

대용량 온톨로지 구축에 있어서 협력적 개발 방법+

문홍구*, 박동훈*, 조이현*, 권혁철*

*부산대학교 컴퓨터 공학과

{mhg09, douguzzang, navigator, hckwon}@pusan.ac.kr

Cooperative Development Method for Construction of Large-scale Ontology

Hong-Goo Moon*, Dong-Hun Park*, Yi-Hyon Cho*, Hyuk-Chul Kwon*

*Dept of Computer Science and Engineering,

Pusan National University

요 약

최근 시맨틱웹의 중요성이 두드러지면서 다양한 분야에서 연구가 활발하게 진행되고 있다. 이렇듯 중요하게 두드러지고 있는 시맨틱웹의 연구가 진행되기 위해선 다양한 분야의 온톨로지가 필요하다. 현재 많은 온톨로지 구축 관련 도구들이 있어 온톨로지 구축에 편리함을 제공하고 있지만, 여전히 온톨로지의 구축은 많은 노력과 시간이 필요하다. 특히, 대용량의 온톨로지 구축은 더 많은 노력과 시간이 필요하다. 따라서 본 논문은 기본적으로 제시된 온톨로지 구축방법 모델들의 비교 분석을 통해 대용량 온톨로지 구축에 협력적 개발 과정을 제시하고자 한다.

1. 서론

기존 웹의 문제점을 보완할 수 있는 차세대 웹으로 시맨틱웹이 등장했다. 이러한 시맨틱웹을 구현하는 데 있어 중요한 요소 중 하나가 온톨로지이다.

시맨틱웹 환경에서는 정보들 간의 의미와 관계가 잘 표현되어 있어야 한다. 정보를 표현하는 데 있어서 각각의 정보가 갖고 있는 의미를 정의하고, 그 의미에 기반을 두어 또 다른 정보와의 연결고리를 유지한다면 사용자는 수많은 정보 속에서 필요한 정보를 보다 효율적으로 검색하고 처리할 수 있게 된다. 또한, 온톨로지는 공유된 개념들의 관계를 형식적으로 표현함으로써 사람뿐만 아니라 컴퓨터도 정보의 의미를 해석하여 효과적인 정보처리가 가능하게 한다. 따라서 차세대 시맨틱웹 환경에서는 의미 기반 서비스를 위한 다양한 온톨로지가 요구된다. 하지만, 온톨로지를 구축하는 데에는 많은 시간과 노력이 필요하다. 특히 대용량 온톨로지의 구축은 더욱더 그러하다.

본 연구는 현재까지 여러 관련연구를 통해 제시된 온톨로지 구축방법 모델을 비교, 참조하여 대용량 온톨로지를 협력적으로 구축함에 있어 효율적인 온톨로지 구축방법을 제시하는데 연구의 목적이 있다.

2. 관련연구

2.1 온톨로지 구축방법 모델

온톨로지 구축방법은 이론적, 경험적으로 다양하게 제안되어 왔다. 일반적으로 온톨로지 구축방법은 목적확인, 개념화, 기호화, 기존 온톨로지 통합, 평가, 문서화와 같은 기본과정으로 이루어져 있다. 현재까지 여러 관련 연구를 통해 많은 구축방법 모델들이 제시되었다. 제시된 온톨로지 구축방법 모델 중 온톨로지 구축과정에서 가장 기본적인 과정을 따르고 있는 모델들로서 OTK 구축방법 모델과 METHONTOLOGY 구축방법 모델이 있다. 그리고 온톨로지 구축에서 통합적 방법을 제시하고 있는 구축방법 모델로 ONIONS 모델이 제시되어 있다. 본장에서 이들 구축방법 모델들의 특징을 살펴본다.

2.1.1 OTK

+ 본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도기반 기술개발사업의 연구결과로 수행되었다.

