비쿼터스 컴퓨팅 환경에서 단절없고 개인화된 자동화 서비스(seamless personalized automated service)를 제공하기 위해서는 지능형 에이전트가 필수적이고, 차세대 웹 환경에서 지능형 에이전트가 동작하기 위한 필수적인 소프트웨어 인프라가 국가 IT 온톨로지이다. IT839 전략을 진행하는 과정에서 지능적 서비스를 제공하기 위해서 각 분야에서 독자적으로 개발하고 있는 온톨로지를 표준화할 필요가 있는데, 본 과제에서 구축하는 국가 IT 온톨로지는 IT839 전략 각 분야에서 공통적으로 활용될 수 있는 지식 형태로 표현되기 때문에 표준화 및 상호 운용성에 중요한 역할을 할 것으로 기대된다.

#### 4. 온톨로지 구축 방법

IT 온톨로지를 체계적으로 구축하기 위하여 5-계층으로 된 온톨로지 구축 단계를, 기본 온톨로지부터 IT 라이프 온톨로지까지 순차적으로 각 온톨로지를 개발한다(그림 3). 각 단계는 온톨로지를 구성할 용어를 추출하는 **【용어 선정 단계】**, 용어를 같은 의미를 가진 개념 단위로 묶어주는 **【용어 개념화 단계】**, 개념 사이의 관계를 맺어주는 **【개념간 관계설정 단계】**, 구축된 기존온톨로지와 새로 구축한 온톨로지를 통합하는 **【기존 온톨로지 통합 단계】**, 구축된 온톨로지를 검증하는 **【온톨로지 검증 단계】**로 이루어져 있다.

IT 온톨로지 설계: 온톨로지 구축에 들어가기에 앞서, 구축할 온톨로지가 갖춰야 할 모습을 설계하고 구축범위를 결정하는 과정이 필요하다. 국가 IT 온톨로지는 IT 분야뿐만 아니라 일반 분야에 적용 가능한 온톨로지가 되어야 하므로, Object, Process, Schema, Script 등의 12가지 Top-level category[14]와 Attribute, Time, Space 등의 기본 요소를 갖추도록 설계한다.

IT 전 분야와 일반분야를 포괄하는 분야 독립적 상위 개념 온톨로지에 앞으로 구축되는 이동통신, 디지털 TV 등의 IT 세부분야의 부분적 온톨로지들을 붙여나갈 수 있도록 한다.

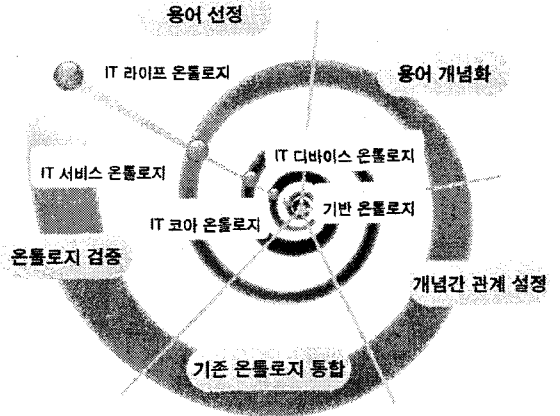


그림 3 온톨로지 구축 방법

**용어 선정:** 온톨로지를 구성할 용어를 수집, 선정하는 단계이다. 대상 용어는 정보통신용어사전<sup>7</sup>과 같은 IT 관련 모든 사전을 모아 수집한다. 또는 IT 관련 웹 신문 기사나 IT 특허 문서와 같은 말뭉치에서 추출한 통계적 정보를 이용하여 IT 분야의 용어를 추출한다. 추출한 용어는 수작업에 의한 검수를 통하여 교정 및 최종 선정을 하며, 이태한-영 대역어, 약어, 한국어/영어 정의문 등의 부가정보도 웹과 사전을 이용하여 같이 수집한다. 영어 대역어와 영어 정의문은 국가 IT 온톨로지를 국제 표준으로 만들기 위해 필요한 항목이며, 정의문은 용어를 정의하고 용어간 관계 설정이나 속성 정의를 위해 필요한 정보이다.

**용어 개념화:** 용어 개념화는 용어를 그 성질, 공통성, 본질에 착안하여 파악하는 과정으로, 본 과제에서는 용어들을 분류/군집화하여 같은 의미를 가지는 개념(클래스)으로 묶어 정의하고 그들 사이의 계층관계를 구축하는 단계까지를 포함한다. 용어 그 자체가 개념명이 될 수도 있지만, 여러 용어가 하나의 개념으로 묶이는 경우에는 그 개념을 대표하는 개념명을 정의할 필요가 있다.

개념 사이의 계층관계를 구축할 때는 어휘-구문 패턴, 정의문에 등장하는 패턴, 용어 구성 단어 사이의 수직 관계를 이용하는 규칙기반 방법과 문맥에 기반한 통계 기반 방법을 조합하여 구축한다.

**개념간 관계 설정:** 온톨로지의 개념들 사이에는 많은 관계가 존재하는데, 이들 사이의 유용한 관계를 찾아 맺어주는 단계이다. 가능한 모든 관계를 찾아 맺어 줄 경우 방대한 관계 수로 인하여 추론이 어려워져 맺은 관계들이 쓸모 없게 되기 때문에 설정할 관계를 먼저 정의하는 과정이 필요

<sup>7</sup> <http://word.tta.or.kr/index.jsp>

하다.

