

씨앗 용어 피드백 관련 검색에 근거한 온톨로지 구축

이인근 황도삼^o 권순학
영남대학교 전기공학과, 전자정보공학부^o
iklee@ynu.ac.kr {dshwang^o, shkwon}@yu.ac.kr

An Ontology Construction using Seed Term Feedback Relevance

In K. Lee, Dosam Hwang^o, Soon H. Kwon
Dept. of Electrical Engineering, Yeungnam University
Dept. of Electrical Engineering and Computer Science, Yeungnam University^o

요 약

컴퓨터를 활용한 지식과 정보의 고도 처리를 위해 온톨로지를 구축하고 활용하려는 요구가 강해지고 있다. 이런 요구에 부응하여 온톨로지의 구축 방법에 대한 많은 연구가 이루어지고 있다. 그러나 현재까지의 온톨로지 구축 이론과 온톨로지 개발 도구가 실제 구현 목적과 사례에 따라 제한적인 부분에서 개발되어 사용되고 있다. 그러므로 구축하고자 하는 분야의 온톨로지에 맞는 적절한 구축 프로세스와 도구가 필요하다. 분야 온톨로지(domain ontology)를 구축함에 있어 특정 분야의 전문가 또는 비전문가가 언어자원으로부터 지식을 개념화(conceptualization), 형식화(formalization)하여 온톨로지를 구축할 수 있도록 분야 관련 언어자원에 근거하여 온톨로지를 구축할 수 있는 온톨로지 구축 프로세스(OntoProcess1))를 제안한다. 그리고, 다수의 온톨로지 구축자가 동시에 같은 분야의 온톨로지를 구축할 경우, 개념화 과정에서 서로 다른 구축자가 동일한 개념을 중복 정의하거나, 형식화 과정에서 형식언어 이해 부족으로 인한 구축 능력 저하 문제가 발생할 수 있다. 이를 위해 메타 온톨로지(meta ontology)를 이용하여 다중 온톨로지를 구축할 때 발생하는 문제를 해결하는 다중 온톨로지 구축 프로세스를 제안한다. 현재 이 프로세스에 근거하여 온톨로지 구축 시스템(OntoCS2))을 개발하였고, 국가 IT 온톨로지 인프라 기술 개발 프로젝트에서 IT분야 온톨로지의 개발에 활용되고 있다.

1. 서론

근래에 들어 컴퓨터를 공유한 정보의 고도 처리를 위해 온톨로지를 구축하고 활용하려는 요구가 강해지고 있다. 이런 요구에 부응하여 온톨로지의 구축 방법에 대한 많은 연구가 이루어지고 있으며, 또한 온톨로지를 구축하기 위한 온톨로지 구축 도구들도 다양하게 개발되고 있다.

온톨로지의 일반적인 구축 방법은 목적 확인, 개념화, 형식화, 온톨로지 통합, 평가, 문서화와 같은 과정으로 이루어져 있다[8][9]. 온톨로지의 일반적인 구축 방법 중 개념화(conceptualization)는 어떤 주제 영역

에서 핵심적인 개념이 무엇이며 개념간의 관계가 어떻게 설정될 수 있는지를 확인하는 과정이다. 즉, 어떤 분야(domain)의 지식으로부터 개념을 정의하고 개념 간의 관계를 설정하는 과정으로서 분야 전문가의 개입이 필요한 단계이다. 그리고, 형식화(formalization)는 자연언어로 작성된 메타 정보로부터 온톨로지의 표현 언어로 기호를 생성하는 과정이다. 즉, 개념과 개념간의 관계를 컴퓨터가 이해할 수 있는 정규화된 언어로 표현하는 단계로써 컴퓨터 전문가의 개입이 필요한 단계이다. 그러므로 온톨로지를 구축하기 위해서는 각 분야 전문가가 필요하다. 그러나 분야 전문가가 충분하지 않은 집단에서의 온톨로지 구축 속도나 범위는 제한적일

* 본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도 기반기술개발사업의 연구결과로 수행되었습니다.

1) Ontology Construction Process의 약자임
2) Ontology Construction System의 약자임

수밖에 없다. 그러므로, 온톨로지를 구축할 때 부족한 분야 전문지식을 보완하기 위해 분야 전문가가 구성된 언어자원(사전, 코퍼스 등)으로부터 분야 전문지식을 추출하여 온톨로지를 구성한다면 비전문가도 충분히 분야 온톨로지를 구축할 수 있다.

그리고, 다수의 온톨로지 구축자가 동시에 같은 분야의 온톨로지를 구축할 경우, 개념화 과정에서 서로 다른 구축자에 의한 동일한 개념을 중복 정의하거나, 형식화 과정에서 형식언어 이해 부족으로 인한 구축 능력 저하 문제가 발생할 수 있다. 이런 문제는 온톨로지 구축 환경을 공유하여 중복 구축을 막고, 간단한 형태의 지식 표현방법을 사용함으로써 복잡한 형식언어에 대한 이해 부족 문제를 해결할 수 있다.

본 논문에서는 일반적인 온톨로지 구축 방법 중 개념화, 형식화 단계에서 비전문가가 온톨로지를 구축하기 위한 방법으로서, 분야에 관련된 언어자원에 기반한 온톨로지 구축 프로세스(OntoProcess)를 제안한다. 그리고, 메타 온톨로지를 이용하여 다중 온톨로지 구축 시 발생하는 문제를 해결하는 다중 온톨로지 구축 프로세스를 제안한다. 또한, 제안한 프로세스로 온톨로지를 구축할 수 있는 온톨로지 구축 시스템(OntoCS)을 개발하고, 개발한 시스템을 이용하여 실제적인 온톨로지 구축 실험을 통해 개발한 시스템의 효용성을 확인한다.