OTKM(On-To-Knowledge Methodology for development and evolution of Ontology based Tool for knowledge Management)[1] 방법론은 EU IST-1999-10132 프로젝트인 On-To-Knowledge에 의해 개발된 방법론으로 온톨로지 기반의 지식 관리 프레임워크에 적용된다. 타당성 연구(Feasibility study), 개시(Kickoff), 정제(Refinement), 평가(Evaluation), 유지보수(Maintenance) 5단계로 구성되어 있다. 타당성 연구는 온톨로지의 적용 범위를 정의하며, 개시를 통해 요구명세서와 초기 온톨로지를 구축한다. 정제는 초기 온톨로지를 검증하고 정제하여 공식온톨로지를 생성하고, 평가는 생성된 공식온톨로지를 분석하는 단계이다. 마지막으로 구축된 온톨로지를 유지보수한다. 온톨로지 구축과정에서 정제, 평가, 유지보수 단계는 필요하면 반복적으로 수행하게 된다.

2.1.2 ONIONS

ONIONS(Ontological Integration Of Navie Source)[2] 방법론은 1990년대 초 다른 개념상의 온톨로지들의 통합 문제를 설명하기 위해 개발되었다. 주로 의학 분야의 도메인 온톨로지의 개념 통합을 지원하는 통합 온톨로지 개발, 도서관 정보검색과 디지털 콘텐츠의 통합 등을 지원하기 위해 사용되었다. ONIONS 방법론은 도메인에서 유효한 전문용어 말뭉치(corpus)를 수집하고, 용어가 가진 개념을 구체화한다. 그리고 각 항목을 명세화한다. 다음은 명세화를 일반적인 온톨로지 라이브러리로 구성하여 도메인 온톨로지를 구현한다. 이 방법론은 전문 용어 수집을 통해 온톨로지 구축이 가능하다.

2.1.3 METHONTOLOGY

METHONTOLOGY[3] 방법론은 AIU(Artificial Intelligence at Universidad Politecnica)의 연구실에서 개발하였고 소프트웨어 공학과 지식 공학 분야의 방법론에서 많은 영향을 받아, 온톨로지 개발에 관한 포괄적인 방법론을 제공하고 있다. 이 방법론은 온톨로지 개발 환경을 위한 생명주기 모형을 지향하며, 온톨로지 개발에 프로토타이핑 기법을 사용한다. 명세화, 개념화, 형식화, 구현 유지보수 5단계로 구성되어 있으며, 명세화 단계에서는 사용목적 및 구축범위를 정한다. 개념화 단계에서 용어사전을 구축하고 개념 분류 트리를 작성하여 개념화하고, 형식화 및 구현단계를 통해 ODE나 WebODE를 통해 온톨로지 언어를 자동생성하게 된다. 마지막으로 온톨로지를 정정, 개정, 보강하는 유지보수단계를 가진다.

3. 대용량 온톨로지의 협력적 구축

3.1 온톨로지 구축방법 모델 비교 분석

온톨로지 구축방법 모델에서 OTK 구축방법 모델과 METHONTOLOGY 구축방법 모델은 기본적으로 소프트웨어 공학론의 방법론을 따르고 있다. 온톨로지 역시 소프트웨어처럼 요구사항을 분석하고 개발, 정제, 유지 보수 단계를 거치게 된다. 따라서 일반적인 온톨로지 구축에서 OTK 구축방법 모델과 METHONTOLOGY 구축방법 모델이 효과적이다.

반면 ONIONS 구축방법 모델은 동일한 주제 분야의 용어나 온톨로지 통합을 목적으로 만들어진 구축방법 모델로, 용어를 활용하기 위한 통합방법을 제시하고 있다.

아래의 표는 온톨로지 구축에서 전체적인 온톨로지 관리 절차와 개발 과정에서의 차이점, 지식획득 과정의 유무, 개념사전 구축 과정의 유무 등을 비교 정리한 표이다.

<표 1> 온톨로지 개발 방법론 비교

특 정	OTK	ONIONS	METH(*)
온톨로지 관리	○	×	○
개발 과정	명세화	○	○
	개념화	○	○
	형식화	×	○
	구현	○	○
	유지보수	○	○
지식획득	×	○	○
개념사전	×	○	○
온톨로지통합	×	○	×

○:지원, ×: 미지원

METH(*): METHONTOLOGY

3.2 대용량 온톨로지 구축을 위한 협력

온톨로지 개발은 많은 시간과 노력이 요구된다. 특히 대용량의 온톨로지를 개발하는 것은 두말할 나위가 없다. 따라서 대용량 온톨로지 개발에서 협력적 개발은 온톨로지 개발의 시간과 노력을 줄일 수 있는 하나의 방법이다.

