

# 시맨틱 웹에서의 도메인 온톨로지 구축 및 적용

공현장\*, 정관호\*, 신주현\*, 김원필\*, 김판구\*

\*조선대학교 컴퓨터공학부

e-mail : {kisofire, khjung, jhshin, wpkim, pkkim}@mina.chosun.ac.kr

## Building Domain Ontology for Semantic Web

Hyunjang Kong\*, Kwanho Jung\*, Juhyun Shin\*, Wonpil Kim\*, Pankoo Kim\*

\*School of Computer Engineering, Chosun University

### 요 약

1990년대 중반부터 최근까지 시맨틱 웹에 대한 많은 관심과 더불어 많은 연구가 진행중이다. 무한한 정보 자원을 가지고 있는 인터넷에서 자원에 대한 효율적 처리가 더욱더 강조된다. 그렇지만 시맨틱 웹에 대한 뚜렷한 결론을 내리기 힘들뿐만 아니라, 지금의 연구들에서는 시맨틱 웹에 대한 전체적 구상에 치중하고 있을 뿐, 세부적인 기술에 관한 연구는 미흡하다. 최근까지의 연구의 초점은 주로 XML, XML Schema에서 RDF, RDF Schema 그리고 DAML+OIL에 이르기까지 다양한 마크업 언어의 개발 및 적용에 대한 연구이다. 이러한 연구의 결과 시맨틱 웹에서의 표현을 위한 마크업 언어에 대한 많은 성과를 가져왔지만, 시맨틱 웹의 핵심이 되는 정보의 의미적 표현은 더 많은 연구들이 요구된다. 본 논문은 시맨틱 웹의 핵심적인 부분을 차지하고 있는 온톨로지에 대한 연구이다. 최근 널리 사용되어지고 있는 온톨로지 중 하나인 WordNet을 시맨틱 웹의 온톨로지로 적용함에 있어, 발생하는 문제점을 해결하기 위한 방법을 제시한다. WordNet에 기반 한 도메인 온톨로지의 구축 및 적용에 대한 내용이 이 문제점을 해결하기 위한 본 논문의 요지이다.

### 1. 서론

인터넷은 그 안에 무한한 정보를 가지고, 광범위하게 사용된다. 사용자는 인터넷상에서 많은 정보를 효과적으로 창출하고자 한다. 따라서 1990년대 중반 웹의 창시자 팀버너스리는 컴퓨터가 웹 상의 정보를 이해하여 사람과 서로 자유로운 정보의 전달이 가능한 시맨틱 웹(Semantic Web)을 차세대의 웹으로 제안하였다.[1] 시맨틱 웹은 사용자가 필요로 하는 정보를 컴퓨터가 이해하여 정보를 검색하고 구성·제공하는 등의 일련의 작업을 수행하는 지능적인 웹이다. 현재까지 시맨틱 웹의 시스템에 대한 구조는 구체적이지 않으며, 시맨틱 웹에 대한 최근의 많은 연구는 마크업 언어에 대한 연구가 주를 이룬다. 현재까지의 시맨틱 웹에 대한 연구를 보면 만족할 만큼의 결과도 있지만, 아직까지는 미흡한 것으로 보인다. 그래서 시맨틱 웹에 대한 지속적인 연구가 요구된다 할 수 있겠다.

본 연구에서는 현재 사용되어지는 웹에서 발생한 문제점을 해결하고 지능적인 정보의 처리가 가능하도

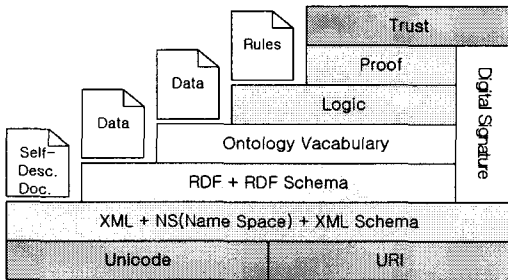
록 하는 온톨로지 부분에 대한 연구에 그 초점을 두고 있다. 팀버너스리가 처음 시맨틱 웹에 대한 계층구조를 발표했을 때, 온톨로지는 구조의 핵심부분을 차지하고 있으며, 온톨로지 대한 중요성과 문제점 그리고 문제점을 해결하기 위한 도메인 온톨로지의 구축과 적용까지의 전반적인 연구에 관한 내용이다. 본 논문에서는 온톨로지의 일종인 WordNet을 기반으로 진행중인 연구 내용을 제시한다.