2. 관련 연구

온톨로지 구축 방법은 이론적, 경험적으로 다양하게 제안되어 왔다. 온톨로지 구축 방법으로는 SENSUS, KACTUS, OTK, CommonKADS, Tove, Methontology, Mikrokosmos, ONIONS, PHYSSYS 등과 같은 여러 종류가 있으며, 이론을 배경으로 한 실제 구현 사례를 통해 제한적인 부분에서 개발되어 사용되고 있다[9]. 그러나, 국내는 물론 국제적으로도 현재 온톨로지의 표준 개발 방법이 없다. 온톨로지를 구축하기 위해서 지금까지 제안된 온톨로지 구축 방법을 온톨로지 구축 목적에 맞게 보완하여 사용하고 있다.

2.1. OTK(On-To-Knowledge)

어플리케이션 지향의 온톨로지 개발에 중점을 두고 있다. On-To-Knowledge는 타당성 조사, 요구명세서 작성을 통한 시작, 지식의 공식화를 통한 정제, 반복적인 온톨로지 분석과 수정을 통한 평가, 유지 및 개선의 다섯 단계로 구성된다. 이 방법에서는 지식을 공식화 하여 온톨로지를 구축하기 위해 지식 도출 과정을 거쳐 온톨로지의 초안을 수정하고 확장한다. 전 과정에 걸쳐 전문가의 개입이 필요하다.

2.2. CommonKADS

대규모의 산업기반 지식 시스템을 만들기 위한 방법으로 유럽의 ESPRIT IT 프로그램 개발사업의 한 부분이다. 지식모델 구축, 지식 상세, 지식 정제의 세 단계로

구분하여 사전에 작업 및 조직 분석을 위해 타당성, 효과 조사를 수행한다. 지식의 확인 단계에서는 실제 지식모델 상세를 위한 준비로서 지식 모델링에 사용될만한 정보원을 확인한다. 정보원으로는 용어사전이나 어휘목록 등이 있으며, 정보원의 작업 템플릿, 영역 스키마 같은 기존 모델링 요소를 조사한다.

2.3. ONIONS

정보 입수에 있어 외래의 정보원을 통합하는 방법과 같은 지식 통합 문제를 해결하기 위해 시작되었다. 검증받은 언어자료를 수집, 분류 구조의 적절성 분석, 협의 정보원 분석, 다중 협의 정보원 분석, 통합된 온톨로지 라이브러리 생성, 온톨로지 라이브러리 분류과정을 통해 온톨로지를 구축한다.

일반적인 온톨로지 구축 방법은 목적 확인, 개념화, 형식화, 온톨로지 통합, 평가, 문서화와 같은 과정으로 이루어져 있다[8]. 목적 확인은 온톨로지를 구축하는 목적을 분명히 함으로써 특정 분야 시스템에서의 온톨로지 사용 여부를 결정하는 과정이다. 개념화는 특정 분야에서 핵심적인 개념이 무엇이며 개념간의 관계를 어떻게 설정할 수 있는지를 확인하는 과정이다. 특정 분야에서의 가장 기본이 되는 단어를 정의하고, 중복된 개념의 단어는 배제하고 유사어는 통합한다. 자연언어로 작성된 분야 지식으로부터 온톨로지 표현 언어로 지식을 형식화한다. 온톨로지 표현언어는 현재 OWL(Web Ontology Language)을 표준으로 권고하고 있다[3]. 온톨로지 구축 후 온톨로지의 기술적인 측면을 판별해 보고 관련 소프트웨어 환경을 살펴보는 과정을 통해 구축된 온톨로지를 평가한다. 온톨로지 구축 과정에서의 효율적인 지식 공유를 위해 문서화 과정이 필요하다.

3. 온톨로지 구축 프로세스

자연언어 처리 기술을 이용하여 사전의 용어를 정의한 문장으로부터 용어 사이의 관계를 자동으로 설정하여 시소러스와 같은 단어의 분류체계(taxonomy)를 구성하려는 연구가 있었다[10]. 그러나 이런 연구들은 최종 완성된 온톨로지를 자동으로 개발하려는 것이 아니라 사람의 수정 및 보완이 필요한 온톨로지 초안을 만드는 과정이다. 또한 이 연구에서는 `equivalentClass`(동언어), `complementOf`(반의어), `subclassOf`(상하위어), `disjointWith`(동위어), `unionOf`(부분어)와 같은 용어 관계만을 고려한다.

다양한 용어 간의 관계를 가진 온톨로지를 구축하기 위해서는 아직까지 인간의 주관적인 개입이 불가피하다. 분야 전문지식을 보유한 전문가가 분야 온톨로지를 구축한다면 구축된 최종 온톨로지의 질적 향상은 기대할 수 있으나 개발 속도가 매우 느리다. 또한 쉽고 편리한 인터페이스의 온톨로지 구축 도구를 활용하여 온톨로지를 구축한다 하더라도 소수의 전문가 집단이 구

축할 수 있는 온톨로지의 크기는 제한적이다. 그러므로 분야 언어자원을 활용하여 다수의 전문가나 비전문가가 대량의 분야 온톨로지를 구축하고, 구축한 온톨로지를 전문가가 검증하는 방법이 효율적이라 하겠다.

