

온톨로지 자동 구축을 위한 서술어 온톨로지

민영근, 이복주
단국대학교 컴퓨터공학과

e-mail : minyk@dankook.ac.kr, blee@dku.edu

Predicate Ontology for Automatic Ontology Building

Youngkun Min, Bogju Lee
Dept. of Computer Engineering, Dankook University

요 약

시맨틱 웹의 기반인 온톨로지는 검색, 추론, 지식표현 등 다양한 분야에서 사용하고 있다. 하지만 잘 구성된 온톨로지를 개발하는 것은 시간적, 물질적으로 많은 자원이 소모된다. 온톨로지를 자동으로 구축하면 이러한 소모를 줄일 수 있는 장점이 있다. 본 논문에서는 자연어처리를 온톨로지 자동 구축에 사용하기 위하여 자연어의 서술부분을 온톨로지의 서술어로 변환할 수 있는 서술어 온톨로지를 제안한다. 그리고 제안된 서술어 온톨로지를 사용하여 자연어 문장의 서술어 부분을 온톨로지의 predicate 로 변환하는 알고리즘을 소개한다. 또한 제안된 온톨로지를 온톨로지 언어인 OWL 을 사용하여 구축하였다.

1. 서론

시맨틱 웹[2]은 이제 검색, 추론, 지식표현분야에서 광범위하게 사용되고 있다. 시맨틱 웹의 기반인 온톨로지는 어떤 특정 도메인에서 사용되는 개념들과 개념들간의 관계를 정의해 놓은 것으로서, 이러한 온톨로지는 특정 도메인의 지식을 보유한 도메인 전문가와 온톨로지 엔지니어가 협업하여 개발하고 있다. 하지만 이러한 개발을 위하여 소모하는 시간적 물질적 자원은 점점 증가하고 있다. 이러한 소모를 줄이기 위하여 온톨로지 자동 구축은 여러 연구에 의하여 시도되어 왔으며 많은 성과를 거두었다. 하지만 대부분의 구축방법은 반 자동구축에 그치고 있으며 온톨로지를 구축하기 위하여 선 처리 작업이나 미리 구축되어 있는 정형적인 자료들을 필요로 한다[2, 3, 5, 7, 8, 11]. 이에 비하여 자연어 처리를 이용한 방법[12]은 일반적으로 사용하는 자연어 문장으로부터 온톨로지를 구축할 수 있으나 구축된 온톨로지의 구조와 온톨로지 구축에 사용된 키워드들이 일정하지 않는다는 단점이 있다.

본 논문은 자연어 처리를 도입하여 자연어를 온톨로지 로 변환하는데 사용하기 위한 서술어 온톨로지를 제안한다. 서술어 온톨로지를 사용하여 자연어 처리된 형태소들에서 서술어를 결정하고 결정된 서술어를 적당한 온톨로지 predicate 로 변환 할 수 있다.

2. 관련연구

2.1 온톨로지

온톨로지는 “공유화된 개념화의 형식적 명시적 기

술(a formal explicit specification of a shared conceptualization)[1]”을 일컫는 말로써 하나의 영역에 존재하는 사물들의 개념과 그 개념들 사이의 관계를 체계적으로 정의한 것이다. 정의한 범위에 따라서 온톨로지는 개념모델(Conceptual Model), 분류(Classification), 데이터베이스 스키마(Database schemas) 등의 형태가 된다. 체계적으로 기술하기 위하여 온톨로지는 하나의 지식을 정의하는데 필요한 세 부분으로 되어 있다. 이러한 형태를 트리플 구조(RDF triple[14])라고 하며 이 구조는 사람의 생각을 기계적으로 해석 가능하도록 서술 할 수 있게 한다.

세 부분 중 첫 번째는 “subject”로 사물의 이름에 해당하는 역할을 한다. 지식을 기술할 때 무한한 영역을 유한하게 만들자면 경계선이 필요하고 온톨로지에서는 “subject”가 그 역할을 한다. 두 번째 부분은 “predicate”로서 기술하고자 하는 분야를 명시하게 된다. 마지막 부분은 “object”로서 “predicate”가 명시하는 분야의 값을 나타낸다. 이러한 트리플 구조를 자연어 문장과 비교하면 “subject”는 문장의 주어부, “object”는 목적어부, “predicate”는 서술어부에 해당한다.

