

# 이질의 도메인 온톨로지내의 개념간 병합 방법 연구

공현장\*, 황명권\*, 김판구\*

\*조선대학교 전자계산학과

e-mail:{kisofire, hmk2958, pkkim}@chosun.ac.kr

## A Study on the Methodology for Merging the Heterogeneous Domain Ontologies

Hyun-Jang Kong\*, Myung-Gwun Hwang\*, Pan-Koo Kim\*

\*Dept of Computer Science, Chosun University

### 요 약

시맨틱 웹의 개발에 대한 연구가 많이 진행됨에 따라 많은 도메인 온톨로지들이 구축되고 있다. 많은 개발자들이 연구 결과를 바탕으로 온톨로지를 제작하고 있으나, 개발자들의 개인 관점의 차이, 명확한 온톨로지 구축 방법론의 부재, 온톨로지 구축 도구의 다양성 등의 문제에 따라 동일한 주제의 온톨로지일지라도 내용에서 많은 차이점이 발생하고 있다. 이러한 온톨로지 언어적 수준, 온톨로지적 수준의 차이 때문에 각각 온톨로지의 상호운용성이 떨어진다. 또한 현재 온톨로지 구축 도구들의 기능들은 온톨로지 생성, 편집, 추론 등이 주를 이루고 있고, 동일한 주제에 대한 이질의 도메인 온톨로지를 병합하는 기능은 부족한 실정이다. 이에 본 논문에서는 각각의 개발자에 의해 개발된 동일한 도메인의 온톨로지들을 병합하여 더욱 완전한 온톨로지 구축을 위한 병합 방법을 연구하였다. 병합 방법은 대형의 온톨로지인 워드넷(WordNet)을 기반으로 하여, 이러한 이질의 도메인 온톨로지들을 공통된 주제를 중심으로 세부적인 단계를 거쳐 병합을 꾀하였다.

### 1. 서론

시맨틱 웹에 대한 연구가 가속화됨으로서, 온톨로지에 대한 관심이 더욱 커지고 있다. 그리하여 많은 온톨로지들이 구축되고 있다. 그러나 현재 온톨로지 구축을 위한 온톨로지 언어의 정립이 되지 않은 상황이고, 온톨로지 언어에 대한 사용자 지침도 많이 부족한 상태이다. 그리고 온톨로지 개발자들이 온톨로지 구축에 서로 다른 도구를 사용하고, 현재에도 그러한 많은 다른 도구들이 개발 되고 있다. 이에 따라 개발자들이 구축한 온톨로지는 같은 주제일지라도 개발자의 관점에 따라 서로 다른 도메인 온톨로지가 개발되고 있는 실정이다.

이렇게 제작된 온톨로지는 동일한 주제에 대하여 구축되었을지라도 개발자의 관점에 따라 서로 다른 관점에서는 다루지 못한 영역이 존재하게 된다. 그렇지만 이러한 이질의 온톨로지들의 처리가 서로 다른 별개의 온톨로지로서 처리되어서는 안 될 것이다.

이와 같은 온톨로지들에 대하여 더욱 완벽한 온톨로지 구축을 위해, 온톨로지들의 지침이 될 수 있는 검증된 대형의 온톨로지인 워드넷[1][2]을 중심으로 병합을 꾀하였다. 기존의 온톨로지 개발 도구에서도 온톨로지들을 병합 하는 방법에 대한 많은 연구가 이루어지고 있지만, 아직까지 부족한 현실이다.

본 논문에서는 이러한 같은 주제에 대하여 제작된 이질의 도메인 온톨로지를 위한 병합 방법에 대하여 연구하였다. 여기에서 연구된 새로운 방법은 대형의 워드넷을 기반으로 이용하는 것이다. 워드넷은 단어들 사이의 관계에 대해 상세하게 서술해 놓은 사전과 같은 것이다. 특히 동의어, 반의어를 비롯한 단어간의 상하관계에 대해 상세하게 서술되어 있어 개념들 사이의 관계를 명확히 할 수 있는 지침이 된다. 이러한 워드넷을 기반으로, 이질의 온톨로지들에서 사용된 개념들의 동일함을 비교하여 용어들 사이의 계층을 다시 구조화하여 병합한다.

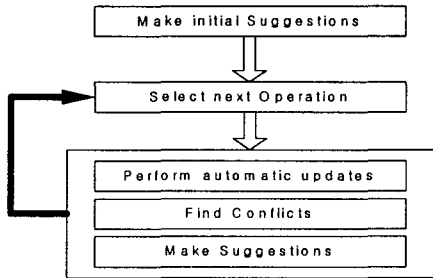
본 논문의 2장에서는 현재 온톨로지의 여러 병합 알고리즘에 대한 관련 연구를 설명한다. 그리고 3장에서는 본 논문의 핵심 주제인 이질의 도메인 온톨로지들의 병합 방법에 대하여 자세하게 설명하였다. 마지막으로 4장에서는 결론과 향후 연구 방향을 제시하였다.