3.1. 온톨로지 구축, 통합, 검증, 평가 및 유지관리 프로세스

(그림 1)은 전체 온톨로지 구축, 통합, 검증, 평가 및 유지관리 프로세스를 보인다. 구축하려는 분야 온톨로지에 대한 명세화(specification)를 통해 온톨로지 구축 목적과 범위, 사용 방법 등을 결정한다. 분야 전문지식을 표현한 언어자원으로부터 지식을 개념화(conceptualization)하고, 온톨로지 표현 문법으로 지식을 형식화(formalization)하여 온톨로지를 구축(implementation)한다. 구축한 온톨로지가 분야 지식을 잘 표현하고 있는지 온톨로지 검사(verification)를 통해 구축한 온톨로지를 수정한다. 그리고, 온톨로지를 구축하고 수정하는 과정에서의 관련 기록을 문서화(documentation)한다. 온톨로지를 구축하는 과정에서 구축된 온톨로지가 온톨로지 구축 명세에 맞는지 전문가에 의해 지속적으로 검증(validation)된다. 분야 온톨로지의 구축이 완료되면, 사전 및 시소러스를 참조하여 구축된 분야 온톨로지와 타 분야 온톨로지와의 통합(integration) 과정을 거쳐 최종 온톨로지가 완성된다. 최종 온톨로지는 여러 가지 응용 방법을 통해 일반 사용자로부터 평가(assessment)가 이루어지며, 사용자의 평가를 통해 온톨로지가 유지(maintenance)되고 관리(management)된다.

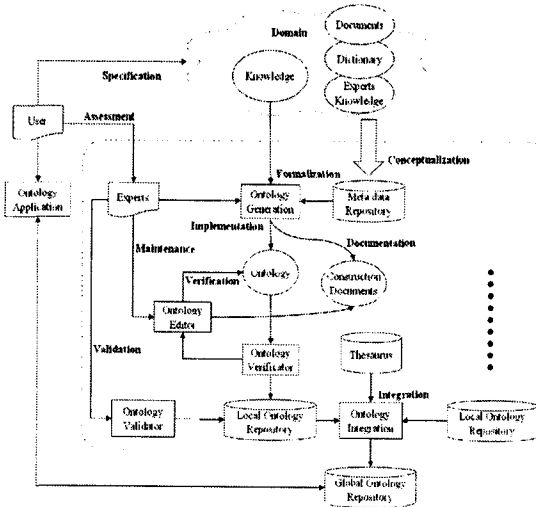


그림 1. 온톨로지 구축, 통합, 검증, 평가 및 유지관리 프로세스

3.2. 온토 프로세스(OntoProcess)

(그림 1)에 제시한 온톨로지 구축 프로세스 중 특정 분야의 언어자원에 기반하여 온톨로지 구축자가 전문지식을 개념화하고 형식화하는 상세 과정인 언어자원에 근거한 온톨로지 구축 프로세스(OntoProcess)를 제안한다.

[온토 프로세스]

- (1) 씨앗 용어(seed term)로 언어자원(코퍼스)으로부터 특정분야 관련 문서-문장 검색.
- (2) 문서-문장으로 부터 온톨로지 구축자가 수동으로 온톨로지를 구축하려는 분야의 분야용어 후보를 추출.
- (3) 추출된 분야용어 후보 중 하나의 용어를 선정.
- (4) 선정된 용어가 온톨로지에 등록되어 있는지 검색.
 - (4-1) 용어가 온톨로지에 등록되어 있으면 (3)과정 실행.
- (5) 용어사전과 인터넷을 통해 용어의 정의 검색.
- (6) 검색된 용어의 정의에 기반하여 온톨로지 구축자가 주관적으로 분야용어 선택.
 - (6-1) 용어의 정의를 알 수 있는 경우를 분야용어로 선정.
 - (6-2) 용어의 정의를 알 수 없으나 분야용어임이 확실한 경우 비정의 용어로 선정.
 - (6-3) 분야용어가 아닌 경우 후보에서 제거.
- (7) 용어의 정의로부터 대상 용어의 상위 용어를 온톨로지서 검색.
 - (7-1) 온톨로지서 상위 용어가 등록되어 있는 경우 용어를 상위 용어의 하위 용어로 등록.
 - (7-2) 온톨로지서 상위 용어가 등록되어 있지 않은 경우 (2)-(7)과정을 반복.
- (8) (2)에서 추출한 분야용어 후보가 모두 온톨로지에 등록될 때까지 (2)-(7)과정 반복.
- (9) (2)-(7)과정에서의 용어의 정의나 문서-문장으로 부터 용어 사이의 관계 추출.
 - (9-1) 용어 사이의 관계가 온톨로지에 등록되어 있지 않은 경우 온톨로지에 관계 등록.
 - (9-2) 용어 관계 설정.
- (10) (1)-(9)의 과정 반복.
 - (10-1) 잘못된 용어나 관계는 (1)-(9) 과정 중에 수정.