2.2 온톨로지 자동구축

온톨로지의 자동 구축에는 다양한 방법들이 시도되고 있다. 기존의 시소러스나 사전 등과 같이 기 구축되어 있는 자료를 확보함으로써 추가의 사전작업 없이 바로 활용할 수 있는 지식베이스를 구축하는 방법 [2]이다. 다른 방법으로 기존의 자료를 활용하지 않고 문서의 분석 결과로 얻어지는 단어들의 분포를 이

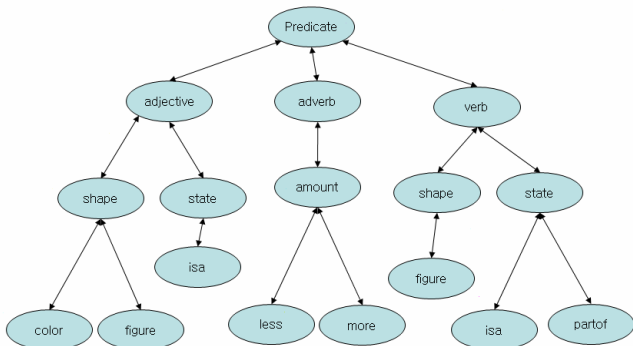
용하여 온톨로지를 구축하고 확장하는 방법[7]으로 개념의 확장이 용이한 장점이 있다. 또 다른 방법으로는 문서에 데이터마이닝 기법을 도입하여 문서로부터 개념을 추출하여 온톨로지를 구축하는 방법[3, 5, 8]이 있다. 또한 문서에 자연어 처리 방법을 적용하여 온톨로지를 구축하는 방법이 있으며 자연어 처리 중에서도 접미사 패턴을 이용한 방법[11]과 대상문을 파싱하여 단어 태그, 품사, 파싱 구조등을 사용하여 온톨로지를 구축하는 방법[6]이 있다.

2.3 자연어 처리를 이용한 온톨로지 자동 구축

자연어 처리를 이용한 온톨로지 자동구축방법은 다음과 같다. 먼저 자연어 문장을 형태소 분석기를 사용하여 문장을 형태소 단위로 구분하고 형태소별로 품사정보를 기록한다. 두 번째로 품사정보를 분석하여 각각의 형태소가 온톨로지 트리플 구조의 어떠한 부분에 적합한지를 결정한다. 세 번째로 서술어 부분으로 결정한 형태소를 적당한 형태의 predicate 로 변환한다. 마지막으로 변환된 결과들을 취합하여 온톨로지를 구축한다. 자연어를 이용한 방법은 좀 더 다양한 부분의 문장들을 온톨로지로 변경할 수 있지만 구축된 온톨로지가 다른 에이전트나 시스템에서 사용하기 적합한 형태의 구조와 predicate 로 구성되도록 하여야 한다.

3. 서술어 온톨로지

서술어 온톨로지는 형태소 분석을 통하여 나온 형태소를 중 온톨로지의 predicate 에 사용될 형태소가 결정되면 그 형태소를 적합한 형태의 predicate 로 변환하기 위하여 사용된다. 이러한 목적에 따라 서술어 온톨로지는 자연어 문장에서 나오는 형태소 중 서술어부(형용사, 부사, 동사)를 포함할 수 있도록 설계되었다. 그리고 적절한 predicate 별로 구분하기 위하여 subject 의 외적 상태를 서술하는 shape 와 subject 의 상태를 서술하는 state 로 세분하였다. 다음의 (그림 1)은 설계된 서술어 온톨로지이다.



(그림 1) 서술어 온톨로지의 클래스 구조

또한 온톨로지의 속성으로는 유일한 구분자인 ID 와 형태소의 원형인 primitive, 그리고 변환할 predicate 로

구성하였다. <표 1> 는 서술어 온톨로지의 인스턴스 예이다.

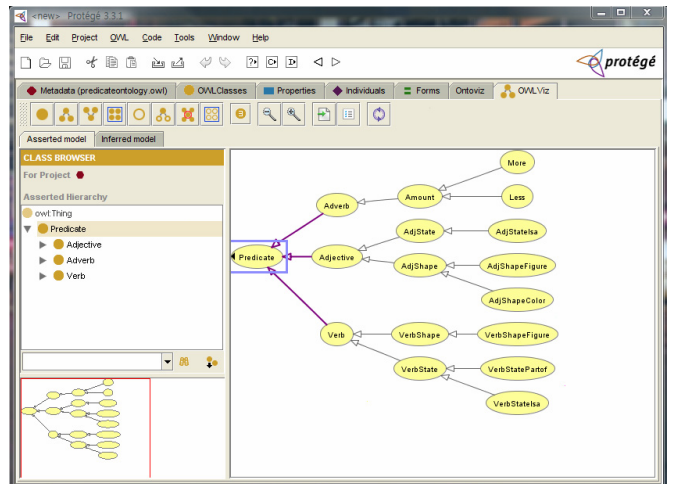
<표 1> Predicate 온톨로지 인스턴스의 예

Property			
Name	Primitive	Part	Predicate
Blue	파라	Adj	Color
Be	이	Verb	Isa
Include	포함	Noun	Partof

