

특정 도메인에 대한 자동 온톨로지 구축 방법에 관한 연구

공현장^o 황명권 김원필 김판구

조선대학교 전자계산학과

{kisofire^o, hmk2958, wonpil, pkkim}@chosun.ac.kr

The Study on the Autometic Ontology Building Methodolgoxy about the Specific Domain Knowledge

Hyunjang Kong^o Myunggwon Hwang Wonpil Kim Pankoo Kim

Chosun University, Dept. of Computer Science

요 약

현재 온톨로지는 많은 분야에서 활발하게 연구되고 있다. 특히 시맨틱 웹에 대한 연구가 활발히 이루어지면서 온톨로지의 활용 범위가 넓어졌다. 그 결과, 많은 온톨로지가 웹 상에 존재하게 되었으며, 이러한 이론적 배경으로 시맨틱 웹의 관점이 아닌, 온톨로지 연구의 측면에서 여러 가지 문제가 발생하게 되었다. 많은 문제들 중에서 동일한 도메인에 대한 온톨로지들 간의 이질성과 온톨로지의 일관성 문제 해결은 현재 가장 이슈가 되는 내용들이다. 이에 본 연구에서는 이러한 온톨로지 구축에서 발생하는 문제점들 중에서 기존의 지식 정보의 효율적인 재사용과 그에 기반한 온톨로지 구축 자동화의 가능성을 보인다. 또 이러한 기존 지식 체계에 기반함으로써 얻을수 있는 온톨로지에 대한 신뢰성의 문제 또한 해결하고 한다. 그리하여, 우리는 범용의 대형 온톨로지인 워드넷을 기반으로 한 자동 온톨로지 구축을 위한 방법론을 제안하고, 이를 적용한 온톨로지 자동 구축 시스템을 설계하고 구현하였다. 그리하여 사용자가 좀더 쉽게 온톨로지를 구축할 수 있기를 기대할 수 있다.

1. 서 론

시맨틱 웹에 대한 연구가 활발히 이루어지면서 온톨로지의 활용 범위 또한 넓어지고 있다. 그리하여 많은 사람들이 온톨로지 구축을 원하고, 많은 온톨로지가 웹 상에 존재하게 되었다. 이러한 배경으로 시맨틱 웹의 관점이 아닌 온톨로지 연구의 측면에서 여러 가지 문제가 발생하게 되었다.[3][4] 특히 온톨로지 구축시에 동일 도메인에 대하여 많은 온톨로지들이 구축되고, 각각의 온톨로지는 동일 도메인을 주제로 구축되었을 지라도, 서로 이질성을 갖는다. 이는 온톨로지의 재사용성의 특징을 해치는 주 요인이 되고 있다. 그리고 온톨로지 구축 방법론의 다양성과 구축 언어 및 구축 도구의 다양성 또한 많은 이질의 온톨로지가 구축된 주 원인으로 작용하고 있다. 이중에서도 특히 문제가 되는 것은 기존의 많은 양의 구조화된 지식들을 활용하기 위한 방법론이 없이, 많은 지식 정보가 그대로 방치되는 것이다. 그래서 사용자가 온톨로지를 구축할 때, 클래스 하나에서 인스턴스 하나까지 사용자가 직접 온톨로지 저작 도구를 사용하여 입력하여, 많은 시간을 온톨로지를 구축하는 데 투자해야만 한다. 또 이렇게 구축되어진 온톨로지를 어떻게 평가 할 것인가? 과연 온톨로지가 완벽하게 구축되었는가?에 대한 의문에 대답해 줄 기준도 아직까진 뚜렷하지 않은 상태다.

이에 본 연구에서는 이러한 온톨로지 구축의 어려움 중에서 기존 지식 정보의 효율적인 재사용과 온톨로지 구축의 자동화 그리고 이러한 기존 지식 체계에 기반함으로써 얻을 수 있는 온톨로지 신뢰성의 문제를 해결하는데 그 목적이 있다. 그리하여 본 연구에서는 대형의

지식 체계인 워드넷을 기반으로 사용자가 원하는 도메인에 대한 자동 온톨로지 구축을 위한 방법론과 도메인 지식 전문가에 의해 제공된 체계적인 데이터를 입력 형식으로 하여, 이러한 데이터 입력에 의한 자동 온톨로지 구축을 위한 방법론을 적용한 자동 온톨로지 구축 시스템을 설계하고 구현하였다. 본 시스템에서는 워드넷 데이터베이스를 입력으로 하고, 사용자가 원하는 각 도메인에 대한 계층 프레임을 1단계로 구축한 후, 도메인 전문가에 의해 준비된 특정 형식의 입력 정보를 추가적으로 처리하여 대표적인 온톨로지 언어인 OWL(Web Ontology Language) 형식의 온톨로지로 구축 할 수 있다.

