

웹 자료 수집을 활용한 언어 지식그래프 개발*

- 정서 장애 주제 분야의 사례 연구 -

조은경 · 강승식**

(서강대학교 조교수 · 국민대학교 교수)

<Abstract>

Jo Eun-Kyoung, Kang Seung-Sik. 2020. Development of Linguistic Knowledge Graph with Collecting Language materials in the Web - A case study of the subject field of Affective Disorder - *Korean Semantics*, 67. This paper aims to show how to develop a knowledge graph as linguistic knowledge-base for linguistic artificial intelligence applications. Firstly, we investigate the previous researches and developments of knowledge graphs which have been developed so far and are the extensions of lexicon and ontology which are well-known forms of linguistic knowledge bases. Secondly, we describe how to develop a knowledge graph in the domain of affective disorder as a case study based on expert knowledge data and general text data. In the process of doing this, we would like to show how to collect language resources from the Web which have popular influences on people and try to make formalisation of representation way of linguistic knowledge in the domain of psychology. Finally, we show the application of using a knowledge graph as a semantic knowledge-base. Overall, we would say that a linguistic knowledge graph still would be one of the knowledge powers which will be linguistically explainable by humans in the world of digital data analysis floods.

핵심어: 지식그래프(Knowledge Graph), 온톨로지(Ontology), 시맨틱 웹

* 이 논문은 2017년 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017M3C4A7068186)

** 제1저자: 조은경, 교신저자: 강승식

(Semantic Web), 연결데이터(Linked Data), 의미관계(Semantic Relation), 언어 자원(Language Resource), 웹 자료(Web data), 언어 지능(linguistic artificial intelligence), 의미 지식베이스(Semantic Knowledge-base)

1. 머리말

이 글은 언어의 의미 정보를 지식 베이스로 개발하는 방법론의 하나인 지식 그래프 개발의 사례 연구를 보이는 것을 목표로 한다. 최근 한국 사회의 발전 속도가 빨라지고 경쟁이 고도화되면서 정서적인 어려움(affective disorder)을 겪는 사람이 많아지고, 이러한 양상은 사회 현상에 관한 이슈를 알리는 여러 매체를 통해 알려지고 있다. 뉴스 기사나 정신 의학 전문가 집단의 뉴스나 기고 글, 그리고 검색 서비스로 제공되는 전문 지식과 용어집 등을 통해 그러한 이슈 정보는 지속적으로 양산되고 있다. 이 글은 정서적인 문제에 관한 지식 정의와 이에 대한 전파의 역할을 하는 국가건강정보포털(<https://health.cdc.go.kr/>)의 정신 의학용어, 정신의학신문(<http://www.psychiatricnews.net>)의 전문가 기사, 네이버 뉴스(<http://news.naver.com>)에서 정서 장애 용어로 검색되는 일반 뉴스를 언어 자원으로 하여, 정서 장애 관련 어휘의 의미를 표현하는 방법의 하나로서 지식 그래프를 소개하고 언어적 의미지식베이스로서 지식 그래프를 개발하는 방안을 제안하고자 한다.

지식 그래프(Knowledge Graph)¹⁾는 2012년 구글이 검색 서비스의 결과를 보다 연관성 있게 표현하기 위해 데이터 정보들의 연관 관계를 잘 갖추게 하는 방법론에 대한 포괄적 이름이다. 지식 그래프라는 용어 정의가 학술적으로 자리잡히거나 저명한 단행본류는 나와 있지 않지만, 전산언어학회(ACL)나 최근의 딥러닝 관련 연구에서 언어 자원 구축이나 지식베이스 개발과 관련된 용어로 꾸준히 사용되어 왔다. 지식 그래프는 언어 지식을 텍스트나 문자열이라는 정적인 표현보다 연관된 정보 개체의 연결(Linked data)로 표현함으로써 인터넷 사회에서 웹으로 통하는 동적인 정보를 제공한다는 데에 그 가치가 있다. 예를

1) https://en.wikipedia.org/wiki/Knowledge_Graph

들어, 2015년 구글 검색에서 토마스 제퍼슨(Thomas Jefferson)에 대한 검색 결과로 제시되었던 지식그래프의 내용은 현재(2020년 2월 10일) 제시되는 검색 결과(<그림 1>)에도 여전히 오른쪽 패널에 제시되고 있다.

Google Thomas Jefferson

About 212,000,000 results (0.98 seconds)

[en.wikipedia.org](#) > wiki > Thomas Jefferson

Thomas Jefferson - Wikipedia

Thomas Jefferson (April 13, 1743 – July 4, 1826) was an American statesman, diplomat, lawyer, architect, philosopher, and Founding Father who served as the third president of the United States from 1801 to 1809. He previously served as the second vice president of the United States from 1797 to 1801.

Political party: [Democratic-Republican](#) Education: [College of William and Mary \(BA\)](#)
 Resting place: [Monticello, Virginia, U.S](#) Spouse(s): [Martha Wayles](#); (m. 1772; died 1782)
[Thomas Jefferson and slavery](#) [Sally Hemings](#) [Monticello](#) [Early life and career](#)

People also ask

- What is Thomas Jefferson most famous for?
- How did Thomas Jefferson die?
- What did Thomas Jefferson do in the Revolutionary War?
- What made Thomas Jefferson a good president?