2장에서는 시맨틱 웹과 관련된 최근 연구내용 중에서 마크업 언어들과 온톨로지에 대한 연구 내용을 살펴본 후, 3장에서는 시맨틱 웹에서 요구되는 온톨로지 시스템과 온톨로지 시스템 구축의 난점과 온톨로지 적용시의 발생하는 문제점 및 그 해결방안을 살펴본다. 4장에서는 WordNet 적용 시 발생 가능한 문제점을 해결하기 위한 새로운 방안으로 도메인 온톨로지 시스템에 대한 내용을 다루고, 다음 장에서는 이를 적용해 시맨틱 웹의 전체적인 구조를 재구성한다. 끝으로 6장에서는 결론 및 향후 연구과제를 기술한다.

## 2. 관련연구

### 2.1 팀버너스리의 시맨틱 웹

웹의 창시자 팀버너스리는 low-level 문법을 사용하는 Syntax 계층과 high-level 문법을 사용하는 Semantic 계층 그리고 데이터간의 상호 이해와 표현을 가능하게 도와주는 온톨로지 계층으로 크게 구성된 시맨틱 웹의 구조를 제안하였다. [그림 1]은 팀버너스리가 제안한 시맨틱 웹의 계층구조이다.[1][2]



[그림 1] 팀버너스리의 Semantic Web "layer cake"

[그림 1]은 시맨틱 웹의 계층 구조와 각 계층에 이용되는 마크업 언어에 대한 내용을 보여주고 있다.

### 2.2 마크업 언어

다음으로 시맨틱 웹에 사용되는 마크업 언어에 대하여 간략히 살펴본다. 그렇지만 소개하는 마크업 언어의 문법은 본 논문의 요지와 벗어나는 내용이므로 언급하지 않았다.

시맨틱 웹은 XML과 네임 스페이스, URI 등의 기술을 기본으로 사용하고 있다. XML은 사용자가 필요에 따라 태그를 정의하여 사용할 수 있는 장점을 가지고 있어서 시맨틱 웹의 Syntax 구조를 구성하는데 사용되는 마크업 언어이다.[3] XML에서 특히, 네임 스페이스는 XML 문서작성 시 발생 가능한 이름의 충돌을 피하기 위한 수단으로 제공되는 중요한 요소이다. XML Schema는 DTD(Document Type Definition)의 개선안으로 문서의 문법, 어휘, 구조, 데이터형 등을 규정하고 있다. 그렇지만 XML은 초기 시맨틱 웹의 모델에서 마크업 언어로 제안되었지만, 언어의 특성상 단순한 계층적 구조를 표현하기에는 적합하지만, 의미적 표현에 미흡한 단점을 가지고 있다.

RDF(Resource Description FrameWork)는 웹 자원을 의미적으로 기술하기 위한 마크업 언어로 현재 시맨틱 웹의 마크업 언어 표준으로 삼고 있다.[4] RDF

는 XML을 기반으로 만들어진 마크업 언어로써 XML의 부족함을 해결할 수 있는 가능성을 제공한다. RDF는 다양한 메타데이터 사이의 연결을 위하여 Semantic, Structure, Syntax에 대한 표준화를 제공한다. RDF는 세 가지의 특성을 표현하고 있으며, 이는 각각 객체(Object), 속성(Attribute) 그리고 값(Value)이다. 그리고 RDF Schema는 자원들의 특성을 기술하며, 자원과 자원사이의 관계와 제약 조건을 기술하고 정의한다.