대용량 온톨로지는 특정 주제 분야에서의 광범위한 개념관계를 포괄하고 있다. 광범위한 개념관계를 표현하고 있는 온톨로지는 특정 주제 분야 내에 있는 협의의 개념관계를 표현한 온톨로지 집합이라 볼 수 있다. 따라서 대용량 온톨로지의 상위 개념 주제를 바탕으로 각각의 상위 주제별 하위 온톨로지를 협력적으로 개발하여 통합한다면 대용량 온톨로지를 구축하는 데 있어 많은 시간과 노력을 줄일 수 있다.

대용량 온톨로지 개발단계에서 가장 중요한 단계는 수집된 용어들에서 상위 개념 클래스를 추출하여 상위 개념 구조를 구성하는 데 있다. 이런 상위의 기본적인 클래스 구조를 바탕으로 각각의 하위 온톨로

지를 협력적으로 구축하고, 구축된 하위 개념관계를 통합한다.

3.2.1 대용량 온톨로지 구축 단계

본 논문에서는 <표 2> 와 같이 대용량 온톨로지 구축단계를 6단계로 구분하여 정의하였다.

<표 2> 협력적 온톨로지 구축단계

단계	작업
1단계	- 사용 목적, 구축 범위 설정
2단계	- 용어 수집, 정의
3단계	- 상위 주제의 개념 트리 작성
4단계	- 주제별 작업자 분배 - 하위 개념 간 기본 속성 정의 - 주제별 하위 온톨로지 작성
5단계	- 하위 온톨로지 병합 - 하위 온톨로지 간 개념 관계 정의
6단계	- 온톨로지 정제

대용량 온톨로지 구축을 위한 첫 번째 단계는 구축할 온톨로지의 사용 목적과 범위 설정 단계로 앞으로 온톨로지 구축과정의 일관성 유지를 위한 준비 단계이다.

2단계는 특정 주제 분야의 용어를 수집하고, 수집된 용어를 바탕으로 개념을 정의하게 된다. 수집된 용어개념 정의 단계에서 하위 온톨로지 구축을 위한 충분한 주제별 사전 분류 작업이 이루어져야 한다. 이는 하위 온톨로지 간의 용어 중복 등의 문제를 줄여준다.

3단계는 수집된 용어들을 바탕으로 상위 주제의 개념 추출을 위해 개념 트리를 작성하는 단계이다. 개념화 단계를 거친 후 추출된 상위 개념들 간에는 상하 구조가 아닌 동등한 수준의 개념구조를 형성하게 된다. 이렇게 형성된 상위 개념 구조를 바탕으로 하위 용어들의 개념관계를 구축하게 된다. 상위 클래스가 명확하게 구분되어 있지 않은 상황에서 하위 온톨로지를 구축하면 각각의 개념관계의 중복현상이 불가피하게 나타난다. 따라서 상위 클래스 간의 개념관계는 최대한 정제되어야 한다.

4단계에서는 추출된 동등한 레벨의 클래스는 각각의 하위 주제범위를 형성하게 되고, 형성된 각각의 주제범위는 구축자들에게 할당되어 하위 온톨로지가 구축된다. 각각의 구축자들이 하위 온톨로지를 구축하는 데 있어서 개념들 간의 속성 정의에 대한 제약이 뒤따라야 한다. 개념들 간의 속성을 정의하는 데 있어서 구축자들의 주관적인 생각이 반영되게 되면

하위 온톨로지 병합과정에서 개념들 간의 속성 정의에서 일관성이 떨어지게 된다. 따라서 각각의 하위 온톨로지 구축은 사전에 제한적으로 정해진 속성 사용 규칙에 의해 구축되어야 한다.