[Feedback](#)

[www.biography.com](#) > us-president > thomas-jefferson

Thomas Jefferson - Quotes, Facts & U.S. Presidency - Biography

Sep 5, 2019 - Thomas Jefferson was a Founding Father of the United States who wrote the Declaration of Independence. As U.S. president, he completed the ...
 Birth Date: April 13, 1743 Education: College of William and Mary
 Death Date: July 4, 1826

Thomas Jefferson 3rd U.S. President

Thomas Jefferson was an American statesman, diplomat, lawyer, architect, philosopher, and Founding Father who served as the third president of the United States from 1801 to 1809. He previously served as the second vice president of the United States from 1797 to 1801. [Wikipedia](#)

Born: April 13, 1743, Shadwell, Virginia, United States
Died: July 4, 1826, Monticello, Virginia, United States
Presidential term: March 4, 1801 – March 4, 1809
Children: [Martha Jefferson Randolph](#), [Madison Hemings](#), [MORE](#)
Vice presidents: [Aaron Burr](#) (1801–1805), [George Clinton](#) (1805–1809)

Books [View 45+ more](#)

- [Notes on the State of Virginia](#)
- [Autobiography](#)
- [Works](#)
- [Draft of the Declaration of Independence](#)

<그림 1>. 검색 결과 중 지식그래프의 활용 예시: 오른쪽 패널

지식그래프는 내용 면에서 풍부한 어휘 정보를 담아내려는 사전의 미시 구조적 특성을 가지면서, 형식 면에서 인터넷 시대의 정보 표현 방식에 맞게 하이퍼링크를 이용하여 의미적으로 연관된 정보를 제시하고 연결된 해당 정보로 이동할 수 있는 기술적 특성을 갖춘 것이라 할 수 있다.

2. 데이터 기반의 어휘 의미 표현 방법

2.1. 어휘부와 온톨로지

어휘부(lexicons)와 온톨로지(ontologies)²⁾는 유기적 관계를 갖는다. 어휘가 어

떤 개념을 나타내고 그 개념은 온톨로지를 구성하기 때문에 어휘부는 잘하면 온톨로지 대용이 될 수 있다.(Hirst 2010:209) 그러나 단어들 간의 의미 관계를 명확하게 규정짓기 어렵듯이 온톨로지 안에서의 개념과 개념 관계도 마찬가지로의 한계가 있다. 그 이유는 단어가 중의성을 띠거나 문맥에 따라 달라지는 다의성을 띠거나 하는 일뿐만 아니라 어떤 언어에는 있는 개념이 다른 언어에는 없는 것과 같은 어휘적 공백도 있기 때문이다. 이렇듯 어휘부만으로 온전한 의미 체계를 개발하는 일이 한계가 있긴 하지만, 어쨌든 우리가 어떤 개념을 언어적으로 표현하기 위해 어휘를 활용하기 때문에 실용적 수준의 온톨로지를 개발하는 데에 어휘부는 그 기초가 된다. 어휘부는 그 구성 내용에 따라 일반(general) 어휘부일 수도 있고, 특정 주제 분야(domain-specific) 어휘부일 수도 있다. 또 어휘부는 그 의미 표현 유형에 따라, 단순한 단어 목록의 형태, 대응 어휘 쌍의 형태, 어휘에 어떤 의미적 기준에 따른 점수를 가진 목록 형태, 어휘에 다양한 층위의 문법 정보를 가진 형태, 거기에 문맥 활용 정보를 가진 형태, 어휘부 안에서 어휘 간의 의미 관계가 명세된 형태 등 다양한 실체가 있을 수 있어서서 어떤 형태로 그 의미 체계를 표현하느냐에 따라 의미 혹은 개념 처리를 위한 온톨로지로서의 다양한 활용도를 가질 수 있다. 워드넷(Wordnet)은 유의어집합(synsets) 때문에 유의어 추론에 활용되는 온톨로지이고, UMLS(Unified Medical Language System)라 불리는 온톨로지는 의학 분야 시소러스를 가지고 의학 분야 전문가 시스템으로 활용되기도 했다.(Udo Hahn and Stefan Schulz 2010:133)

2.2. 알디에프, 아울, 시맨틱 웹

알디에프(RDF:Resource Description Framework)는 국제 표준을 관리해가는 월드와이드웹 컨소시엄(W3C:World Wide Web Consortium)에서 수립된 웹에서 데이터 교환을 위한 표준 모델의 하나이다.³⁾ 알디에프는 시맨틱 웹을 구현하는

2) 온톨로지(ontology)란 어떤 분야 안에서 공감되는 개념(shared concept)에 관한 외적, 형식적 명세이다. - [Handbook on Ontologies]의 서문 -

3) https://en.wikipedia.org/wiki/Resource_Description_Framework,
<https://www.w3.org/RDF/>

데이터 표현 및 교환 체계로서 그 어휘 기술 언어(RDF vocabulary description language)로 사물의 속성(property), 부류(class), 관계(relationships)를 기술할 수 있게 한다. 특히 Subject, Predicate, Object이⁴⁾ 연결되어 하나의 단위를 이루며 그 한 단위는 Triple 이라 불린다. 예를 들어 Eric Miller 라는 사람의 연락처를 구성하는 이름, 이메일주소, 직책의 의미 값 기술은 아래와 같이 세 항목으로 구분하고, <표 1>과 같은 방식으로 표현된다⁵⁾

subject: “http://www.w3.org/People/EM/contact#me”

objects:

“Eric Miller”,

mailto:e.miller123(at)example,

“Dr.”

predicates (for each object):

http://www.w3.org/2000/10/swap/pim/contact#fullName

http://www.w3.org/2000/10/swap/pim/contact#mailbox

http://www.w3.org/2000/10/swap/pim/contact#personalTitle

<표 1> RDF Triples

http://www.w3.org/People/EM/contact#me,	http://www.w3.org/2000/10/swap/pim/contact#fullName,	“Eric Miller”
http://www.w3.org/People/EM/contact#me,	http://www.w3.org/2000/10/swap/pim/contact#mailbox,	mailto:e.miller123 (at)example
http://www.w3.org/People/EM/contact#me,	http://www.w3.org/2000/10/swap/pim/contact#personalTitle,	“Dr.”