(그림 2)의 온토 프로세스는 분야 언어자원이 분류되어 있지 않은 상황에서 온톨로지를 구축하는 것을 전체로 한다. 신문기사, 전문문서, 사전의 용어정의 등 분야별로 분류되어 있지 않은 문서에서 관심 문서를 분류하는 많은 연구가 이루어지고 있으나, 새롭게 생성된 분야용어가 등록되어 있는 분야용어 사전이 없는 상황에서 관심 문서를 분류해 내는 것은 쉬운 일이 아니다. 또한 전문지식이 부족한 비전문가가 관련문서를 분류해 내는 것 또한 어려운 일이다. 따라서 (1), (2)과정과

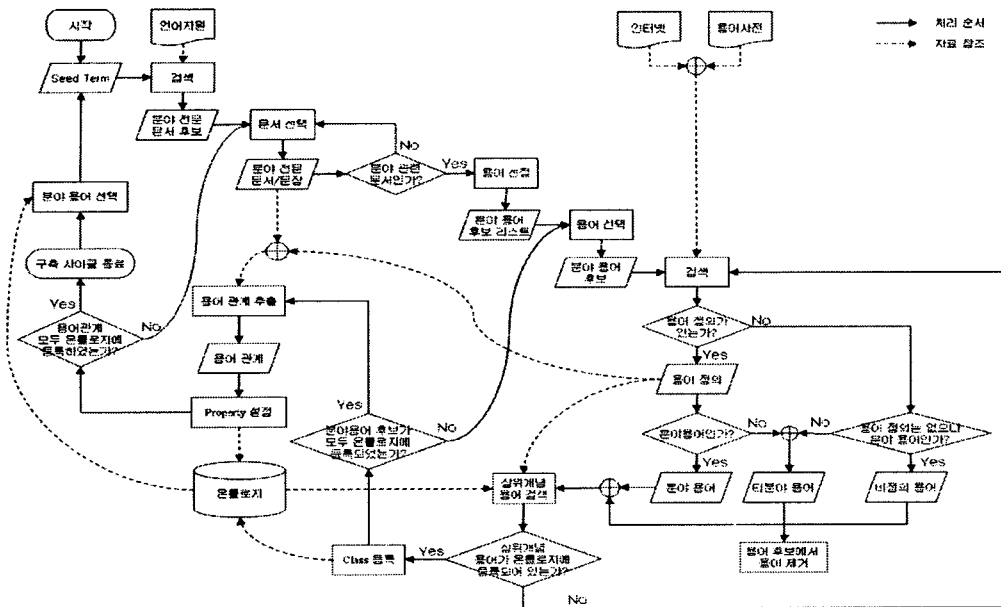


그림 2. 온토 프로세스 - 개념화, 형식화 과정

같이 특정 분야에서 가장 일반적으로 알려진 단어를 씨앗 용어(seed term)로 선정하여 언어자원으로부터 관련 후보 문서-문장을 검색하고, 검색된 문서들 중 온톨로지 구축자의 주관적 판단을 통해 적절한 문서를 선택한다. 선택된 문서로부터 온톨로지 구축자가 직접 분야용어를 추출하고, 추출한 분야용어를 이용하여 다른 관련문서에서 또 다른 분야용어를 추출하는 과정을 통해 분야용어를 구성한다.

(3)-(5)에서는 (1),(2)를 통해 온톨로지 구축자가 선택한 분야용어가 구축 중인 온톨로지에 등록된 용어인지를 확인하고, 온톨로지에 등록되어 있지 않은 용어는 온톨로지에 등록하기 위해 용어사전과 인터넷 등의 자원을 활용하여 용어의 정의를 찾는다. (6)과 같이 용어 정의를 참조하여 선택한 용어가 분야용어인지를 판단하여 온톨로지에 등록할지를 결정하고, 정의를 찾을 수 없지만 온톨로지 구축자가 주관적으로 분야용어임을 확신하는 경우에도 선택한 용어를 분야용어로 결정한다. (7)에서는 선택한 용어를 이용하여 실제로 온톨로지에 등록하는 과정으로서, 선택한 용어를 어떤 용어의 하위 개념에 위치할 것인지를 판단하는 과정이다. 이 과정에서 (3)-(5)에서의 용어 정의를 참조한다. 선택한 용어의 정의에서 상위개념의 용어가 온톨로지에 등록되어 있지 않은 경우 상위개념의 용어도 온톨로지에 등록해야만 선택한 용어도 온톨로지에 등록 가능하다. 따라서 온톨로지에 등록하고자 하는 상위개념의 용어 또한 (3)-(7) 과정을 통해 온톨로지에 등록하는 절차를 따른다. 용어 정의에서 상위개념의 용어뿐만 아니라 (2)의

과정을 통해 분야용어를 선택하여 온톨로지에 등록할 수 있다. (2)-(7) 과정에서 온톨로지 구축자가 선택한 분야용어 후보들이 모두 온톨로지에 등록되거나 후보에서 제거될 때까지 (2)-(7)과정을 반복한다.

지금까지는 언어자원으로부터 분야용어를 수집하고, 온톨로지에 등록하는 과정이다. 하지만 이 과정을 통해 구축한 온톨로지는 용어 의미의 상하 관계만을 표현하므로 진정한 온톨로지라 할 수 없다. 따라서 용어 사이의 상하 관계 이외의 관계를 설정하기 위해서 (2)-(7) 과정에서 참조한 문서-문장 및 용어의 정의를 이용한다. 문서-문장 및 용어의 정의로부터 온톨로지에 등록된 용어 사이의 관계를 추출하여 온톨로지에 용어 사이의 관계를 설정한다. 이 과정에서 온톨로지에 이미 정의된 관계의 경우에는 용어 관계를 설정만 해 주면 된다. 그러나, 온톨로지에 정의되어 있지 않은 관계의 경우에는 온톨로지에 관계를 추가하여 정의해 주어야 한다. (1)-(9) 과정을 반복하면서 잘못된 부분은 수정해 나간다. 이 과정을 반복할수록 온톨로지의 규모가 커지게 된다.

