

한국어 품사 기반 온톨로지 구축 방법 및 차량 서비스 적용 방안*

차 시 호** · 류 민 우***

Constructing Ontology based on Korean Parts of Speech and Applying to Vehicle Services

Si-Ho Cha · Minwoo Ryu

〈Abstract〉

Knowledge graph is a technology that improves search results by using semantic information based on various resources. Therefore, due to these advantages, the knowledge graph is being defined as one of the core research technologies to provide AI-based services recently. However, in the case of the knowledge graph, since the form of knowledge collected from various service domains is defined as plain text, it is very important to be able to analyze the text and understand its meaning. Recently, various lexical dictionaries have been proposed together with the knowledge graph, but since most lexical dictionaries are defined in a language other than Korean, there is a problem in that the corresponding language dictionary cannot be used when providing a Korean knowledge service. To solve this problem, this paper proposes an ontology based on the parts of speech of Korean. The proposed ontology uses 9 parts of speech in Korean to enable the interpretation of words and their semantic meaning through a semantic connection between word class and word class. We also studied various scenarios to apply the proposed ontology to vehicle services.

Key Words : Knowledge Graph, Lexical Dictionary, Ontology, Semantics, Vehicle Service

I. 서론

지식 그래프(Knowledge Graph)는 다양한 리소스를 기반으로 시맨틱 정보를 사용하여 검색 결과를 향상시키는 기술로서, 최근 AI 기반의 서비스를 제공하기 위한 핵심 연구 기술로 대두되고 있다[1]. 이러

한 지식 그래프들은 웹 및 다양한 도메인으로부터 평문 형식의 지식을 수집하고, 지식과 지식들을 온톨로지를 사용하여 의미적 연결을 통해 다양한 형태의 지식들을 생성 및 연결 할 수 있다. 따라서, 이러한 지식 그래프의 장점들로 인해 지식 그래프는 검색 기반의 다양한 서비스에 이용되거나 음성을 이용한 Q&A 시스템의 기반 기술로 활용되고 있다[2, 3].

최근에는 다양한 음성 스피커 기반의 서비스가 확대되면서 택내뿐만 아니라 가상 상담원과 차량내 음

* 이 논문은 2020년도 청운대학교 연구년 지원에 의한 것임

** 청운대학교 멀티미디어학과 교수 (제1저자)

*** 현대자동차 자동차부분 연구개발본부 책임연구원 (교신저자)

성 기반 검색 서비스 등 다양한 도메인에서 활용되고 있다. 그러나 한글이 갖는 단어의 포괄적인 의미와 문장 내에 위치한 품사의 형태에 따라 단어의 의미가 달라지는 특징이 있기 때문에, 한국어 기반의 지식 그래프를 구축할 경우에는 이를 해석하기 위한 보조 도구의 필요성이 높아지고 있다. 뿐만 아니라 한글의 경우 영어로 표현된 단어와의 사용을 함께 하기 때문에, 문장의 구성 성분에 따른 의미 해석은 매우 중요하다. 이를 위하여, 본 논문에서는 온톨로지를 이용하여 문장의 구성 요소를 한국어의 9품사를 이용하여 품사와 품사간의 관계를 정의할 수 있는 온톨로지 구축 방법에 대하여 제안한다. 제안하는 품사 기반 온톨로지 구축 방법은 한국어 문장이 구성될 수 있는 9개의 품사를 정의하고, 각 품사와 품사간의 연결 구조를 온톨로지로 정의한다. 또한, 차량 서비스 내 질의 응답 서비스를 제공하기 위하여 본 논문에서 제안하는 품사 기반의 온톨로지가 사용될 수 있는 서비스 시나리오를 기술한다.

본 논문의 구성은 다음과 같다. 2장에서는 온톨로지 기반의 질의 응답 시스템 및 언어 사전과 관련된 기존의 연구 결과들에 대하여 기술하고, 3장에서는 본 논문에서 제안하는 품사 기반의 온톨로지 구축 방법에 대하여 기술한다. 또한, 본 논문에서 제안하는 품사 기반의 온톨로지가 차량 서비스에 적용될 수 있는 시나리오에 대하여 기술한다. 마지막으로, 4장에서는 결론 및 향후 과제에 대하여 기술한다.