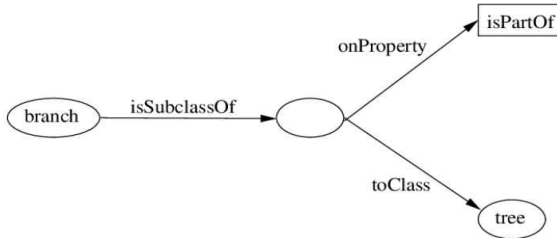
RDF는 XML 형식으로 변환이 되고 W3C에서 지원하는 웹 문서 간의 의미 관계를 표현하는 시맨틱 웹(Semantic Web)의 발전과 함께 지속적인 확장과 갱신이 있어왔다. 오늘날 강력한 인터넷 지식 서비스를 하고 있는 DBpedia,

4) RDF에서 Subject, Predicate, Object는 문법 성분 명칭이 아니라, 개념적 뜻의 이름이다.

5) 예시 출처: http://en.wikipedia.org/wiki/Resource_Description_Framework#rdfs

Wikidata, YAGO 등이 RDF 기술 형식을 쓰고 있다.

아울(OWL:Web Ontology Language)은 평면적, 기술적(descriptive) 표현을 하는 RDF의 한계를 개선하여 사물, 사물의 집합, 그것들의 관계에 관한 논리적 지식을 더 잘 표현하기 위해 W3C에서 만들어진 또 하나의 시맨틱 웹 언어이다. 특히 IsA, PartOf와 같은 의미 관계를, 예를 들어 <그림 2>에서 tree와 branch 간의 의미 관계를 노드와 노드의 연결로 표현하기 위해 <표 2>와 같은 형식 구조를 갖춘다.