3.3. 메타 온톨로지

다수의 온톨로지 구축자가 제안한 온톨로지 구축 프로세스로 온톨로지를 구축할 때 개념화 과정에서 서로 다른 구축자에 의한 동일한 개념을 중복 정의하거나, 형식화 과정에서 형식언어 이해 부족 문제로 인한 구축 능률 저하 문제가 발생할 수 있다. 온톨로지를 동시에

구축할 때 동일한 개념의 중복 정의를 막기 위해 하나의 온톨로지 구축 환경을 네트워크를 통해 공유함으로써 문제를 해결할 수도 있다. 그리고, Protege[1] 같은 온톨로지 구축 도구를 이용하여 형식화 과정에서의 문제를 해결할 수 있지만, 온톨로지 구축 도구 사용에 대한 사전 훈련이 필요하다. 이런 문제를 해결하기 위해 전문지식을 <표 1>과 같은 간단한 형식의 메타 온톨로지(meta ontology)로 작성하여 네트워크로 공유하고, 온톨로지 구축 도구 사용에 숙련된 온톨로지 구축자가 작성된 메타 온톨로지를 온톨로지 구축 도구를 이용하여 형식화한다.

<표 1> 메타 온톨로지 형식과 예

Class Name	digital television
Korean	디지털 TV
Definition	흑백시대 · 컬러시대를 거친, 이른바 제3세대 텔레비전이다.
Synonym	디지털 텔레비전, digital TV

(a) 클래스 정의

Property	isSubclassOf (이다)
Restriction	someValuesFrom
Domain	digital television (디지털 TV)
Range	television (텔레비전)
Sentence	디지털 TV는 흑백시대 · 컬러시대를 거친, 이른바 제3세대 텔레비전이다

(b) 클래스 속성

Instance	Pavv SVP-42Q2HL1
SuperClass	digital television

(c) 인스턴스 정의

Property	isProducerOf
Domain	삼성 (Samsung)
Range	Pavv SVP-42Q2HL1
Sentence	http://blog.naver.com/ecoms?Redirect=Log&logNo=80027474383

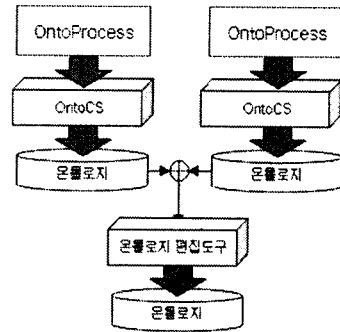
(d) 인스턴스 속성 (object)

Property	panelSize
Domain	Pavv SVP-42Q2HL1
Data value	42inch
Data type	integer
Sentence	http://blog.naver.com/ecoms?Redirect=Log&logNo=80027474383

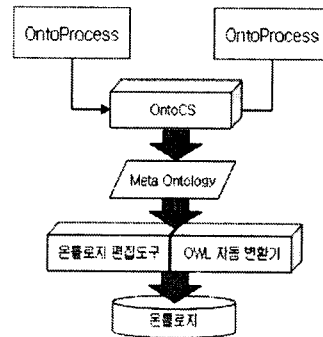
(e) 인스턴스 속성 (data type)

<표 1(a)>는 개념을 나타내는 클래스(class)의 형식이다. <표 1(b)>는 개념의 속성을 나타내는 domain-property-range의 triple 형식을 나타낸다. <표 1(c)>는 개념의 구체물인 인스턴스(instance)이다. 클래스와 클래스 사이의 속성(property)을 설정할 수 있는 것과

달리, 인스턴스는 다른 인스턴스나 자료형(data type)에 대해 속성을 설정할 수 있다. <표 1(d)>과 <표 1(e)>은 인스턴스에 대한 속성 형식을 나타낸다.



(a) 온톨로지 구축 프로세스



(b) 다중 온톨로지 구축 프로세스

그림 3. 제안한 온톨로지 구축 프로세스 비교

제안한 온톨로지 구축 프로세스에서 메타 온톨로지를 이용한 온톨로지 다중 구축 프로세스는 (그림 3(b))와 같다. 제안한 온톨로지 구축 프로세스를 통해 온톨로지를 직접 구성할 경우 온톨로지 다중 구축자에 의해 구축된 온톨로지를 통합해야 한다. Prompt plugin[2]과 같이 온톨로지 통합(merging) 도구를 이용할 수도 있으나, 도구를 이용하더라도 온톨로지 통합 과정은 단순하지 않다. 그러나 온톨로지 구축 프로세스에서의 출력으로 직접 온톨로지를 구축하지 않고 메타 온톨로지를 구축하고 네트워크로 공유하면, 다중 구축자들이 중복 작업을 막을 수 있고, 구축된 메타 온톨로지를 기반으로 자동, 반자동으로 온톨로지를 구축할 수 있다.

3.3. 온톨로지 구축 시스템(OntoCS)

제안한 온톨로지 구축 프로세스를 온톨로지를 구축하기 위해서는 전문가가 작성한 문서와 사전, 그리고 온톨로지 구

축 도구 등을 동시에 사용해야 한다. 많은 언어자원을 동시에 활용한다면 온톨로지 구축자가 언어자원의 사용에 혼란을 겪게 되고, 온톨로지 구축 과정에서 언어자원을 참조하기 위한 참조 이동은 온톨로지 구축에 부가적인 소요 시간과 혼란을 발생시킨다.