본 논문의 제 2장에서는 관련연구인 워드넷과 온톨로지 구축 시스템에 대하여 조사하여 설명하고, 3장에서는 본 시스템에서 제안하는 온톨로지 자동 구축 방법론에 대하여 자세하게 소개한다. 그리고 제 4장에서는 실제 구현 시스템의 활용에 대하여 기술하고, 5장에서는 결론과 향후 연구과제를 제시한다.

2. 관련연구

2.1. 워드넷(WordNet)

워드넷은 현재까지 가장 널리 사용되는 범용의 대형 온톨로지로서 실제 그 내용은 6개의 데이터베이스 테이블들로 구성되어져 실제계에 존재하는 어휘에 대해서 체계적으로 정의하고 있다. 워드넷에서 중요한 내용은 바로 개념간의 관계를 정의하고 있는 부분이며, 사전과 가장 큰 차이점이 또한 바로 이러한 부분이다.[1] 워드넷 내의 6개의 테이블의 연결 내용을 살펴보면 다음 [그림 1]과 같다.

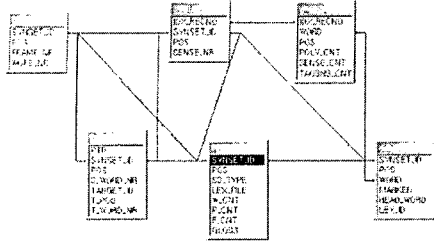


그림 1. WordNet 데이터베이스 관계도

워드넷은 크게 4개의 카테고리(명사, 형용사, 부사, 동사)로 분류되고, 그 안에는 다시 45개의 소카테고리로 분류되어져 있다. 그리고 워드넷내의 모든 개념들은 특정의 심볼들을 사용하여 각 개념들간의 관계를 표현하고 있다. 워드넷에서 사용되는 개념들간의 관계를 정의해 놓은 심볼들과 그 의미 그리고 워드넷의 내부 구조에 대한 내용은 다음 [그림 2]에서 잘 설명하고 있다.

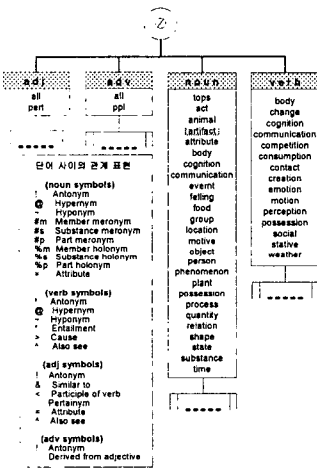


그림 2. 워드넷 구성도

이렇게 워드넷은 개념들간의 관계를 체계적으로 정의해 놓은 대형의 범용 온톨로지이지만 실제 내용은 데이터베이스 기반으로 구성되어져 있어서 모든 처리는 데이터베이스 기반 특정 쿼리와 데이터베이스 기반 특정 어플리케이션을 통하여서만 이루어져야 한다. 그래서 이를 시맨틱 웹 환경에 사용하기에는 우리가 따른다. 그리하여 본 연구에서는 이러한 데이터베이스 기반의 내용을 분석하여 OWL 온톨로지로 변형함으로써 높은 재사용성을 기대할 수 있다.

2.2 온톨로지 구축 도구와 그 한계점

온톨로지에 대한 연구가 활발히 진행되면서, 많은 온톨로지 구축 도구들이 개발되었다. 대표적으로 KAON, Protege 3.1, OilEd, OWL Editor 등과 같은 온톨로지 구축 도구가 개발되어져 사용되고 있으며, 이들 도구들은 모두 GUI 기반의 시스템들로 쉽게 온톨로지를 구축할 수 있도록 사용자 인터페이스를 제공하고 있다.[2] 그렇

지만 이렇게 많은 도구들에서 제공하지 않는 부분이 바로 온톨로지를 자동으로 구축 할 수 있는 방법론과 기존의 지식 정보를 재사용하기 위한 방법론의 접목이다. 그리하여 기존의 온톨로지 구축 도구를 사용하면, 사용자는 온톨로지 구축을 처음부터 끝까지 직접 해야 한다. 본 연구에서의 온톨로지 구축 시스템은 온톨로지 자동 구축을 위한 방법론을 적용하여 사용자가 좀더 쉽게 온톨로지 구축을 위한 개념들을 수집하고 계층을 구축 할 수 있어 시간적 측면과 일관성의 측면에서 큰 효율성을 기대할 수 있다.