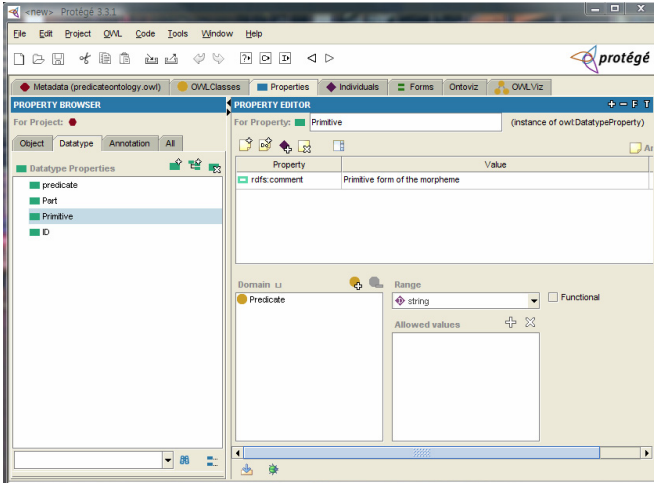
서술어 온톨로지를 사용하여 형태소가 predicate 로 변환되는 과정은 다음과 같다. 첫 째, 서술어부로 결정된 형태소의 원형과 형태소의 품사로 질의를 구성하여 온톨로지를 검색한다. 둘째, 검색결과가 존재하면 검색결과와 predicate 속성을 구축된 온톨로지의 predicate 로 사용한다. 셋 째, 검색결과가 존재하지 않는다면 형태소의 품사정보를 사용하여 각 품사별 대표 predicate 를 사용한다.

4. 구현

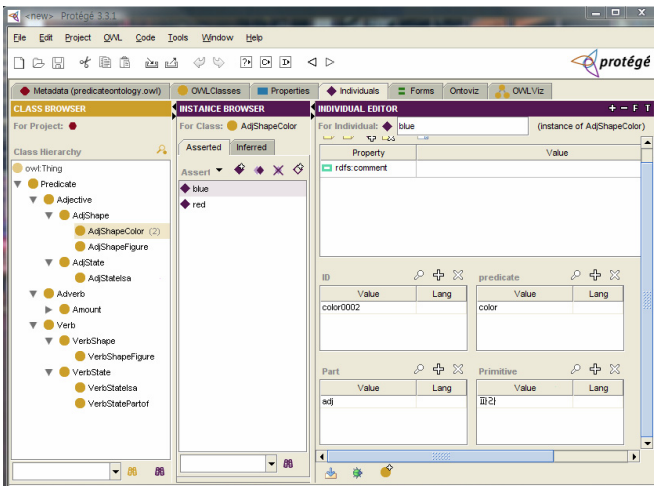
본 논문에서 제안한 서술어 온톨로지를 구현하기 위하여 온톨로지 언어인 OWL(Web Ontology Language)[14]를 사용하였으며 세부적으로 OWL 의 서브 셋인 OWL Lite 를 사용하였다. 다음의 (그림 2), (그림 3), (그림 4)는 온톨로지 구축도구인 Protégé 3.3[15]을 사용하여 서술어 온톨로지를 구축한 그림이다.



(그림 2) Protege 로 구축한 서술어 온톨로지 구조



(그림 3) Protege 로 구축한 서술어 온톨로지의 속성



(그림 4) Protege 로 구축한 서술어온톨로지 인스턴스

<표 2>, <표 3>은 구축된 온톨로지를 OWL 로 표현한 것이다.