<그림 2> Ontology: ‘나뭇가지와 나무’의 의미 관계 노드

<표 2> OWL의 개념 관계 표현 (Antoniou and van Harmelen 2010:84)

```
<owl:Class rdf:ID="tree">
  <rdfs:comment>Trees are a type of plants</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#plant"/>
</owl:Class>
<owl:Class rdf:ID="branch">
  <rdfs:comment>Branches are a type of trees</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#is-part-of"/>
      <owl:allValuesFrom rdf:resource="#tree"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

OWL은 세상 지식과 세상 사물, 개념을 잘 연결하여 표현하기 위한 것이지만 <표 2>에서 볼 수 있듯이 그 내용면에서는 언어의 의미 기술을 풍부하게 하는 갖춘 사전으로서의 어휘부와 다르지 않고 XML이나 RDF라는 형식적인 틀과 결합되어 있다는 도구적 장점을 갖추고 있는 것이다.

월드와이드웹(WWW)의 창시자 팀버너스리가 고안한 시맨틱 웹이란 인터넷으로 유통되는 데이터를 데이터의 웹(Web of data)으로 즉, 데이터 간의 관계를 기계가독형(machine-readable)의 의미있는 연결로 표현하는 데에 그 목적이 있었고, RDF와 OWL은 그러한 데이터 표현을 가능하게 하는 형식적인 기제로 지속적으로 활용되고 있다. 그리고 이러한 표현 형식들은 웹에서 데이터 간의 의미 있는 연결 관계를 표현하기 위한 데에 그 목적이 있다.

2.3. 지식그래프(Knowledge Graph)

언어학 혹은 의미론 연구에서 위와 같은 RDF, OWL은 사실, 형식일 뿐이다. 어떤 일의 내용과 형식을 구분하여 생각하면, 형식을 잘 표현해야 내용이 더 가치있게 쓰일 수 있는 점에서 형식이 중요하고 또 이때 더 많이 활용될 수 있는 형식을 활용하는 것이 유리한 점이 있다.

그런데 보다 최근에 의미적 지식베이스의 형식을 일컫는 명칭인 지식그래프는 그 표현 형식에 제한을 두지 않는다. 특정한 형식적 제한을 두기보다는 내용에 중점을 두고 개체와 개체간의 표현을 위한 연결 관계에 중점을 둔다. 그렇기에 지식그래프는 오히려 형식적 제약을 규명하지 않는 어휘부와 가깝다. 그렇지만, 어휘부가 문자열(string) 혹은 텍스트라는 표현 형태에 의존성이 큰 반면, 지식그래프는 문자열보다는 사물 자체의 개념과 그 연결에 중점을 둔다.(Singhal 2012) 즉 언어 형태보다는 의미에 중점을 둔다고 보아야 한다.

언어학의 어휘부가 문자열이라는 표현 형식 안에 갇히고 XML이라는 형식적 표현에 갇히게 되면, 언제나 형식적 한계로 인한 비판의 여지가 있다.(조은경 2016) 언어의 뜻은 변화 여지가 많고, 아주 잘 정제된 형식 의미라 할지라도 특정한 문맥에서의 의미를 살리는 데에는 여전히 한계가 있을 것이며, 특정한 문자열의 단어도 맥락에 따라 의미가 다르기 때문이다. 그래서 어휘부이든 지

식그래프이든 온톨로지이든 언어가 나타내는(denote) 의미는 텍스트 말뭉치에서 학습된 의미가 지속적으로 유입되어야 하며, 단일 어휘 항목 자체에 모든 가능한 의미를 기술하기보다는 지식그래프라는 큰 체계 안에서 관계에 따라 다른 의미를 획득할 수 있게 하는 것이 맞겠다. 또한 인터넷과 디지털 문명의 고도화에 따라 우리의 지식은 많은 부분 인터넷으로 유통이 되므로, 그 표현 형식의 면에서 개체간의 관계와 연결에 중점을 두는 지식그래프는 더 오랜 기간 표현력이 인정될 것 같다.

지식그래프로 자주 언급되는 야고(YAGO: Yet Another Great Ontology)는 막스플랑크연구소에서 개발되어 2008년에 배포를 시작하여 꾸준히 배포되고 있는 대용량 의미 지식베이스이며 IBM 왓슨의 인공지능시스템에 꾸준히 활용되고 있다⁶⁾. 초기에 야고는 워드넷(Wordnet)과 위키피디아(Wikipedia)의 언어 자원을 많이 활용하였고(Suchanek et al. (2007)), 위키피디아(Wikipedia)의 지속적인 성장에 따라 위키피디아의 자원을 꾸준히 활용하고 있다. 워드넷의 자원을 활용한 이유는 워드넷이 가진 유의어집합(Synset)과 잘 정립된 의미 관계들 때문이며, 위키피디아의 자원이 활용되는 이유는 한 웹사이트 문서가 한 주제를 잘 정의하면서 다른 웹사이트 문서와 연결 관계가 잘 구축되어 있기 때문이다.

3. 정서 장애 주제 분야 지식그래프 개발

이제 이 글이 보이고자 하는 사례 연구인 정서 장애 분야의 언어 지식 개발로서 국가건강정보포털 정신 의학용어, 정신의학신문의 기사, 정서 장애 용어로 검색되는 일반 뉴스를 자료로 하여 지식그래프 개발의 과정을 보이고자 한다.

3.1. 웹 데이터 수집

웹에서 언어 자료를 수집하여 활용하는 국어학적 선행 연구로는 말뭉치의

6) [https://en.wikipedia.org/wiki/YAGO_\(database\)](https://en.wikipedia.org/wiki/YAGO_(database)), Retrieved January, February 2020.
<https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>