실제로 언어자원을 이용하여 온톨로지를 구축하기 위해서는 코퍼스, 사전, 인터넷, 온톨로지 구축 도구, 그리고 별도의 기록 장소 등 많은 도구와 자원을 이용해야 한다. 그러나 많은 언어자원과 도구를 동시에 사용하거나 언어자원에 대한 사용 정보를 별도로 기록하고 수정하는 과정이 매우 번거롭다. 이런 구축 과정에서의 번거로움을 줄이기 위해 온톨로지 구축 시 다양한 언어자원을 효율적으로 활용하고 관리하여 온톨로지를 쉽게 구축하고, 구축된 온톨로지를 웹 환경에서 간단하게 확인하고 비교할 수 있는 온톨로지 구축 시스템(OntoCS)을 개발하였다. (그림 4)에 OntoCS의 구조를 보인다.

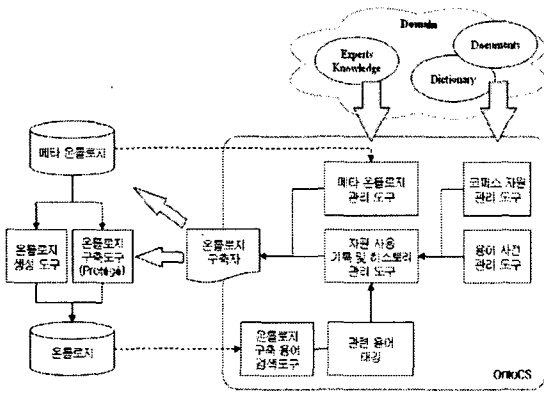


그림 4. OntoCS 구조

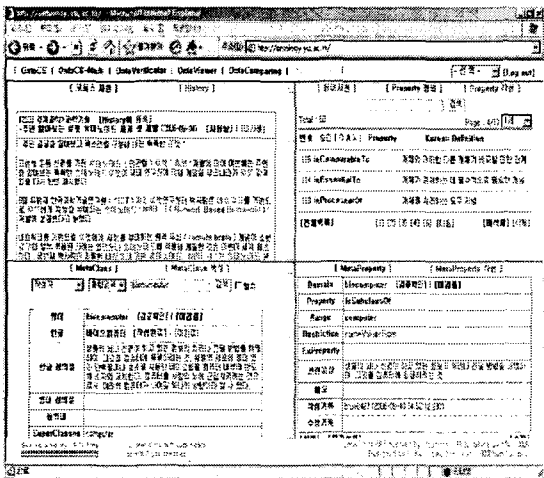


그림 5. OntoCS 실행 화면

OntoCS는 "코퍼스 자원 관리 도구", "용어 사전 관리 도구"를 통해 온톨로지 구축에 필요한 언어자원을 관리한다. 그리고 "자원 사용 기록(log) 및 히스토리(history) 관리 도구"를 이용하여 사용한 자원에 대한 목록을 관리하도록 하였다. 개념화한 지식을 "온톨로지 구축 도구"를 이용하여 직접 온톨로지에 구축할 수 있고, "메타 온톨로지 관리 도구"를 통해 다수의 온톨로지 구축자가 온톨로지 구축 전단계의 "메타 온톨로지"를 작성할 수도 있다. 그리고, "온톨로지 구축 용어 검색 도구"를 통해 구축한 온톨로지의 용어를 검색할 수 있으며, "관련 용어 태깅"을 통해 문장으로부터 관련 용어를 쉽게 추출하여 온톨로지 구축에 도움이 될 수 있도록 한다. (그림 5)는 OntoCS의 실행 화면이다.

OntoCS에서 구축한 메타 온톨로지는 OntoCS에서 제공하는 OWL 변환기를 통해 OWL 파일로 자동 변환되고, 변환된 OWL 파일은 (그림 7)과 같이 Protege와 같은 온톨로지 편집 도구를 통해 확인 및 관리할 수 있다.

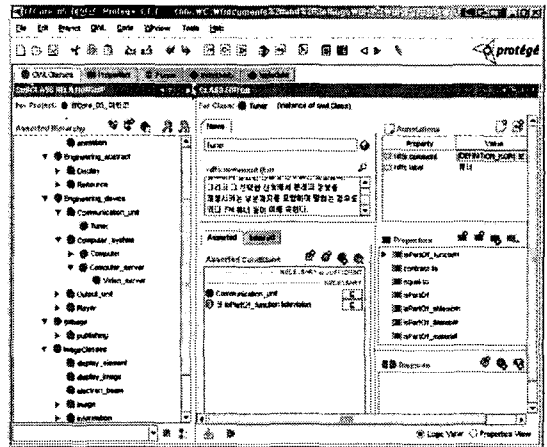


그림 7. "디지털 TV 및 방송" 분야 온톨로지

4. 온톨로지 구축 실험

4.1. 온토 프로세스에 기반한 온톨로지 구축 실험

온토 프로세스에 기반하여 개발한 온톨로지 구축 시스템(OntoCS)을 사용하여 "디지털 TV 및 방송" 분야의 온톨로지를 구축하였다. 온톨로지 구축 그룹을 구성하기 위하여 "디지털 TV 및 방송" 분야와 무관한 전공의 대학생 17명으로 온톨로지 구축 그룹을 구성하였다. 온톨로지 구축자에게는 온톨로지의 개념과 온톨로지 구축 프로세스, OntoCS의 사용법에 대한 기본 교육을 실시하고, 각 개인별로 한달동안 10회~20회의 온톨로지 구축 연습을 하도록 하였다.