<표 2> 서술어 온톨로지의 클래스 선언부분

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns="http://ai.dankook.ac.kr/ontology/ontobuild/predicateontology.o
wl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://ai.dankook.ac.kr/ontology/ontobuild/predicateontology
.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="AdjShapeFigure">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="AdjShape">
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Adjective">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Predicate">
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="AdjShapeColor">
    <rdfs:subClassOf>
```

```
<owl:Class rdf:about="#AdjShape"/>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="VerbShape">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Verb"/>
  </rdfs:subClassOf>
</owl:Class>
```

표 3 서술어 온톨로지의 인스턴스 부분

```
<AdjShapeColor rdf:ID="blue">
  <Part rdf:datatype="http://www.w3.org/2001/XMLSchema#string">adj
</Part>
  <predicate rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >color</predicate>
  <Primitive rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >파라</Primitive>
  <ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >color002</ID>
</AdjShapeColor>
<AdjShapeColor rdf:ID="red">
  <Primitive rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >빨</Primitive>
  <Part rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >adj</Part>
  <predicate rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >color</predicate>
  <ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >color001</ID>
</AdjShapeColor>
```

5. 결론 및 향후과제

본 논문에서는 자연어 처리를 이용한 온톨로지 자동구축 방법을 위하여 자연어 문장의 서술어부를 온톨로지의 predicate 로 변환하기 위한 서술어 온톨로지를 제안하였다. 제안된 온톨로지를 사용하면 자연어 문장의 서술어부를 효과적으로 온톨로지의 predicate 로 변환 할 수 있을 것이다. 하지만 변환하고자하는 서술어가 서술어 온톨로지에 존재하지 않을 때에는 품사에 기초한 대표 predicate 가 선택되기 때문에 자동구축된 온톨로지의 품질이 저하될 수 있을 것이다. 이러한 경우를 위하여 대표 predicate 를 좀 더 자세히 구분하고 서술어의 품사정보만이 아니라 다른 정보들, 즉 문장 내에서 서술어의 위치, 서술어가 서술하고 있는 형태소의 품사 등의 다른 정보까지 포함하여 서술어 온톨로지에 존재하지 않는 서술어를 predicate 로 변환하기 위한 연구가 진행되어야 할 것이다.

참고문헌

[1] T.R. Gruber, "A Translation Approach to Portable Ontologies", Knowledge Acquisition, vol. 5, no. 2, pp. 199 ~ 200, 1993
 [2] Kang, S.J. and Lee, J.H., "Semi-Automatic Practical Ontology Construction by Using a Thesaurus", Computational Dictionaries, and Large Corpora, ACL2001, Workshop on Human Language Technology and Knowledge Management, Toulouse, France, 2001
 [3] Patrick Clerkin, Pdraig Cunningham, Conor Hayes, "Ontology Discovery for the Semantic Web Using Hierarchical Clustering", Trinity College Dublin Computer Science Dept., Technical Reports, April 2001

- [4] T. Berners-Lee, J. Handler and O. Lassila, "The Semantic Web", Scientific American, May, 2001
- [5] Armin Wrobel, Oliver Wurmli, "Data Mining for Ontology Building", Diploma Thesis-Dept. of Computer Science WS 2002/2003
- [6] Jose Saias and Paulo Quaresma, "Using NLP techniques to create legal ontologies in logic programming based Web information retrieval system", In Proceedings of the ICAIL-International Conference on Artificial Intelligence and Law, 2003
- [7] Lim, S.Y., Koo, S.O., Song, M.H., Lee, S.J., "Hub word based on Ontology Construction for Document Retrieval", IC-AI'03, Las Vegas, USA, 2003
- [8] Mario Cannataro, Carmela Comito, "A Data Mining Ontology for Grid Programming", 1st Workshop on Semantic in Peer-to-Peer and Grid Computing at the Twelfth International World Wide Web Conference, 2003
- [9] Shelley Powers, "Practical RDF", O'REILLY, 2003
- [10] 김혜정 강보영, 황선옥, 이상조, "질의응답 시스템에서 의미 연관성 참조를 위한 온톨로지의 자동구축", 한국정보과학회 2003년 추계학술대회, Vol.30 No.2-1 pp. 109 ~ 111, 2003. 10
- [11] 임수연, 구상옥, 송무희, 이상조, "접미사 패턴을 이용한 온톨로지의 구축방안", 한국정보과학회 2003년 추계학술대회, Vol.30 No.2-1 pp.547 ~ 549, 2003, 10
- [12] 민영근, 이복주, "자연어 처리를 이용한 온톨로지 자동구축방법 제안", 한국멀티미디어학회 추계학술대회, Vol.10 No.2 pp.136, 2007. 11
- [13] 세종계획, <http://www.sejong.or.kr>
- [14] OWL Web Ontology Language Overview, <http://www.w3.org/TR/owl-features>
- [15] The Protege Ontology Editor and Knowledge Acquisition System, <http://protege.stanford.edu>