2. 관련연구

동일한 주제에 대한 이질의 도메인 온톨로지의 병합을 위하여 기존에 연구된 온톨로지 병합 방법에 대하여 조사하였다.

2-1. The PROMPT Algorithm

다음의 [그림 1]은 PROMPT[3]의 온톨로지 병합 알고리즘이다.

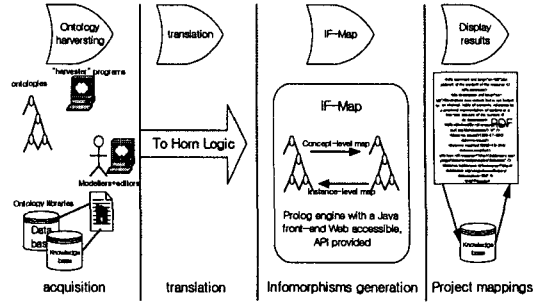


[그림 1] PROMPT 온톨로지 병합 알고리즘

PROMPT는 두개의 온톨로지를 그림과 같은 처리과정을 통하여 하나의 통합된 온톨로지를 만든다. PROMPT 알고리즘은 병합에 관하여 가능한 자동적으로 처리를 가능하도록 설계되었으며, 자동으로 처리하기 힘든 상황에서는 사용자가 개입하여 처리한다. PROMPT는 Protege-2000의 지식 모델링 환경에 기반하여 플러그인 형태로 개발되었으며, 여기에 기반이 되는 지식은 OKBC-compatible 지식 모델이다.

2-2. Information-Flow-based Ontology Mapping

IF-Map(Information-Flow-based Ontology Mapping)[4] 방법은 Information 흐름과 Channel 이론에 그 이론적 기반을 두고 있다. IF-Map의 전반적인 처리과정은 다음 [그림 2]와 같으며, IF-Map에서는 logic infomorphisms을 매핑을 위한 수식으로 사용하고 있다.



[그림 2] IF-Map 처리 과정

2-3. Formal Concept Analysis Merge

FCA-Merge(Formal Concept Analysis Merge) [5]는 Stumme와 Maedche에 의해 개발되었으며, 그들에 의해 개발되어진 개념의 격자를 조사하는 FCA-based 방법을 자연어 처리 기술에 접목시킨 방법이다. 이러한 격자는 FCA-Merge를 이용하여 반자동의 방법으로 온톨로지를 병합하기 위한 온톨로지 개발자에게 유용하다.

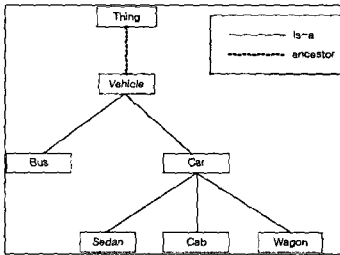
3. 이질의 온톨로지 병합을 위한 방법

본 연구는 온톨로지 병합을 위해 대형의 온톨로지인 위드넷을 이용한다. 위드넷은 현재까지는 전 세계적으로 가장 널리 사용되어지고 있는 대형의 온톨로지이다. 만약 서로다른 사람이 같은 도메인에 대한 온톨로지를 각각 만든다면, 많은 외부적인 요인에 의하여 상이한 온톨로지가 개발될 것이다. 물론 비슷한 부분도 많이 있겠지만, 똑같은 온톨로지를 기대하기는 힘들 것이다.

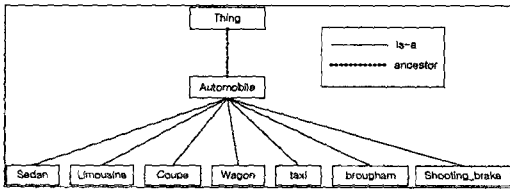
서로 상이한 온톨로지를 구축하는 현재의 가장 큰 이유는 첫째, 온톨로지 개발자의 개인적인 관점의 차이이다. 모든 사람들은 각각 사물을 바라보는 관점이 다르고, 환경적 차이가 있기 때문에 발생하는 요인이다. 둘째, 현재까지 명확한 온톨로지 구축을 위한 방법론이 존재하지 않는다. 온톨로지에 대한 관심이 커지면서 온톨로지를 중점적으로 연구하게 된 건 불과 몇 년되지 않는다. 또한 온톨로지를 구축하기 위해 과거에 철학적관점의 존재론에서 제기된 구축 방법론을 이용하여 온톨로지를 구축하는 것은 시맨틱 웹 온톨로지의 이상과 거리가 있기 때문이다. 셋째, 온톨로지 구축 도구의 다양성이다. 현재 온톨로지에 대한 활발한 연구는 다양한 온톨로지 구축 도구를 살펴봄으로써 알수 있다. 그렇지만 이 모든 온톨로지 구축 도구가 같은 온톨로지 언어를