온톨로지 구축 시 2005년 1월부터 2006년 5월까지의 전자신문 기사를 수집한 코퍼스와 TTA정보통신용어사전을 언어자원으로 이용하였고, OntoCS에서 언어자원을

관리하고 이용할 수 있도록 하였다.

실험에 앞서 온톨로지 구축 그룹에서 5명을 피실험자로 선별하고, 10회의 온톨로지 구축 실험에서 피실험자의 온톨로지 구축량과 온톨로지 구축시 소요시간을 측정하였다. 그리고, 클래스와 프로퍼티의 등록시 소요되는 평균 시간을 계산하였다. <표 2>는 실험 종료 후 온톨로지 총 구축량과 전체 실험시간을 보인다. (그림 6)은 온톨로지 구축자별로 각 회차별 실험에서 클래스와 프로퍼티의 평균 등록시간을 그래프로 보인 것이다.

<표 2> 온톨로지 구축 소요시간 측정 결과

피실험자	A	B	C	D	E
구축개수(개)	525	515	507	586	1458
작업시간(분)	1139	763	738	1032	1059
평균등록시간(분)	2.17	1.48	1.45	1.76	0.72

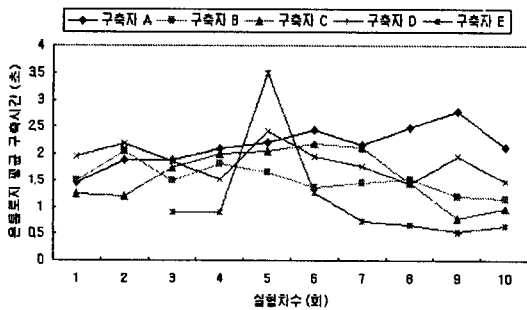


그림 6. 온톨로지 구축 소요시간 측정 그래프

본 실험 결과는 온톨로지에 대한 이해와 도구 사용법을 숙지한 이후 실시한 실험이므로 실험 회차를 거듭하면서 작업 속도가 빨라졌다고 판단하기는 어렵다. 전반적으로 온톨로지 평균 등록시간이 단축되기는 하였으나, 피실험자 A의 경우에는 오히려 증가한 경우를 볼 수 있었다. 따라서 5명의 피실험자의 실험결과 중 상반되는 두개의 실험결과를 <표 3>과 <표 4>에 보이며, 그 차이점을 분석한다.

두 피실험자의 실험 결과를 볼 때, 피실험자 A의 경우는 실험 회차를 거듭할수록 클래스의 등록 비중이 높아진 반면, 피실험자 E의 경우는 프로퍼티의 설정 비중이 높아진 것을 알 수 있다. 프로퍼티 설정 비중이 높아짐에 따라 평균등록시간이 단축되는 원인은 피실험자의 인터뷰를 통해 알 수 있었다. <표 1>에서와 같이 클래스를 등록하기 위해서는 많은 정보를 입력해야 하지만, 프로퍼티 설정은 이미 등록된 클래스를 참조만 하므로 프로퍼티의 설정 시간이 클래스 등록 시간보다 더 짧다. 그리고, 온톨로지로 표현할 수 있는 개념의 수가 제한되어 있으므로 실험이 진행됨에 따라 클래스를 등록하는 수가 줄어드는 반면, 이미 등록된 클래스를 참조하는 프로퍼티의 설정 개수는 늘어난다.

<표 3> 온톨로지 구축 실험결과 (피실험자 A)

차수(회)	클래스(개)	프로퍼티(개)	등록비율 클래스:프로퍼티	작업시간(분)	평균등록 시간(분)
1	13	38	1 : 2.9	74	1.45
2	9	32	1 : 3.5	77	1.87
3	17	35	1 : 2.0	98	1.88
4	13	38	1 : 2.9	107	2.09
5	21	35	1 : 1.6	124	2.21
6	24	37	1 : 1.5	149	2.44
7	22	39	1 : 1.7	132	2.16
8	15	29	1 : 1.9	109	2.47
9	22	38	1 : 1.7	167	2.78
10	8	40	1 : 5.0	102	2.12

<표 4> 온톨로지 구축 실험결과 (피실험자 E)

차수(회)	클래스(개)	프로퍼티(개)	등록비율 클래스:프로퍼티	작업시간(분)	평균등록 시간(분)
1	-	-	-	74	-
2	-	-	-	77	-
3	6	14	1 : 2.3	98	0.90
4	12	15	1 : 1.2	107	0.89
5	3	5	1 : 1.6	124	3.50
6	34	69	1 : 2.0	149	1.26
7	143	376	1 : 2.6	132	0.73
8	76	192	1 : 2.5	109	0.67
9	83	232	1 : 2.7	167	0.53
10	61	137	1 : 2.2	102	0.63

4.2. 온톨로지 구축결과 비교

제안한 온톨로지 구축 프로세스의 타당성을 검증하기 위해 실험에서 구축한 온톨로지와 외부의 온톨로지 구축 그룹에서 구축한 온톨로지를 대조하여 비교하고자 한다.