수집과 구성이라는 문제를 다룬 한영균·이두행(2014), 말뭉치 구성과 디자인의 문제를 다룬 남길임(2018)이 있다. 두 연구 모두 어떤 전문 영역을 상정하지 않고 일반어 연구 목적에 맞는 말뭉치 수집과 구성 상의 쟁점에 관한 것으로, 언어 연구에 있어 웹 자료의 수집의 타당성과 활용성 측면의 의의를 살핀 바 있으며, 일반어를 연구함에 있어 말뭉치 언어학의 미래 방향성에 대한 논의를 담고 있다.

이 글은 일반어 연구라기보다는 정서 장애라는 주제 분야의 사례 연구에서 필요로 하는 언어 자료 수집과 구성을 위해 웹 데이터를 수집하고 활용하는 방안을 제안하고자 한다. 어떤 주제 분야의 핵심 의미는 전문가 집단에서 사용하는 용어일 것이다. 그리고 그러한 용어가 많이 사용되는 전문가가 작성한 글은 용어와 용어 사용의 의미를 살아있게 한다. 나아가 일반인들이 어떤 전문 용어를 쓰는 맥락이나 글은 해당 용어의 의미를 세상 지식과 정보를 소통하는 데에 필요한 개념 의미를 형성하는 데에 필요하다. 무엇보다 오늘날 이러한 언어적 자원이 인터넷 공간에 서비스되는 일이 많고 약간의 프로그래밍을 활용하면 전문 지식을 수집하여 지식베이스 구축에 활용할 수 있다.

3.1.1. 전문용어집

어떤 분야의 전문용어는 해당 분야의 핵심적 개념 단위에 관한 언어적 표현 형태라 할 수 있다. 정서장애 분야를 대상으로 한다고 할 때 어떤 언어 자료를 수집할 것이냐의 문제에 있어, 많은 사람의 언어적 개념 형성에 영향력이 있는 자원을 활용하는 것이 의미가 있을 것이다. 국가건강정보포털의 의학정보는 오늘날 한국인의 수많은 생활에 영향을 미치고 있는 네이버 포털의 지식백과의 한 부분으로 서비스되고 있다. 따라서 이 자원은 많은 노력을 들여서 새로이 구축해야 하는 자원이 아니면서도 언어적 영향력이 큰 자료라 할 수 있다.

<그림 3>은 수집 대상으로 삼은 정신의학용어에 관련된 정보 제공의 예시이다. 용어 항목 하나를 클릭하면 제시되는 항목의 기술에는 개요, 정의, 원인, 증상, 치료 방법 등의 상세한 전문 정보가 포함되어 있다. 따라서 웹에서 이러한 자료를 수집하면 정서 장애 용어 이르면 ‘공황장애’의 정의, 증상, 원인,

치료 방법 등에 관한 전문가 지식에 해당하는 언어적 표현물을 획득할 수가 있다. 또한 검색 포털을 통해 나열되는 정보의 순서는 대개 많은 사람들이 찾아본 로그 기록에 토대하여 순위화 모델(ranking model)이 적용된 것으로 그 제시 순서를 활용할 수도 있다.



<그림 3> 웹으로 서비스되는 국가건강정보포털의 의학 용어 자원.

출처: <https://terms.naver.com/list.nhn?cid=51004&categoryId=51004>

<그림 3>에서 보이듯이 여러 건(2020년 1월 기준, 671건)의 의학 정보가 있지만 그 중 정신의학과 범주의 용어만, 그리고 조회 순으로 30 페이지 안에 있고 정신의학 범주 순위로 상위 27 개만 수집하면, 그 목록은 <표 3>과 같다.

<표 3> 정신의학 용어 수집: 조회순 상위 27개

강박증	인격장애	언어장애
수면장애	외상 후 스트레스 장애	도박중독
성폭력	노인 우울증	말더듬
아동학대	인터넷 중독	식이장애
알코올 중독	불면증	신경증
치매	자살 예방	틱 장애

학습장애	자폐증	공포불안장애
조울증	스트레스	공황장애
조현병	산후 우울증	주의력결핍 과잉행동장애

3.1.2. 전문 분야 신문 기사

개념 단위를 표현하는 용어라기보다는 용어를 포함하여 텍스트 수준에서 그 의미를 활용하여 표현하는 언어 자료는 당연히 그 분야의 텍스트일 것이다. 이러한 텍스트 자원 확보를 위해 정서 장애 관련한 전문 분야의 언어적 자원으로서는 정신의학신문 사이트의 기사를 대상 출처로 삼았다. 정신의학신문 사이트에는 트랜딩토픽이라는 탭이 있고 하위에 자리잡은 주요 기사 주제는 <그림 4>와 같이 주로 정서 장애에 관련한 기사들(불안, 우울, 분노, 치매, 망상, 공포, 중독, ...)로 구성되어 있다.



<그림 4> 전문가 커뮤니티 신문의 하위 주제 범주

출처: <http://www.psychiatricnews.net/>

7) 용어집 표제어가 ‘자살 예방’이라 되어 있는데, 맞는 표제어는 ‘자살’이어야 할 것 같다.

불안·걱정~스트레스에 이르기까지의 기사들은 모두 정서 장애를 표현하는 의미를 풍부하게 표현할 수 있는 언어적 자료이다. 즉, 전문용어집에 제시된 ‘공황장애’에 관한 정의, 증상, 원인, 치료 방법이라는 비교적 제한된 의미를 벗어난 보다 풍부한 내용을 얻어낼 수 있다. 수집한 자료의 예를 들면 <그림 5>와 같다.

공황장애, 6가지 오해와 진실 [정신의학신문 : 김영미 정신건강의학과 전문의] 늘어만 가는 공황장애, 그러나... 임상 현장에서 진료를 보면서 공황장애를 겪는 수많은 환자분을 접하게 됩니다. 그리고 자신이 공황장애라는 사실을 인지하고 외래를 찾아오시는 분들이 꽤 많아졌음에 놀라게 됩니다. 중략 모든 정신과적 질환이 그렇듯 한, 공황장애에 동반되는 이차적인 증상들 (secondary symptoms)-우울, 알코올 중독 등- 을 치료하는데도 약물치료 외의 심리치료가 큰 도움이 됩니다. ...하략
--

<그림 5> ‘공포·두려움’ 주제에 있는 전문가 기사 예시

출처: <http://www.psychiatricnews.net/news/articleView.html?idxno=12594>

3.1.3. 일반 뉴스 기사