3. 워드넷 기반 자동 온톨로지 구축 방법론

워드넷에 기반한 자동 온톨로지 구축 방법은 워드넷 내의 6개의 테이블 중에서, Pointers, Synsets, Words 테이블을 사용하고 있다. 여기에서는 워드넷에서 정의되어 있는 명사, 동사, 형용사, 부사 중에서 명사 부분에 대해서만 자동 구축의 입력 양식으로 사용하였다. 그래서 명사들간의 관계를 정의하는 워드넷 심볼들과 OWL Vocabulary들을 최대한 매칭하여 자동 구축의 형식을 만들었다. 워드넷 심볼과 OWL Vocabulary 매칭의 내용은 다음의 [표 1]과 같다.

표 1. OWL 어휘들과 매칭

워드넷	OWL 어휘
!	owl:disjointWith
@	owl:superClassOf
~	owl:subClassOf
동일 synset ID	owl:equivalentClass

워드넷의 Words 테이블내의 명사들은 OWL에서의 Class들로 각각 대체가 되고 각 심볼들은 OWL Vocabulary들과 매칭하고 대체되어 사용자가 특정 도메인에 대한 OWL 온톨로지의 큰 구조를 쉽게 구축할 수 있다. 이렇게 구축된 특정 도메인에 대한 프레임 온톨로지는 본 연구에서 구축된 온톨로지 저작 도구를 통하여 개념들을 추가하고 수정하는 작업을 처리할 수 있다.

이렇게 워드넷에 기반하여 사용자가 원하는 도메인에 대한 온톨로지 프레임을 1단계로 구성하고, 다음 단계로 도메인 전문가에 의해 제공되는 개념들의 집합이나 그들 사이의 관계에 대한 특정 형식의 입력을 받아서 1단계로 구축된 프레임 온톨로지에 추가함으로써 좀더 상세한 온톨로지를 자동으로 구축할 수 있다. 여기에서는 가장 기본적인 예로 사용자가 컴퓨터에 관한 도메인 온톨로지를 구축하고자 할때, 사용자는 자신이 온톨로지 구축시 필요한 개념들과 개념들 사이의 관계에 대하여 일정 포맷으로 사전에 준비하여야 한다. 그리고 사용자는 본 시스템에 접근하여, 먼저 워드넷에 기반한 컴퓨터 도메인에 대한 프레임 온톨로지를 구축한후, 사용자가 준비한 특정 도메인의 개념들을 입력형식으로 받아서, 프레임 온톨로지에 자동으로 추가할 수 있다. 끝으로 이렇게 자동으로 구축된 온톨로지에 대해서는 사용자가 온톨로지 저작 도구에 접근하여 쉽게 내용을 수정할 수 있으며, 최종 온톨로지 파일은 OWL 언어를 사용하여 사용자에게 제공된다. 전체 과정을 살펴보면 다음 [표 2]와 같다.

표 2. 자동 온톨로지 구축 과정

1. 온톨로지 저작 도구에 접속
2. 사용자가 원하는 개념 선택(워드넷 기반)
3. 워드넷 기반 도메인 프레임 온톨로지 구축
4. 사용자 특정형식의 입력(도메인 전문가)
5. 자동 온톨로지 구축(프레임 온톨로지 기반)
6. 사용자 수정, 추가, 삭제(온톨로지 저작 도구 사용)
7. OWL 온톨로지 생성

4. 본 시스템 구성 및 활용

3장에서 소개한 방법론에 대하여 본 연구에서는 실제 시스템 설계를 구현하였으며, 이러한 자동 입력 방법론을 구현한 후, 이를 온톨로지 저작 도구의 기능으로 추가하여, 전체적인 시스템을 재구성하였다. 전체 시스템의 흐름을 살펴보면 다음 [그림 3]과 같다.