실험 대조를 위한 온톨로지는 4명의 온톨로지 구축자가 Protege를 사용하여 각기 다른 방법으로 구축하였고, 구축작업 중간 결과물로서 프로퍼티는 설정하지 않고 클래스와 클래스의 계층구조만으로 구성되었다. 본 논문에서 제안한 온톨로지 구축 프로세스와 달리 온톨로지 구축자가 각기 다른 방법으로 온톨로지를 구축하였으며, 구축 중간 결과 4개의 OWL 파일이 생성되었다. <표 5>와 <표 6>은 실험 대조 온톨로지를 분석한 결과이다.

<표 5>에서도 알 수 있듯이 서로 다른 분야의 온톨로지를 구축하였음에도 다수의 중복되는 클래스가 존재하였다. 또한 <표 6>에서와 같이 두명의 온톨로지 구축자가 동일한 클래스를 등록한 회수가 31번이고, 세명의 온톨로지 구축자가 동일한 클래스를 등록한 회수가 2번

이다. 다중 상속이 가능한 OWL의 특성에서 흔히 일어날 수 있는 일이며, 온톨로지 구축자 개인별로 온톨로지를 구축시에는 프로퍼티를 등록하는 과정에서 더 심한 클래스의 중복 등록이 예상된다.

<표 5> 실험 대조 온톨로지 분석 결과

구축자	a	b	c	d
구축개수	493	480	60	696
중복개수	20	23	4	21

<표 6> 실험 대조 온톨로지에서의 클래스 중복 회수

전체 중복회수	0번	1번	2번
계	1661	31	2

메타 온톨로지를 이용한 다중 온톨로지 구축 프로세스에 따라 온톨로지를 구축시 클래스와 프로퍼티의 중복등록을 방지하고, 다른 구축자에 의해 등록된 클래스를 이용함으로써 온톨로지 구축 능력이 향상됨을 알 수 있다.

5. 결론

온톨로지의 활용에 대한 필요성은 높아지고 있으나 실제적으로 온톨로지를 구축하여 사용하기에는 많은 어려움이 따른다. 특정 분야에 대한 온톨로지를 구축하기 위해서는 분야 전문가와 컴퓨터 관련 전문가의 개입이 필수적이라 하겠다. 하지만 소수의 전문가 집단이 대량의 온톨로지를 구축하기는 쉬운 일이 아니다. 따라서 본 논문에서는 일반적인 온톨로지 구축 방법 중 개념화, 형식화 과정에서 특정 분야에 대한 전문가나 비전문가가 온톨로지를 구축하기 위한 온톨로지 구축 프로세스(OntoProcess)를 제안하였다. 그리고, 다수의 온톨로지 구축자가 동시에 온톨로지를 구축할 경우 발생할 수 있는 문제점을 해결하기 위해 메타 온톨로지를 사용한 다중 온톨로지 구축 프로세스를 제안하였다. 그리고, 제안한 프로세스를 기반으로 온톨로지를 쉽게 구축할 수 있는 온톨로지 구축 시스템(OntoCS)을 개발하였고, 국가 IT 온톨로지 인프라 기술 개발 프로젝트에서 IT분야 온톨로지의 개발에 활용되고 있다. 그리고, IT분야 비전문가로 구성된 온톨로지 구축팀이 구성되어 제안한 프로세스에 따라 IT분야 언어자원에 기반한 온톨로지 구축이 진행되고 있다.

전체 온톨로지 구축 프로세스에서 보듯이 온톨로지 구축은 단순하지 않은 과정이다. 지식의 개념화, 기호화를 통해 온톨로지를 구축하였다 하더라도, 구축한 온톨로지에 대한 검증과 평가를 통한 온톨로지의 유지관리가 매우 중요하다. 따라서, 구축한 온톨로지에 대한 검증과 수정을 통해 양질의 온톨로지를 유지하기 위한 연구가 필요하다. 또한, 구축한 온톨로지의 활용을 통해 실제 사용 가능한 온톨로지의 평가에 대한 연구도 필요하다.

참고 문헌

- [1]Protege, <http://protege.stanford.edu>, 2006.7.
- [2]Prompt plugin, <http://protege.stanford.edu/plugins/prompt/prompt.html>, 2006.7.
- [3]OWL Web Ontology Language Guide, <http://www.w3.org/TR/owl-guide>, 2006.7.
- [4]Michael Denny, "Ontology editor survey results", http://xml.com/2002/11/06/Ontology_Editor_Survey.html, 2002.6.
- [5]Natalya Fridman Noy, and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, 2001.3.
- [6]Hai Wang, Matthew Horridge, Alan L. Rector, Nick Drummond, Julian Seidenberg, "Debugging OWL-DL Ontologies: A Heuristic Approach", International Semantic Web Conference 2005, 2005, pp.745-757.
- [7]Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang and Chris Wroe, "OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns", 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW), Whittlebury Hall, Northamptonshire, UK, 2004.
- [8]"웹 온톨로지 개발지침 연구", 한국전산원, NCA IV-RER-04059, 2004.
- [9]김은경, 남영준, "시맨틱 웹을 위한 온톨로지 구축 방법에 관한 비교 연구", 정보관리연구, vol.35, no.2, 2004, pp.57-85.
- [10]송도규, "대용량 OWL 온톨로지 자동구축을 위한 세종전자사전 활용 방법론 연구", 한국언어정보학회지:언어와정보, 1226-7430, 제9권1호, 2005, pp.19-34.
- [11]황도삼, 최기선, "차세대 웹을 위한 온톨로지 공학", 두양사, 2006(발간예정).