DAML+OIL은 웹 자원에 대한 온톨로지의 적용과 표현을 위한 언어이다.[5] DAML+OIL은 단어 사이의 명확한 의미를 규정한다. OIL은 시맨틱 웹에서의 사용을 목적으로 온톨로지를 표현하기 위한 수단이며, RDF와 비교적 호환성을 지닌다. 최근 W3C는 DAML+OIL을 업그레이드 한 OWL(Web Ontology Language)[6]을 발표하였으며, 이는 DAML과 OIL의 장점을 수용하여 온톨로지를 간결하게 표현하는 의미적 마크업 언어이다.

### 3. 시맨틱 웹에서 온톨로지 중요성

시맨틱 웹에서 가장 핵심적인 부분을 차지하고 있는 부분 바로 온톨로지이다. 온톨로지는 과거 수십 년 전부터 활발히 연구되어져 온 분야로써, 주로 자연어 처리나 인공지능 분야, 지식표현의 분야에서 많은 연구가 진행되어져 왔다. 온톨로지에 대한 많은 정의가 있지만 그 중에서 시맨틱 웹에 적합한 정의는 '개념들간의 의미적 표현'이 가장 잘 어울리는 정의일 것 같다. 시맨틱 웹은 그 목적에서 바로 웹 상의 정보들에 대한 의미적 결합이므로 온톨로지의 이론과 일맥상통하는 부분이라고 생각된다. 온톨로지에 관한 여러 연구로 온톨로지가 구축되었으며, 대표적인 예로, WordNet, UMLS, CYC 등이 있으며, WordNet은 각 개념들에 대하여 체계적으로 구성되어 있으며, RDF에서 표현하고자 하는 개념들간의 관계를 잘 표현하고 있는 대표적인 온톨로지의 예이다.

이에, 본 논문에서는 온톨로지의 대표적인 예로 WordNet을 시맨틱 웹의 온톨로지 계층에 적용하는 방안을 중점적으로 다루고 있다. 주요 내용은 WordNet을 적용하였을 때, 발생하는 문제점들에 대한 것으로 이를 해결하기 위해서 도메인 온톨로지를 생성, 적용하는 방법을 제시하고자 한다.

WordNet은 단어들간의 관계를 잘 표현하고 있을 뿐만 아니라, 많은 개념을 체계적으로 포함하고 있는 대형의 온톨로지이다.[7] 그렇지만 많은 개념을 가지

고 있고, 관계가 잘 표현되어 있는 대형의 온톨로지라도 몇 가지 문제점은 가지고 있다. 먼저, 개념들의 시대적 낙후성이다. WordNet에서 포함하고 있는 개념들은 많은 개념들이 최근에 발생한 새로운 개념들을 포함하고 있지 않다. 자동차의 예를 들어보면, 첫째로, "chairman"이라는 단어는 WordNet내의 개념은 "회사의 최고 위치에 있는 사람"과 같은 개념으로 표현된다. 그렇지만 "chairman"은 "자동차의 이름"으로 최근에 많은 사람들에게 인식되어져 있다. 둘째로, 시대가 변화에 따라 새로이 발생한 개념들은 기존의 WordNet에서 포함하지 않는 문제점이다. 예를 들어, 대표적인 자동차 회사인 "volvo"는 WordNet내의 개념들에 포함되어 있지 않다. 이러한 문제점을 안고 있는 WordNet을 그대로 Semantic Web의 구축에 적용을 한다면, 분명 예기치 않은 문제점이 발생할 것이다. 그러므로, 본 논문에서 주장하고자 하는 내용은 이러한 WordNet을 좀더 효과적인 방법으로 적용하는 방법이다. 이를하여 새롭게 생겨난 개념들의 묶음을 WordNet 구축의 방법에 따라 특정 도메인 온톨로지를 구축하여 WordNet에 결합하여 사용하는 것이다.