II. 관련 연구

본 절에서는 본 연구와 관련된 연구로 시맨틱 기술을 이용하여 구축된 사전과 관련된 기존의 연구에 대하여 살펴본다. 시맨틱 기술은 다양한 리소스를 온톨로지 형태로 표현하고, 리소스와 리소스를 의미적 관계 정보로 표현하여 사람이 이해하는 형태의 의미를

기계가 처리할 수 있는 하나의 프레임워크이다[4]. 시맨틱 기술은 RDF(Resource Description Format) 및 OWL(Web Ontology Language)의 형태로 표현되며 Subject, Object, Predicate 형태로 3가지의 데이터를 하나의 데이터 형태로 표현한다.

SKOS(Simple Knowledge Organization System)은 W3C에서 개발한 시맨틱 웹 기반의 지식 구조화 프레임워크 표준으로서, 지식 어휘 체계를 정의하기 위하여 사용된다[5]. SKOS는 개념모델을 정의하여, 기존 ISO 2788 시소러스 표준[6]을 지원한다. 따라서, SKOS는 시소러스, 택소노미, 분류체계 등과 같은 지식 어휘 체계를 표현하기 위한 RDF를 지원한다.

Hugo는 포르투갈어를 위한 사전 기반의 어휘 온톨로지인 PAPEL(Palavras Associadas Porto Editora Linguatca)를 제안하였다[7]. PAPEL은 반자동화적으로 단어와 단어간의 관계를 추출할 수 있으며, 이를 통하여 포르투갈어의 자연어 처리를 위한 어휘 리소스를 제공한다. PAPEL의 경우, 반자동화적으로 단어와 단어 관계를 추출하는 장점은 가졌지만, 포르투갈어를 위한 어휘만을 지원하는 단점이 있다.

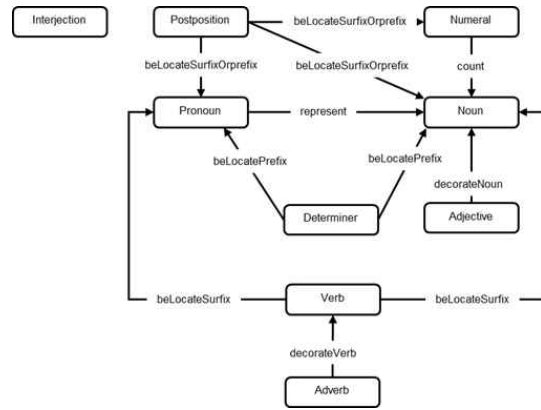
류기동은 N은행 콜센터 사례를 중심으로 AI 기반 콜센터 실시간 상담 도우미 시스템을 개발하는 방법을 제안하였다[8]. 이 방법은 AI 기술과 PBX, CTI 등의 콜센터 정보 시스템을 결합하여 실시간 상담 연구 센터에 대한 사례를 분석하였으며, 자연어 시스템과 온톨로지를 활용한 개체명 인식기를 통하여 질의 응답 시스템의 성능이 약 31% 증가하는 사례 결과를 분석 및 제시하였다.

WordNet은 널리 알려진 영어의 의미 어휘목록으로, 영어 단어의 유의어 사전을 제공한다[9]. 또한 WordNet에서는 유의어 집단으로 정의된 어휘 목록간 다양한 의미 관계를 온톨로지를 이용하여 구축하였다. WordNet은 널리 알려진 어휘 목록이지만, PAPEL과 같이 한국어이 아닌 다른 언어만을 지원하기 때문에, 한국어 기반의 지식 평문을 지식 그래프

에 입력할 경우 정확한 처리가 어렵다는 단점이 있다.