5단계에서는 각각의 주제범위별로 형성된 하위 온톨로지는 최상위 클래스를 중심으로 병합 과정을 거치게 된다. 병합과정을 수행하는 데 있어 개발자들 간의 중복 사용된 용어의 삭제, 빠진 용어 추가 과정, 하위 온톨로지 간 개념 관계 정의 작업등이 행해진다.

마지막 6단계에서는 구축된 온톨로지를 검증하고 온톨로지 정제를 되풀이하면서 대용량 온톨로지가 구축된다.

3.2.2 대용량 네트워크 온톨로지 구축

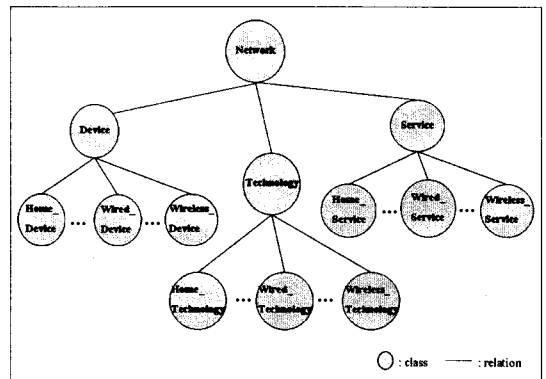
본 논문에서 제시한 대용량 온톨로지의 협력적 구축방법을 토대로 네트워크 분야의 개념을 포괄하는 대용량 온톨로지를 구축하였다.

웹을 비롯한 다양한 문서로부터 네트워크 분야와 관련된 용어를 수집하고, 개념을 정립하는 단계를 수행하였다. 아래의 표는 수집된 용어의 일부분이다.

<표 3> 용어 수집 개념정의 예

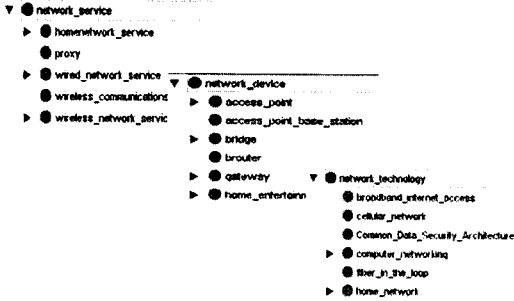
KorTerm	EngTerm	정의	분류
무선1394	Wireless 1394	유선 영역을 IEEE1394로 접속하고 무선 영역을 UWB로 접속하는 방식	네트워크기술
초광대역 무선	LWB	메우빌은 주파수 대역을 이용하여 대용량의 데이터를 초고속으로 전송할수 있는 전송기술	네트워크기술
지그비	ZigBee	데이터 보안, 기기들의 상호 복잡을 갖는 mesh, peer-to-peer, cluster tree network 표준 정의	네트워크기술
홈 PNA	HomePNA	가정에서 전화선을 이용하여 2대 이상의 컴퓨터들을 서로 공유할 수 있도록 하는 네트워크 기술	네트워크기술
Home Control 기술	Home Control	홈 네트워크 기기를 제어하고 모니터링 하는 기술	네트워크기술
빌딩 자동화	Building Automation	빌딩 자동화에 관련 기기들을 자동화하기 위한 기술	네트워크기술

다음 단계로 수집된 용어를 바탕으로 상위 개념 구조를 구축하였다. (그림 1) 과 같이 Network을 중심으로 Device, Technology, Service로 분류하였다.



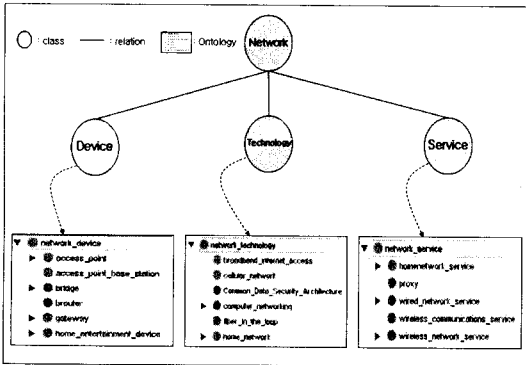
(그림 1) 개념 트리 작성 예

분류된 상위 개념 구조를 바탕으로 각각 구축자들은 하위 온톨로지를 구축하였다. 구축자들은 하위 온톨로지 구축 시 개념들 간의 관계를 정의할 때 일련의 개념 간 관계 정의에 제한적인 규칙을 적용하였다. 본 네트워크 분야 온톨로지에선 1차적 개념 간 관계 정의에서 is-a 관계만을 사용하도록 제한하였다.