#### 4. WordNet에 기반한 도메인 온톨로지 구축

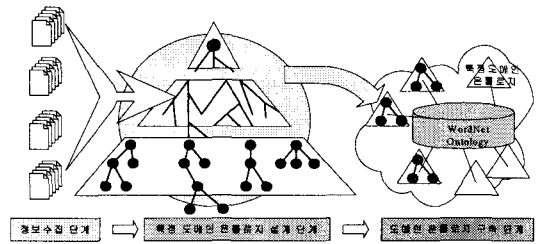
본 논문에서 제안하고자 하는 특정 도메인 온톨로지의 구축 방법은 이렇하다. 일반적으로 온톨로지의 구축 단계는 다음과 같다. 첫째로, 구축하고자하는 온톨로지를 대표할 수 있는 몇 개의 문장을 작성한다. 다음으로, 온톨로지가 어떠한 내용을 포함하길 원하는지 리스트를 만들고, 온톨로지에서 포함할 개념들의 목록을 만든다. 그리고 다양한 방법으로 포함될 개념들의 정보를 수집한다. 끝으로 위의 방법을 반복 수행하면서 온톨로지를 완성한다. 그리고 이렇게 구축된 특정 도메인 온톨로지를 그 뼈대가 되는 WordNet과 결합함으로써 시맨틱 웹의 구축에서 더욱더 큰 기능을 기대할 수 있다.

도메인 온톨로지의 구축은 일반적으로 조립(Composing), 가지치기(Pruning), 뽑아냄(Extracting) 그리고 병합(Merging)의 네 가지의 구축 모듈을 이용한다.[8]

조립(Composing) 모듈은 인공의 물체, 시간 혹은 공간과 같은 실세계의 모델화를 위해 요구되어지는 개념들의 형식화와 표현을 위한 기초적 묶음을 포함하고 있다. 가지치기(Pruning) 모듈은 구축하고자 하는 도메인 온톨로지에 적합하지 않은 개념들을 대형 온톨로지로부터 나무의 가지치기를 하듯 잘라낸 개념

들이다. 뽑아냄(Extracting) 모듈 역시 가지치기 모듈과 비슷하지만 이 방법은 도메인 온톨로지에 키워드가 될 단어들을 우선 대형의 온톨로지서 선정할 후 그 단어에 연결된 모든 링크와 그에 해당하는 개념을 뽑아내서 새로운 도메인을 형성하는 방법이다. 끝으로, 병합(Merging) 모듈은 새로이 구축할 도메인 온톨로지에 적합하여 선정된 부분과 새로이 추가할 개념들을 병합하는 모듈이다.

본 논문에서 제안한 도메인 온톨로지의 구축과 구축된 특정 도메인 온톨로지들과 WordNet과의 결합을 통한 완벽한 구축과정을 간단히 살펴보면 다음과 같다. [그림 2]는 도메인 온톨로지의 구축 방법과 WordNet과의 결합을 보여준다.



[그림 2] 도메인 온톨로지 구축 과정

[그림 2]의 첫 번째 정보수집 단계에서는 특정 도메인 온톨로지 구축을 위한 다양한 정보를 수집하는 단계이다. 이 단계에서 중요한 점은 일반적이면서 모두가 공감하는 많은 정보를 수집하는 것이다. 다음은 특정 도메인 온톨로지 설계 단계로 맨 위의 상위에 특정 주제를 선정하고 그에 따른 수집된 정보들을 계층적으로 구성하고 연결한다. 끝으로 각각의 구축된 특정 온톨로지를 WordNet의 해당 부분에 결합하여 완벽한 도메인 온톨로지를 구축한다.

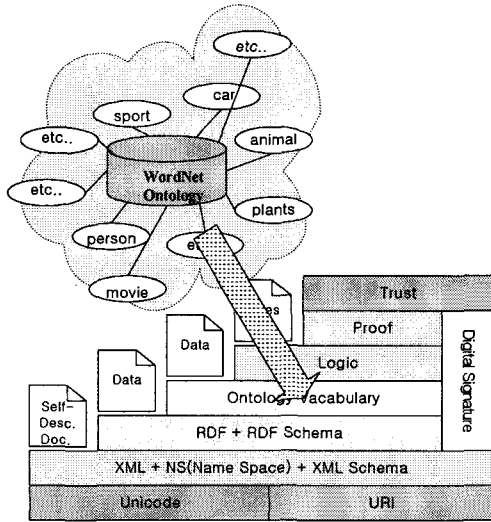
#### 5. 제안한 시맨틱 웹

본 논문에서는 시맨틱 웹의 계층구조 안에 WordNet에 기반한 도메인 온톨로지를 구축·적용함으로써 온톨로지 적용의 이점을 극대화하고자 한다.