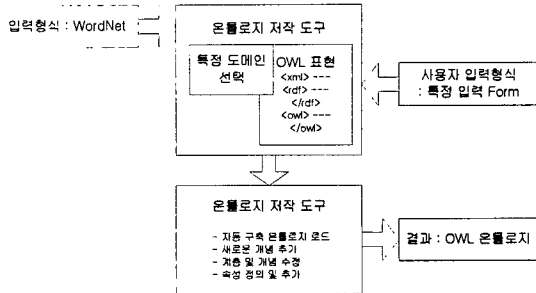


그림 3. 전체 시스템 구성도

[그림 3]에서와 같이 전체 시스템은 워드넷을 기반으로 사용자가 특정 도메인을 선택하여 프레임 온톨로지를 구축하고, 도메인 전문가에 의해 준비된 특정 형식을 입력으로 받아서 특정 도메인에 대한 온톨로지를 자동으로 구축할 수 있다. 그런다음, 온톨로지 저작도구를 통해서 전체 온톨로지를 수정하고 OWL 온톨로지를 사용자에게 제공한다. 전체적인 시스템 화면은 다음 [그림 4]와 같다.

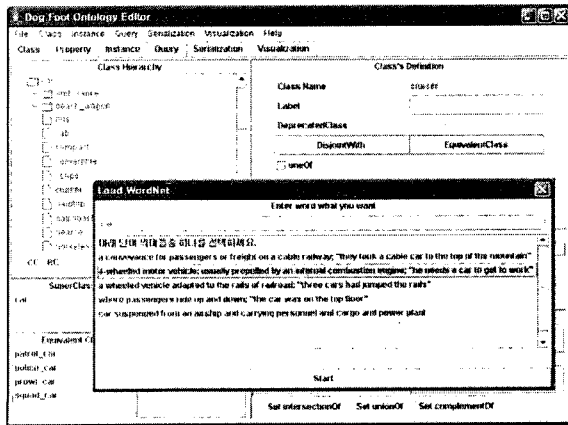


그림 4. 워드넷 자동입력 인터페이스

이러한 자동 구축의 방법과 이를 적용한 시스템을 사용하여 사용자가 온톨로지를 구축하고자 할때, 쉽게 개념의 수집과 관계를 적용한 도메인 계층을 바탕으로 좀더 빠른 시간안에 정확하고 표준적인 온톨로지 구축을 기대할 수 있다.

5. 결론 및 향후 연구 방향

시맨틱 웹에 대한 활발한 연구는 온톨로지에 대한 많은 발전을 가져온데 반해, 많은 해결해야할 문제점을 야기시켰다. 특히 온톨로지 구축의 어려움과 시간적 투자, 표준적 형식의 부재에 대한 내용으로 수많은 이질의 도메인 온톨로지가 생겨나게 되었다. 이에 본 연구에서는 이러한 문제들을 해결하기 위해 워드넷을 기반으로 한 기존의 지식 베이스를 재사용하는 자동 온톨로지 구축을 위한 방법론과 이를 적용한 온톨로지 자동 구축 시스템을 구현하였다. 그리하여 사용자가 좀더 쉽게 온톨로지를 구축할 수 있기를 기대한다. 그렇지만 본 시스템에서 일반적인 지식의 계층은 제공되지만 더욱 풍부한 개념의 정의는 사용자의 몫으로 남아있다. 이러한 부분에 대해서는 더 다양한 지식 베이스를 이용하여 해야 하며, 완벽한 자동화 시스템의 구현은 향후 연구 과제로 남아 있다.

참고문헌

- [1] <http://wordnet.princeton.edu/>
- [2] Michael Denny, "Ontology editor survey results", http://xml.com/2002/11/06/Ontology_Editor_Survey.html, 2002. 06
- [3] Natalya Fridman Noy, Deborah L. McGuinness. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05. March 2001
- [4] Berners-Lee, T., Hendler, J., Lassila, O., "The Semantic Web", Scientific American, 2001
- [5] Maedche A., Ontology Learning for the Semantic Web, Academic Publishers, 2002
- [6] Latifur Kahn, Feng Luo, "Ontology Construction for Information Selection", Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence, pp. 122-127, 2002
- [7] 이재호, "시맨틱 웹의 온톨로지 언어", 정보과학회지, 제21권, 제3호, pp. 18-27, 2003
- [8] 양정진, "시맨틱 웹에서의 온톨로지 공학", 정보과학회지, 제21권, 제3호, pp.28-35, 2003
- [9] 정도현, "시맨틱 웹을 위한 온톨로지 언어와 구현사례 연구", 정보관리연구, 제34권, 제3호, pp. 87-109, 2003