사용하지 않고, 같은 방법으로 온톨로지를 만들고 있지 않으므로 각각의 온톨로지에 대한 상호운용성이 떨어진다.

본 논문에서는 이러한 다양한 이질의 온톨로지 개발에 대한 배경에 전제하여, 이러한 온톨로지의 병합을 위한 새로운 방법을 연구하였다. 앞에서 제시한 세 가지의 배경에서, 특히 본 연구에서는 첫 번째로 제시한 온톨로지 구축자들 사이의 관점의 차이에 초점을 두고 있다. 예를 들어, 자동차 도메인에 대한 온톨로지를 두 명의 구축자가 구축을 하였다고 가정하여 보자. 한명의 개발자는 자동차에 대한 전문적인 지식을 가지고 그 분야의 전문가로서 전문가용 온톨로지 구축을 목적으로 온톨로지를 구축하였고, 다른 한명은 자동차에 대한 가장 기본적인 지식 체계를 온톨로지로서 구축하고자 한다. 첫 번째 개발자의 온톨로지에는 두 번째 개발자가 알지 못하는 많은 정보가 들어있을 것이지만, 물론 두 번째 개발자에 의해 구축된 온톨로지에도 첫 번째 개발자가 다루지 않은 면이 존재할 것이다. 다음은 위의 개발자들이 개발한 두개의 온톨로지의 일부분들이다.



[그림 3] Car 온톨로지

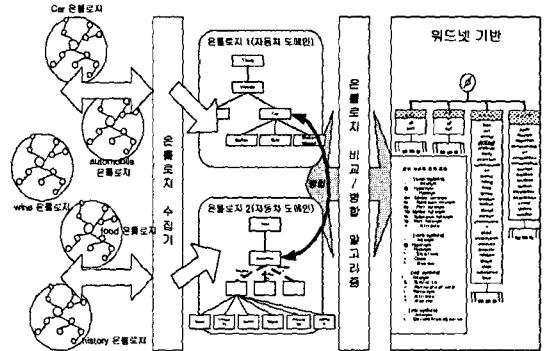


[그림 4] Automobile 온톨로지

간단하게 위의 두 온톨로지는 같은 자동차에 대한 온톨로지를 한 개발자는 "Car"로, 또 다른 개발자는 "automobile"로 작성하였다. 이러한 간단한 차이는 갈수록 많아지는 온톨로지에서 더욱 많아져서 앞으로는 큰 차이를 보이며 발생 할 것이다. 이러한 수많은 이질의 도메인 온톨로지를 병합하여 좀더 완벽한 온톨로지 구축을 위한 병합 방법에 대한 연구

가 요구된다.

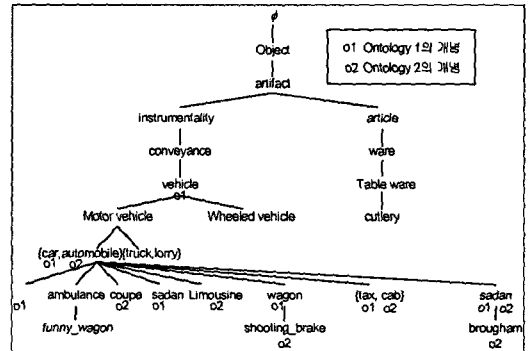
본 연구에서 제안하는 이질의 도메인 온톨로지 병합 방법은 워드넷에 기반하고 있으며, 워드넷의 내부적 특성을 최대한 이용하여, 이질의 도메인 온톨로지 내의 개념들 사이의 동의관계와 상하위관계를 재조정함으로써, 새로운 통합 온톨로지 구축을 꾀한다.



[그림 5] 병합을 위한 처리 과정

위의 전체적인 처리과정은 다음의 몇 단계로 세분화하여 설명할 수 있다. 본 연구에서 제시하는 병합의 단계는 다음과 같다.

[단계 1] 각 온톨로지1 과 온톨로지2의 모든 개념들을 기반 온톨로지인 워드넷에 매칭 시킨다.