논문에서 제안한 시맨틱 웹의 새로운 구성도는 초기 시맨틱 웹의 온톨로지 계층에 WordNet에 기반하여 구축된 도메인 온톨로지를 적용하여 그 기능면에서 좀더 향상된 성능을 기대한다.

구성도는 다음의 [그림 3]과 같다.

참고문헌



[그림 3] Semantic Web 계층구조에 WordNet을 적용한 새로운 모델링

위의 계층도는 초기의 시맨틱 웹의 구성에서의 온톨로지 적용 시 발생하는 문제점을 보충하기 위함의 목적을 두고 있다.

6. 결론

본 논문의 연구에서 시맨틱 웹에 WordNet과 이에 기반하여 구축된 도메인 온톨로지를 결합한 형태의 좀더 완벽한 온톨로지를 구축하여 시맨틱 웹의 온톨로지 계층에의 적용에 관한 내용이다. 본 연구는 도메인 온톨로지 구축과 평가에 대한 연구는 많이 이루어졌지만, 구축된 온톨로지의 적합성을 평가하기가 힘들어, 실질적인 시맨틱 웹의 구축에 적용하기엔 아직까지 거리가 있다. 그리고 도메인 온톨로지를 구축하는 방법상의 정확한 절차에 대하여서도 문제점이 발생하였으며, 구축하였을 경우, 도메인 온톨로지가 과연 적합하게 구축이 되었는가? 에 대한 의문점도 가지고 있다. 그렇지만 본 논문에서는 도메인 온톨로지의 구축에 대한 내용을 목적으로 한 것이 아니라, 시맨틱 웹의 구축에서 온톨로지를 적용할 때, WordNet과 같은 대형의 온톨로지의 적용 시 발생가능한 문제점의 해결을 도메인 온톨로지의 개념으로 해결하고자 제시한 것이다.

향후, 도메인 온톨로지의 구축에 대한 적합성 판단 여부와 또 이렇게 완성된 도메인 온톨로지의 시맨틱 웹에의 적용방안 등 많은 연구가 앞으로 요구된다.

- [1] Tim Verner-Lee, James Hendler, and Ora Lassila. The Semantic web. *Scientific American*, May 2001
- [2] S.A. McIlraith, T.C. Son, H.Zeng. Semantic Web Service, *IEEE Intelligent System*, March/April 2001
- [3] Tim Bray, Jean Paoli and Eve Maler. Extensible Markup Language(XML) 1.0 (Second Edition). 2000, *W3C: Recommendation(October 2000)*
- [4] O.Lassila and R.Swick, Resource Description Framework(RDF) Model and Syntax Specification, W3C Recommendation, *World Wide Web Consortium*, Boston, 1999, [www.w3.org/TR/REC-rdf-syntax](http://www.w3.org/TR/REC-rdf-syntax)
- [5] Dieter Fensel, IanHorrocks, Frank van Harmelen, Deborah L. McGuinness, and Peter F. Patel-Schneider. OIL:An ontology infrastructure for the semantic web. *IEEE Intelligent Systems*, 16(2), May 2001
- [6] Michael K. Smith, Deborah McGuinness, Raphael Volz and Chris Welty, Web Ontology Language (OWL) Guide Version 1.0, 2002
- [7] R. Mihalcea and D. Moldovan. Semantic indexing using WordNet senses. *In Proceedings of ACL Workshop on IR & NLP*, Hong Kong, October 2000
- [8] Robert MacGregor & Ramesh S. Patil, Tools for Assembling and Managing Scalable Knowledge Bases , CA 90292