전문가 작성 용어나 기사는 해당 분야의 전문 지식인이 작성을 한다. 그런데 정서 장애와 관련한 일은 결국 일반인의 영역에서 벌어지는 일이고 사회적 이슈이기 때문에 일반 신문 기사의 언어를 살필 필요도 있다. 여기서는 일반 신문 기사 수집을 위해 모든 신문사 웹사이트에서 검색하는 것이 아니라 네이버 뉴스 검색을 활용했다. 검색 포털에서 뉴스를 검색하여 나오는 기사는 여러 신문사 출처의 기사를 제공 받아서 글자 수가 많고 광고성 기사가 아니며 사람들에게 많이 읽힌 기사일수록 상위 결과로 제시되기 때문에 비교적 품질 좋은 텍스트로 간주될 수가 있다. 여기서는 27개 정서 장애 용어를 검색어로 하여 검색된 뉴스 기사를 수집하였으며, ‘공황장애’로 검색된 기사 텍스트의 예를 들면 <그림 6>과 같다.

쪽방에 없는 4가지, 온수·부엌·안전·사생활, “홈리스로 분류하자” 인권위 비주택 거주자 실태 조사김모(33)씨는 고시원에 살면서 <u>공황장애</u>를 겪었다. 마트에서 일할 때 손님이 주위로 몰려들면 손이 떨리고 불안했다. 박모(51)씨도 고시원에 살면서 아무 이유 없이 짜증이 났다. 4~5일씩 집밖에 나가지 않는 일이 잦아져 우울증 약을 먹기 시작했다. -하락- 국민일보(www.kmib.co.kr), 무단전재 및 재배포금지
--

<그림 6> ‘공황장애’로 검색된 일반 뉴스 기사 예시

출처: <http://m.kmib.co.kr/view.asp?arcid=0924042664>

3.2. 의미 관계 추출과 연결

용어집은 용어의 정의, 증상, 원인, 치료 방법이라는 비교적 제한된 의미를 상세히 기술하고 있지만, 기계 가독형의 지식은 아니기 때문에 일정 수준의 정보 처리를 요구한다.

전문가 기사나 일반 기사는 텍스트이기 때문에 구조화된 지식을 추출하기 위한 언어 처리가 있어야 하고, 시의성을 갖는 뉴스이기에 얻어낼 수 있는 언어적 정보가 있고 이를 추출할 수 있는 방안이 있어야 한다.

3.2.1. 의미 관계의 유형을 드러내는 어휘와 표현

‘조현병’이라는 한 용어에 대한 기술은 서술식이고 전문은 4,266 어절이다. 용어집을 활용해서 정서 장애 용어의 의미 관계 추출을 한다면, 단서 단어나 특정한 구문적 표현을 활용할 수 있다. 특히 ‘증상’, ‘진단’, ‘치료’, ‘처방’, ‘요인’, ‘원인’ 등의 단어는 용어집에서 해설의 항목으로 쓰이며, ‘발병’, ‘부작용’, ‘작용’, ‘특징’, ‘결과’, ‘효과’, ‘~원인’, ‘~ 경우’, ‘영향을 미치다’, ‘중요한 역할을 하다’ 등이 항목 기술에 자주 쓰인다. 이러한 단서 단어와 구문 패턴을 활용하여 그 주요 내용을 축약하여 제시하면 <그림 7>과 같다. 이 축약 내용은 129 어절로 핵심적인 내용만을 갖춘 간략한 형태로, 지식그래프로 구조화되기 전에 중간 단계라 할 수 있다.

이제 이렇게 보다 핵심적인 내용으로만 구성된 전문가 용어 해설집을 간략

한 구조의 지식그래프의 테이터로 구조화하려고 한다면, ‘개요’는 수많은 온톨로지나 의미 관계에 관한 연구(최호섭 외(2006), 옥철영(2007), 임지희 외(2008))에서 활용하듯이 ‘IsA’ 관계 추출의 언어이고, ‘증상, 원인, 치료’는 연결 즉 의미 관계의 이름이 되며 각 항목의 내용이 해당 의미 관계로 연결이 되는 식으로 구조화될 수 있다.

조현병(schizophrenia)

개요:조현병(정신분열병)이란 사고(思考), 감정, 지각(知覺), 행동 등 인격의 여러 측면에 걸쳐 광범위한 임상적 이상 증상을 일으키는 정신 질환입니다.

-하략-

증상:여기서는 조현병(정신분열병)의 증상을 이해하기 쉽도록 ‘양성 증상’과 ‘음성 증상’, ‘인지 증상’ 그리고 급성기 이후에도 계속 남아 있는 ‘잔류 증상’의 네 가지로 분류하여 설명하고자 합니다.

-하략-

원인:과학의 발달로 인해 뇌에 대한 연구가 차근차근 진행되고 있어 조현병(정신분열병)의 원인도 하나씩 밝혀지고 있습니다. 현재까지 밝혀진 바로는 조현병(정신분열병)은 어느 한 가지 원인이 아닌 여러 요인, 크게는 유전적 요인과 환경적 요인의 복합적 결과입니다.

-하략-

치료:1. 약물치료 현재까지 밝혀진 다양한 원인만큼 조현병(정신분열병)의 치료에도 다양한 방법이 있습니다. 그러나 만일 단 한 가지 치료 방법만을 선택해야 한다면, 그것은 두말할 것도 없이 “약물 치료”입니다.

-하략-

<그림 7> 용어집 ‘조현병’의 내용 축약: 지식그래프로 구조화되기 이전의 축약 형태

전문 4266 어절의 내용은 <그림 7>과 같은 중간 단계를 거치고, 단서 단어와 구문을 활용하여 <표 4>와 같은 핵심적이고 구조화 가능한 형태로 가공된다. 여기서 ‘<...>’로 싸인 단어나 구문은 지식 그래프에서 연결 관계가 있는 것이다. 예를 들어, <표제어>은 이 명칭에 대한 설명으로 연결될 것이고, ‘환청’, ‘환시’, ‘환각’, ‘도파민’, ‘뇌실’과 같은 단어는 다른 개념 노드로 연결된다. 이처럼 지식그래프는 텍스트로 된 긴 언어적 정보를 구조적인 지식으로 표현하고

연관된 지식을 ‘연결’하게 함으로써 짧은 시간에 핵심적인 내용을 이해할 수 있고 연관 지식으로 이동할 수 있는 이점을 갖는다.

<표 4> 지식그래프로 구조화될 핵심 내용

*다른 개념 표현의 노드로 연결이 될 수 있는 것은 < >로 표현하였다.