따라서, 한국어 기반의 지식 평문을 지식 그래프에 적용하기 위해서는 한국어의 특수성 및 구조를 식별하고, 이를 위한 언어 사전을 적용해야 한다. 본 논문에서는 한국어의 9품사와 품사간의 수식 관계 등을 고려하여, 한국어 지식 평문에 맞는 온톨로지 기반의 한국어 사전 구축 방법을 제안하고, 이를 차량 서비스에 적용할 수 있는 시나리오를 기술한다.

시맨틱 연결을 정의하였다.



<그림 1> 온톨로지 기반 한국어 구품사간의 관계 정의

III. 품사기반 온톨로지 구축 방법

3.1 한국어 품사 기반 온톨로지 구조 설계

본 절에서는 본 논문에서 제안하는 품사 기반의 온톨로지 구축 방법을 기술하기 위하여 한국어 품사 기반의 온톨로지 구조를 설계 한다. 이를 위하여, 한국어의 9품사를 기준으로 각각의 클래스를 정의하고, 품사간의 수식 관계를 의미적 관계로 연결하여, 한국어 지식 평문이 입력되었을 때, 각 단어의 품사를 기반으로 다양한 문장을 형성할 수 있도록 설계 하였다.

<그림 1>은 한국어 9품사간의 관계를 온톨로지로 표현한 것이다. <그림 1>에서 각각의 클래스는 한국어의 9품사를 나타낸다. 따라서 Noun 클래스와 Pronoun 클래스는 한국어의 명사와 대명사를 표현하는 클래스로, represent 오브젝트 프로퍼티를 통하여 시맨틱적 의미 관계를 가진다. Determiner 클래스는 한국어의 관형사를 표현하는 클래스로 명사 또는 대명사의 앞에 위치할 수 있도록 beLocatePrefix 오브젝트 프로퍼티를 이용하여 Pronoun 클래스와 Noun 클래스와의 시맨틱 연결을 갖는다. Adjective 클래스는 한국어의 형용사를 표현한 것으로 명사를 수식할 수 있도록 decorateNoun 오브젝트 프로퍼티를 통하여

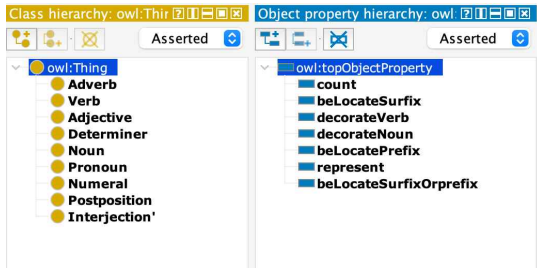
Verb 클래스는 한국어의 동사를 표현한 클래스로서, Pronoun 클래스와 Noun 클래스와의 위치 관계를 나타내는 beLocateSurfix 오브젝트 프로퍼티를 통하여 연결하였다. Postposition 클래스는 한국어의 조사를 표현한 클래스로서, 명사를 표현하는 Noun 클래스와 대명사를 표현하는 Pronoun 클래스의 앞 또는 뒤에 위치할 수 있도록 beLocateSurfixOrprefix 오브젝트 프로퍼티를 이용하여 시맨틱적 연결을 정의 하였다. 또한, 명사 또는 대명사의 앞과 뒤에서 사용될 수 있는 수사의 경우, Numeral 클래스로 표현할 수 있도록 정의하였으며, count 오브젝트 프로퍼티를 통하여 시맨틱 관계 정의를 하였다. 이와 함께, 동사를 수기하는 용도로 사용되는 부사의 경우, Adverb 클래스에서 표현할 수 있도록 정의하였으며, decorateVerb 오브젝트 프로퍼티를 통하여 Adverb 클래스와 Verb 클래스간 관계를 정의하였다.

마지막으로, 한국어의 감탄사를 표현하기 위한 Interjection 클래스는 감탄사의 특징대로 독립적으로 사용이 가능하기 때문에, 다른 클래스와의 시맨틱 관계를 가지지 않도록 독립적으로 클래스만 선언하였다.