관계를 정의하기 위하여 기존 상위 및 도메인 온톨로지에서 사용한 관계를 검토하여 이용하는 방법과 텍스트에서 나타나는 개념 분포에 대한 분석을 통하여 IT 분야에서 필요한 기본 개념 관계를 정의하는 방법을 함께 이용한다. 기본적으로 텍스트에서 관계를 추출하는 방법으로는 문장에서 (주어, 목적어, 동사) 관계에 대한 패턴을 추출하고 이것의 확률적 분포를 이용하여 온톨로지에 나타나는 non-taxonomic 관계들을 후보로 추출하는 방법이 있다.

그 외에도 정의문이나 말뭉치에서 정보를 추출하여 개념의 속성을 정의하고, 개체명 인식 기술을 활용하여 추출한 인스턴스와 그 속성을 개념에 연결하는 과정도 이 단계에 포함된다.

기존 온톨로지 통합: 앞 단계에서 구축된 도메인 온톨로지를 기존의 상위 온톨로지에 통합하는 단계이다. 예를 들어, 개념명의 어휘 비교(string matching)를 통하여 통합할 도메인 온톨로지의 개념 중 통합대상 온톨로지에서 일치하거나 비슷한 개념을 찾아 통합한다. 또는 구조가 서로 달라 구조가 통합이 어려운 경우에는 통합하려는 두 대상의 구조를 변경하여 합치는 방법도 있다.

온톨로지 검증: 마지막으로, 온톨로지가 제대로 구축되었는지와 온톨로지의 유용성을 검증하는 단계이다. 검증, 평가되어야 하는 측면으로는, IT 분야에서 자주 사용되는 개념과 관계들의 포괄성, 개념들의 계층구조의 타당성, 응용분야에서의 효율성 등을 들 수 있다. 구체적으로는, OntoClean[15]과 같은 방법을 이용하여 속성의 변질 없이 개념의 계층관계가 제대로 구축되었는지를 검증할 수 있으며, 온톨로지를 이용한 추론을 통하여 질의응답 시스템에 활용하거나, 시맨틱 포털과 같은 응용서비스에 적용하여 그 유용성을 증명할 수 있다.

## 5. 결론

온톨로지에 대한 인식이 커지면서 전세계적으로 온톨로지의 구축과 응용에 대한 연구가 활발하게 진행되고 있다. 그러나 현재까지 구축된 온톨로지들은 대부분 분야 전문가의 도움으로 수작업으로 구축되어 특정 분야의 소규모 온톨로지가 대부분이다. 특히 IT 분야에 대해서 구축된 대규모 온톨로지는 전무한 실정이다.

본 논문에서 소개한 국가 IT 온톨로지가 구축되면 이는 세계 최초의 대용량 IT 온톨로지라는 큰 의미를 가지게 된다. 더 나아가 OWL을 사용하여 한-영 2개의 언어로 제작하여 국제 표준 온톨로지로서 국내뿐만 아니라 세계적으로 널리 사용될 기반을 마련하고자 한다.

## Acknowledgement

본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도 기반기술개발사업의 연구결과로 수행되었습니다.

## 참고문헌

- [1] Gruber, T. R., "A Translation Approach to Portable Ontology Specifications," Knowledge Acquisition, Vol. 5, No. 2, pp.199-220, 1993.
- [2] Berners-Lcc, T., Hendler, J. and Lassila, O., "The Semantic Web," Scientific American, 2001.
- [3] Mills Davis, Dean Allemang and Robert Coyne, "Evaluation and Market Report," Top Quadrant, 2004.
- [4] Suggested Upper Merged Ontology (SUMO)  
<http://ontology.teknowledge.com/>
- [5] Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)  
<http://www.isib.cnr.it/infor/ontology/DOLCE.html>
- [6] Cyc Upper Ontology  
<http://www.openccvc.org/>
- [7] Harry Chen, Filip Perich, Tim Finin, Anupam Joshi, "SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications," In Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004.
- [8] CoBra (Context Broker Agent Ontology)  
<http://cobra.umbc.edu/index.html>
- [9] FOAF (Friend of A Friend)  
<http://www.foaf-project.org/>
- [10] DAML-Time (DARPA Agent Markup Language-Time)  
<http://www.cs.rochester.edu/~ferguson/daml/>
- [11] 정한민, 강인수, 구희관, 이승우, 성원경, "URI 서버에 기반한 국가 R&D 기반정보 온톨로지 설계 및 구현," 정보관리연구, 37권, 2호, 2006.
- [12] 최호성, 임지희, 배영준, 최수일, 옥철영, "온톨로지 구축 방법과 사례," 정보과학회지, 24권, 4호, pp.31-44, 2006.
- [13] 한동일, 하상범, 최호준, "여우 서비스: 온톨로지 기반 검색 에이전트 구현 사례," 정보과학회지, 24권, 4호, pp.75-81, 2006.
- [14] John F. Sowa, "Knowledge Representation: Logical, Philosophical and Computational Foundations," Brooks/Cole, 2000.
- [15] Nicola Guarino, Christopher A. Welty, "An Overview of OntoClean in Handbook on Ontologies," Springer, pp.151-172, 2004.