<표제어>:조현병 <질병 분류>: 정신질환, <정신분열병> <진단 및 증상> <양성 증상>: <환청>, <환시>, <환촉>, <환각>, 과대 망상, 피해 망상 <음성 증상>: 정상적 감정 반응의 둔화, 의욕 감퇴, 사회적 위축 <인지 증상>: 집중력 저하, 학습 능력 저하 <잔류 증상>: 음성 증상과 인지 증상의 복합 <원인> <생물학적 요인> <유전 요인>: 1차 직계 가족, 쌍둥이 <생화학적 요인>: <도파민>, 세로토닌 불균형. 뇌세포 차이. <뇌실> 크기. <환경적 요인>: 성장 과정의 심리적 요인 <발병 연령>: 15세-25세 <치료> <약물 치료> <항정신병 약물>:<할로페리돌(haloperidol)>, <클로로프로마진(chlorpromazine)>, <플루페나진(fluphenazine)>, <클로자핀(clozapine)>. <부작용>: 졸림, 어지러움, 피부 발진 <항우울제>* <사회 심리 치료>: 규칙적이고 지속적으로 약물을 복용하도록 지원. <기타>: 대부분 가족 대상의 폭력, 자살 시도자가 많음.
--

그런데 4,266 어절의 전문을 다 읽고 구조화된 요약 정보를 갖춘다 해도 여전히 부족한 정보가 있을 수 있다. ‘<항우울제>’에 대해서는 구체적인 약제 이름이 없는 것이 그 예가 된다. 또 얻고자하는 정보가 지식그래프라는 표현 형식을 통해 얻는 단순한 요약적 지식보다는 동적이고 시의성 있는 정보일 수도 있다. 이러한 정보 요구는 사건과 스토리를 담고 있는 뉴스 텍스트를 통해 구해 볼 수 있다.

3.2.2. 전문가 기사에서의 지식 추출

정서 장애 용어의 지식그래프 구성에 있어, 정신의학신문의 콘텐츠는 정서 장애 문제의 처방이나 대응에 관련하여 좋은 사실 지식(factual knowledge)을 얻을 수 있는 출처이다. 용어 해설집의 지식의 어떤 한 용어에 대한 집중적인 지식을 담고 있다면, 전문가 지식은 그 지식을 확장해갈 수 있기 때문이다. <그림 8>은 정신의학신문 기사에서 ‘조현병’으로 검색되는 즉, 조현병과 연관된 기사의 상단을 발췌한 것이다. 여기에는 조현병의 용어 해설에서 다루지 않았던 ‘우울증’과의 연관성이나 처방 약제 관련성까지 담고 있다.

우울증인데 조현병 약을 처방받았어요 - 항정신병약제사용
[정신의학신문 : 임찬영 정신건강의학과 전문의] “우울증상이 있어서 정신건강의학과를 다니고 있습니다. 그런데 약 성분 중에 조현병 약이 들어있어요. ‘혹시 약물이 잘못 처방된 것이 아닌가요?’ ‘전문의가 보기에는 제가 조현병 증상을 보인다 생각하는 건가요?’” 불안, 불면, 우울 등이 주된 증상인데 조현병약이 처방되었다면서 문의하시는 환자 분들이 종종 있습니다. 사실 조현병 약제, 다른 말로는 항정신병약제 중에 일부는 용량에 따라서 우울증/조현병 모두에 사용 가능합니다. 항정신병약제 중 몇 가지 제제는 우울증상, 공황증상 등에 매우 효과적입니다.
-하략-

<그림 8> 정신의학신문 기사: ‘조현병’으로 검색된 기사

출처: <http://www.psychiatricnews.net/news/articleView.html?idxno=12461>

전문가 기사에 기술되어 있는 조현병과 우울증이 연관된 질병이라는 것과 그래서 같은 약제를 쓸 수 있음에 대한 지식을 지식그래프 안에 담아낸다면, <표 5>와 같이 될 것이다.

<표 5> 연관 질병 관계

<연관질병>: <조현병>, <우울증>, <공황장애>
<공통 약제>: 항정신병약제: <아빌리파이>(abilify, aripiprazole), <세로켈>(seroquel, quetiapine)>

<그림 9>는 ‘공황장애’로 검색된 2018년 7월부터 2018년 12월까지의 뉴스 기사 790 건을 수집하고 워드투백을 활용하여 연관도가 높은 상위 100 단어의 분포를 시각화한 것이다. 이 공간 안에 들어와 있는 단어들의 공기 연관도가 이 안에 표현되지 않은 다른 단어 대비 상대적으로 큰 단어들이라는 것이다. 그리고 의미적으로 보다 연관된 단어들 ‘겪다’, ‘당하다’, ‘폭행’이 가까이 있고, ‘사회’, ‘우리’, ‘나’가 가까이 있으며, ‘불안’, ‘스트레스’, ‘정신’, ‘질환’ 등이 가까이, 그리고 ‘공황장애’와 ‘우울증’이 가까이 있다. 즉 대용량 언어 자료를 가지로 의미적 연관도가 큰 단어 집합을 추출하고자 할 때, 워드투백을 활용하여 어떤 개념 단어(‘공황장애’)와 연관도가 큰 단어들과 상대적인 의미 연관도가 큰 단어들을 추려내는 일에 효율을 더할 수 있다. 그리고 여기에 이러한 결과를 해석하고 활용하기 위해 우리의 언어적 직관이 어느 정도 작용을 한다.

그리고 이 그림의 왼쪽 가운데 즈음에는 ‘성폭력’, ‘김기덕’, ‘피해자’라는 단어가 가까이 있고, 이보다 오른쪽에는 ‘이찬오’, ‘폭행’이라는 단어가 있다. 모두 의미적 특정성이 있다. 실제로 해당 시기인 2018년 7월~12월에 사회적 이슈가 되었던 인물이며 ‘공황장애’와 관련된 인물들이다. 이를 통해 워드투백을 적절히 활용하면, ‘공황장애’라는 정서 장애가 사회적 문제와 관련되어 있음을 발견하는 데에 활용될 수 있음을 볼 수 있다.

<그림 10>은 ‘조현병’으로 2018년 7월부터 2018년 12월까지 배포된 뉴스 기사를 검색하여 수집된 71,610건⁸⁾을 가지고 그 분포 의미를 시각화한 것이다. ‘조현병’ 뉴스 텍스트는 형태소 분석기를 써서 어휘 형태를 정규화하였을 때, ‘동물교감치유’는 ‘동물’, ‘교감’, ‘치유’가 각각의 단어가 될뿐만 아니라 그 외에 의미 특정성이 떨어지는 단어가 분포 의미 공간에 자리를 잡았기 때문에 형태소 분석을 하지 않은 텍스트로 분포 의미를 구하였다. 그래서 ‘동물교감치유에’, ‘동물교감치유를’, ‘동물교감치유’가 각각의 노드로 되어 있지만, 각기 다른 조사가 붙었음에도 불구하고 이 공간 안에 자리잡았다는 것은 그만큼 ‘조현병’과의 연관도가 크다는 말이 된다.