3.2 Protege를 활용한 온톨로지 구축

본 절에서는 3.1절에서 설계한 한국어 품사 기반 온톨로지 구조를 기반으로 프로지티(Protege)[10] 툴을 활용하여 구현한 온톨로지에 대하여 기술한다.

<그림 2>는 프로지티 툴을 사용하여 온톨로지 기반 한국어 사전의 클래스와 오브젝트 프로퍼티를 정의한 것이다. <그림 2>에서 왼쪽 부분의 노란색 동그라미로 표기된 것은 클래스를 나타낸 것이며, 오른쪽의 파란색 네모는 오브젝트 프로퍼티를 나타낸 것이다. 품사 기반 온톨로지 구축을 위하여 각각의 클래스는 서로 포함구조가 되지 않도록 구성하였으며, 각각의 클래스는 오브젝트 프로퍼티를 통하여 연결이 가능할 수 있는 구조로 개발하였다.



<그림 2> Protege를 활용한 클래스 및 오브젝트 프로퍼티 정의

<그림 3>은 프로지티 툴을 통하여 정의된 클래스와 오브젝트 프로퍼티를 통하여 OWL(Web Ontology Language)을 구현한 결과의 일부를 보인 것이다.

3.3 차량 서비스 시나리오

본 절에서는 본 논문에서 제안하는 품사 기반의 온톨로지가 차량 서비스에 적용될 수 있는 시나리오에 대하여 논의한다. 최근 자동차에 지능형 서비스가 적용되면서 자율 주행 기술이 적용된 차량이 증가되고 있으며, 이를 기반으로 다양한 인프라를 활용한 커넥티드카(connected car) 서비스의 개발이 활발하게 이루어지고 있다. 이러한 커넥티드카 서비스는 차량이 스스로 인지한 데이터를 커넥티드 서비스 센터에서 해석하고, 이를 통하여 차량을 안전하게 주행시키거나 다양한 부가가치 서비스를 제공할 수 있다. 따라서, 본 논문에서 제안하는 품사기반의 온톨로지를 활용하여 차량 내 탑재된 카메라를 이용하여 도로 내 이정표 및 전광판에 나타나는 문장들을 해석하여 예약, 주문, 서비스 다운로드 등 다양한 분야에서 적용될 수 있다. 이러한 서비스를 제공하기 위해 가장 중요한 기능은 한국어 문장을 식별할 수 있는 기술들이 요구되며 문장의 형태소를 분석하여 각각의 단어들

```

<!-- http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#beLocatePrefix -->
<owl:ObjectProperty rdf:about="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#beLocatePrefix">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#Determiner"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#Noun"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#Pronoun"/>
</owl:ObjectProperty>

<!-- http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#beLocateSurfix -->
<owl:ObjectProperty rdf:about="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#beLocateSurfix">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#Verb"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#Noun"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/m/ontologies/2021/10/untitled-ontology-44#Pronoun"/>
</owl:ObjectProperty>
    
```

<그림 3> Protege로 구현한 OWL 결과

의 품사를 구분해야하는 것이 필요하다. 또한, 품사 기반 온톨로지 시스템은 차량 내 탑승한 운전자의 언어적 해석에도 적용될 수 있다. 최근 음성 스피커를 활용한 비서 서비스 등과 유사하게, 차량 내 탑승한 운전자의 음성을 수집하고, 이에 대한 음성 문장을 품사 기반의 온톨로지를 활용하여 보다 지능적인 서비스를 제공할 수 있다.