(그림 2) 구축된 하위 온톨로지 구축 예

개별 작업자에 의해 구축된 하위 온톨로지들 간의 병합 과정을 거친다. 병합 과정을 통해 중복된 용어 삭제, 빠진 용어 추가 작업, 용어 정의문을 바탕으로 개념들 간의 관계 재정의 등의 작업을 수행하였다. 마지막 단계로서 병합된 온톨로지를 반복적인 정제 작업을 수행함으로써 좀 더 완전한 온톨로지를 구축하게 된다.



(그림 3) 구축된 네트워크 온톨로지

4. 결론 및 앞으로 과제

본 논문에서는 시맨틱웹의 중요한 요소인 온톨로지를 구축하는 데 있어 기존의 구축방법 모델들이 제안하는 기본적인 절차를 비교 분석하여 대용량 온톨로지를 협력적으로 구축하는 방법을 제시하고 이를 네트워크 분야 온톨로지 구축과정에 적용해 보았다. 각각의 구축자들은 하위 온톨로지 개발에서 개념 간 관계 정의를 제한적으로 규제받음으로써 하위 온톨

로지를 병합하는 데 있어 개념 간 관계 정의에 대한 일관성을 유지할 수 있었다. 하지만, 개념 간 관계 정의에서 관계 속성의 부족을 일으켰다. 이러한 개념 간 관계 속성의 부족은 반복적인 정제 과정을 수행함으로써 보완되어야 할 문제라 여겨진다.

앞으로 과제로는 온톨로지 구축과 관련하여 비전문가들도 효율적으로 온톨로지를 구축할 수 있는 지침, 개발방법 및 개발도구와 관련된 연구를 진행할 계획이다.

참고문헌

- [1] "Knowledge Processes and Ontologies (2001)" Steffen Staab, Rudi Studer, Hans-Peter Schnurr IEEE Intelligent Systems 16(1):26-34
- [2] Steve G, Gangemi A, Pisanelli DM, "Integrating Medical Terminologies with ONIONS Methodology", in Kangassalo H, Charrel JP (eds.) Information Modelling and Knowledge Bases VIII, Amsterdam, IOS Press 1997.
- [3] Fernandez, M., Gomez-Perez, A. and Juristo, N. (1997). METHONTOLOGY: From Ontological Arts Towards Ontological Engineering.
- [4] Asuncion Gomez-Perez, Mariano Fernandez-Lopez and Oscar Corcho "Ontological Engineering" Springer
- [5] Pascal Hitzler, Markus Krotzsch "What Is Ontology Merging?", 『AAAI』 2005.
- [6] Natalya Fridman Noy, Mark A. Musen "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment", 『AAAI』 2000.
- [7] Keonsoo Lee, Wonil Kim, Minkoo Kim "정보 공유를 위한 온톨로지 맵핑 방법", 『한국지능 정보 시스템학회』
- [8] 김은경, 남영준 "시맨틱웹을 위한 온톨로지 구축 방법에 관한 비교 연구", 『정보관리연구』 vol. 35. no. 2. 2004. pp. 57-85
- [9] 김재홍, 이상조 "지역 및 전역 의미집합을 이용한 온톨로지 병합 및 정렬 알고리즘", 『전자공학회 논문지』 제 41 권 CI편 제 4 호
- [10] 이지윤, 양진혁, 정인정 "지능형 유비쿼터스 웹을 위한 온톨로지 통합 및 확장"
- [11] 김재홍, 손주찬 "OWL 온톨로지 합병도구의 설계 및 구현", 『한국정보과학회』 가을 학술발표논문집 Vol. 30. No. 2