[그림 6] 워드넷에 각 온톨로지의 개념 매칭

여기에서는 모든 개념들의 워드넷내의 위치 파악이 가능하다.

[단계 2] 각 개념들에 대하여 워드넷의 Synset\_number를 이용하여 동일한 개념을 파악한다.

[표 1] Synset\_number 비교

온톨로지1 \ 온톨로지2	vehicle	bus	car	sedan	cab	wagon
automobile			1			
sedan				1		
limousine						
coupe						
wagon						1
taxi					1	
brougham						
shooting_brake						

각 온톨로지의 동의어를 파악하기 위하여 워드넷 내의 synset\_number를 이용하여 위의 표와 같은 처리를 하였다. '1'의 값을 가지는 개념들은 동의어를 의미한다.

[단계 3] 각 개념들에 대하여 워드넷의 PTR을 이용하여 상·하위 관계를 파악한다.

[표 2] PTR 비교

온톨로지1 \ 온톨로지2	vehicle	bus	car	sedan	cab	wagon
automobile	@	~		~	~	~
sedan	@		@			
limousine	@		@			
coupe	@		@			
wagon	@		@			
taxi	@		@			
brougham	@		@	~		
shooting_brake	@		@			~

PTR은 워드넷내의 개념들 사이의 관계정의를 위한 필드로서, 여기에서 '@'과 '~'을 이용하여 각 개념들 사이의 상·하위 관계를 파악한다.

[단계 4] 명확한 상·하위 관계의 정의를 위하여 개념들중 최상위 개념으로부터의 각각 깊이를 구한다.

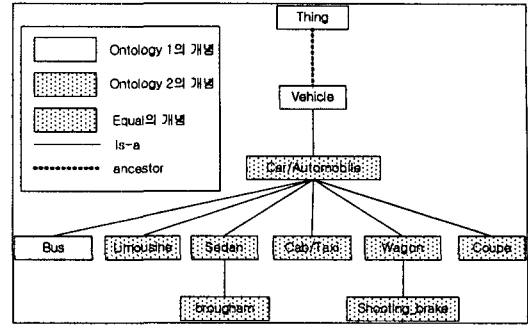
[표 3] 각 개념들의 깊이

온톨로지1	vehicle	bus	car	sedan	cab	wagon		
깊이	0	3	2	3	3	3		
온톨로지2	automobile	sedan	limousine	coupe	wagon	taxi	brougham	shooting_brake
깊이	2	3	3	3	3	3	4	4

[단계 3]에서 상·하위 관계의 파악이 가능하지만, 명확한 계층 구조를 만들기 위해서 각 개념들의 레벨을 파악해야 한다. 이를 위해서 여기에서는 각 개념들에 대하여 도메인의 최상위 개념으로부터 각 깊이를 구한다.

[단계 5] 모든 단계를 거쳐 새롭게 병합된 온톨로

지를 생성한다.



[그림 7] 통합 온톨로지

4. 결론

시맨틱 웹의 발전으로 많은 온톨로지가 웹에 존재하게 될 것이며, 또한 이러한 온톨로지는 여러 가지 요인에 의하여 각각 이질성을 가질 것이다. 그리하여 본 연구에서는 앞에서 설명한 여러 요인 중에서 개발자의 관점의 차이에 의해 서로 다르게 만들어진 온톨로지들을 병합하기 위한 방법을 연구하였으며, 여기에는 워드넷내의 구조적 특성을 이용하는 연구를 하였다.

본 연구에서는 온톨로지 구축자들의 관점 차이에 초점을 맞추어 이질의 온톨로지의 병합을 위한 방법을 연구 하였으며, 향후에는 두 번째, 세 번째 요인을 해결하기 위한 방법과 실제 구현을 통한 입증 이후 연구과제로 남아있다.

참고문헌

[1] George A. Miller "Introduction to WordNet: An On-line Lexical Database" 1993  
 [2] Jay J Jiang, David W.Conrath "Semantic Similarity Based Corpus Statistics and Lexical Taxonomy" 1997  
 [3] N.Fridman Noy and M.Musen. "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment." In Proceedings of the 17th National Conference on Artificial Intelligence, (AAAI'00), Austin, TX, USA, July 2000.  
 [4] Y.Kalfoğlu and M.Schorlemmer. "Information-Flow-based Ontology Mapping" Journal on Data Semantics 1(1):pp.98-127, 2003.  
 [5] G.Stumme and A.Maedche. "Ontology Merging for Federated Ontologies on the Semantic Web." In proceedings of the International Workshop for Foundations of Models for Information Integration(FMII-2001), Viterbo, Italy, September 2001.