V. 결론

본 논문에서는 한국어의 품사를 이용하여 온톨로지를 구축하는 방법에 대하여 제안하였다. 제안하는 온톨로지 기반 한국어 사전은 한글이 가지는 특수성과 다양성을 고려하여 한국어에서 정의한 9품사를 기준으로 품사와 품사간의 관계성을 정의하여, 수집된 문장에 대한 단어가 포함하는 품사의 특징을 기반으로 여러 단어가 결합 및 생성이 가능할 수 있도록 각각의 문장에 대한 문장 성분을 온톨로지 기술을 이용하여 정의하였다. 또한, 제안하는 방법론을 기반으로 차량 서비스에 적용할 수 있는 시나리오를 제시하였다. 그러나 본 논문에서 제안하는 온톨로지 기반의 한국어 사전의 경우, 지식 평문이 입력되었을 때 문장을 구성하는 단어의 형태소가 정의되지 않는다면 제안하는 온톨로지 기반의 한국어 사전을 구축할 수 없는 단점이 존재한다. 이를 해결하기 위하여 향후 과제로는 한국어 형태소 분석기와 결합하여, 형태소 분석의 결과값을 본 논문에서 제안하는 온톨로지 기반 한국어 사전과 결합하여 단어적 데이터를 저장할 수 있는 방법에 대한 연구가 요구된다.

참고문헌

- [1] D. Allemang, J. Hendler, F. Gandon, "Semantic Web for the Working Ontologist: Linked Data, RDFs, and OWL," 3rd Edition, ACM Books, Aug. 3, 2020.
- [2] 권준희, "다중 기기 환경에서 기기 특성과 기기 사용 특성을 활용한 검색 기법," 디지털산업정보학회논문지, 제17권, 3호, 9월 2021, pp.17-26.
- [3] 김은희, 서유화, "토픽 레이블링을 위한 토픽 키워드 산출 방법," 디지털산업정보학회논문지, 제16권, 3호, 9월 2020, pp.25-36.
- [4] T. Berners-Lee, J. Hendler, O. Lassila, "The semantic web," Scientific american, vol. 284, no. 5, 2001, pp.34-43.
- [5] A. Miles, B. Matthews, M. Wilson, D. Brickley, "SKOS core: simple knowledge organisation for the web," In International conference on dublin core and metadata applications, Sept. 2005, pp.3-10.
- [6] S. G. Dextre Clarke, M. L. Zeng, "From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling," Information Standards Quarterly (ISQ), vol. 24, no. 1, 2012.
- [7] H. G. Oliveira, D. Santos, D., Gomes, N. Seco, "PAPEL: a dictionary-based lexical ontology for Portuguese," In International Conference on Computational Processing of the Portuguese Language, Springer, Berlin, Heidelberg, Sept. 2008, pp.31-40.
- [8] 류기동, 박종필, 김영민, 이동훈, 김우제, "AI 기반 콜센터 실시간 상담 도우미 시스템 개발: N 은행 콜센터 사례를 중심으로," 한국산학기술학회논문지, 제20권, 제2호, 2019, pp.750-762.
- [9] G. A. Miller, "WordNet: a lexical database for English," Communications of the ACM, vol. 38, no. 11, 1995, pp.39-41.

[10] protege, Stanford University, Web Site:
<https://protege.stanford.edu/>.

■ 저자소개 ■



차 시 호
Cha, Si-Ho

2009년 3월~현재
청운대학교 멀티미디어학과 교수
2020년 3월~2021년 2월
Auckland University of
Technology 방문교수
2004년 2월 광운대학교 컴퓨터과학과(공학박사)
1997년 7월~2000년 2월
대우통신 종합연구소 선임연구원

관심분야 : 네트워크 관리, 차량통신 네트워크,
시맨틱웹, 머신러닝
E-mail : shcha@chungwoon.ac.kr



류 민 우
Ryu, Minwoo

2021년 7월~현재
현대자동차 자동차부분
연구개발본부 책임연구원
2017년 1월~2021년 6월
KT AI연구소 선임연구원
2012년 2월 광운대학교 컴퓨터과학과(공학박사)
2011년 2월~2016년 12월
전자부품연구원 선임연구원

관심분야 : 사물인터넷, 시맨틱스, 코그니티브
컴퓨팅, 지능 네트워크, 차량통신
E-mail : mw.ryu@kt.com

논문접수일 : 2021년 11월 30일
수 정 일 : 2021년 12월 13일
게재확정일 : 2021년 12월 16일