8) ‘공황장애’의 뉴스 기사 검색 기간과 같은 기간이지만, ‘조현병’에 관한 이슈가 많았고 뉴스 기사 수도 훨씬 많다. 설명의 편의를 위해 그림의 왼쪽만 잘라서 넣었다.



<그림 10> ‘조현병’의 분포 의미 시각화

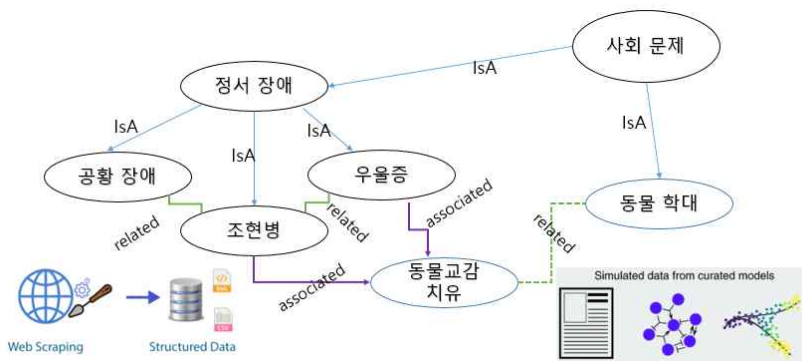
즉, 이 결과는 ‘조현병’이 사회적 문제 측면에서는 ‘동물교감치유’와 연결됨을 드러낸다. 다시 말해 정서 장애, 반려 동물, 동물 학대는 우리 사회에 유관한 사회적 문제임을 드러낸다고 할 수 있다. 이러한 점을 볼 때 정서 장애 용어와 연관된 일정 시점의 사회 문제를 일반 뉴스 기사를 활용하여 그 언어적 표현을 지식그래프에 노드로 나타내고 연결할 수 있다.

요컨대 지식그래프 구성에서, 전문가 용어집이나 전문가 기사를 활용하여 주로 전문 지식에 의존한 개념 관계와 언어적 지식을 표현할 수 있다면, 어떤 시점의 일반 뉴스 기사를 반영하면 해당 시점에 유의미했던 연관 개념들을 언어적으로 표현할 수 있다.

3.3. 전문용어, 전문가 기사, 일반 사회 뉴스의 지식그래프

지식그래프는 용어집의 구조적 설명력, 용어 지식과 연결되는 전문가 기사의 풍부한 지식, 개념 용어와 연관된 사회적 이슈의 발견을 엮어낼 수 있다.

<그림 11>은 지금까지 살펴온 방법론과 ‘공황장애’, ‘조현병’, ‘동물교감치유’의 의미 연결을 중심으로 간략하게 표현한 것이다.



<그림 11> 지식그래프 개발과 의미 표현

<표 6>은 조현병을 중심으로 조현병의 의학 용어적 지식, 연관 질병과 약제

에 관한 전문 지식, 사회적 이슈와 관련된 지식을 언어적으로 표현한 것이다. 여기에서는 텍스트로 표현이 될 수밖에 없지만, 웹 사이트에서 이러한 지식의 내용을 표현한다면, 연결되는 데이터는 다른 웹사이트 문서와 연결되어서 입체적 양상을 보일 수 있다.

<표 6> 지식그래프의 ‘조현병’ 노드의 중심 내용

<표제어>:조현병		
<질병 분류>: 정신질환, <정신분열병>		
<진단 및 증상>		
..... 중략		
<치료>		
<약물 치료>		
<항정신병	약물>:<할로페리돌(haloperidol)>,	<클로로프로마친(chlorpromazine)>,<플루페나진(fluphenazine)>,<클로자핀(clozapine)>.
... 중략		
<연관질병>: <우울증>,<공황장애>		
<공통 약제>: 항정신병약제: <아빌리파이(abilify, aripiprazole)>,<세로켈(seroquel, quetiapine)>		
<사회적 이슈>:<동물교감치유>,<동물학대>		

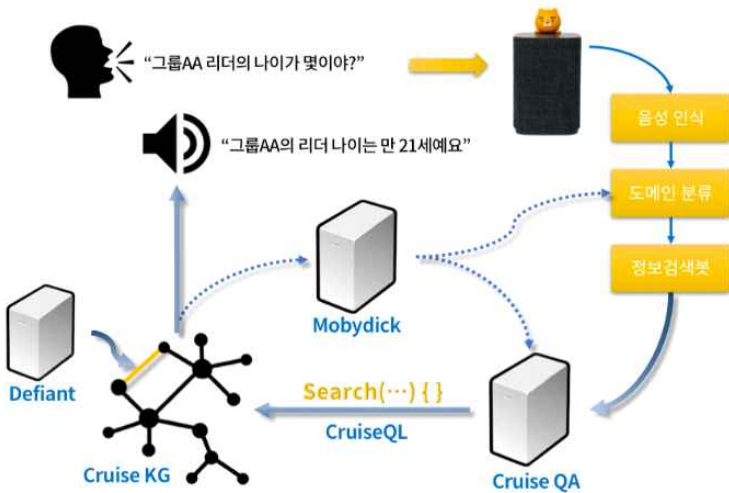
<표 6>과 같은 사전적 내용은 ‘공황 장애’의 내용으로 채워서 표현될 수도 있고, ‘조현병’이나 ‘공황 장애’의 지식 노드는 <그림 11>에서와 같은 연결 관계가 있는 그래프 구조 안에 있게 된다.

4. 지식 그래프의 응용과 맺음말

지금까지 지식그래프를 구성하기 위해 용어해설집, 전문가 텍스트로서 전문영역 기사, 일반인 텍스트로서 신문 기사를 수집하여, 언어적 지식과 연결 관계를 구축하는 방법에 대해 살펴보았다. 특히 나뭇가지가 나무의 일부분이라는 자명한 사실이나 유명 인물의 출신 학교나 나이가 어떻다는 자명한 사실보다는, 물론 자명한 사실이 의미적 추론의 기본이 되는 분명하지만, 발견적 사실이나

사건을 지식그래프 안에 담아낼 수 있다는 점에서 뉴스 기사는 언어적 지식 발굴에 유용한 자원이며, 이때 그 지식의 발견에 있어서는 실제 세계에서 자주 공기하는 언어적 표현이라는 점과 이에 대한 필자의 세상 지식을 결합한 설명이 개입되었던 것처럼 지식그래프 구성에는 언어 현상과 이에 관련된 언어 의미적 연관성, 그리고 그에 관한 설명력도 필수적이다.

지식그래프는 인간의 지식 체계를 모방하고자 만든 것이기 때문에 관계 기반, 의미 기반 검색이나 질의응답 시스템의 자연어 지식 처리에 활용되기도 한다. <그림 12>는 실제 상용화된 언어 처리 서비스에서 지식그래프가 대화형 상호작용을 가능케 하는 언어지식으로 쓰이는 사례이다.



<그림 12> 대화형 검색(카카오 미니)에서의 지식그래프 활용 사례

출처: <https://brunch.co.kr/@kakao-it/281>

과거, 1900년대 후반에서 2000년대 초반까지, 인공지능 시스템은 전문가 시스템(Expert system)이라 불릴 만큼 전문가의 지식이 중요했다. 딥러닝 기법이 쓰나미 같은 물결로 전산언어학뿐만 아니라 많은 분야 연구자들의 연구 방법론을 바꾸고 있고, 기계독해(MRC:Machine Reading Comprehension)⁹⁾이라는 연구

주제로는 긴 특정 전문 분야 텍스트를 학습하여 질문의 빈칸에 답을 찾아내는 일도 활발히 진행되고 있기도 하지만, 많은 실제 응용 시스템에서는 여전히 전문가 지식이 유효하며 설명가능한(explainable) 시스템 결과를 내기 위해 전문 영역 지식이 요구되고 있다. 그래서 딥러닝의 최근 응용 주제 중 하나는 지식 그래프 구축과 개발을 위해 딥러닝 기법을 활용하는 일도 많아졌다.(Trouillon et al.(2017), Nathani et al.(2019), Bonatti et al.(2019), Gesese et al.(2019)).

지식그래프라 불리는 지식 표현 방법은 숫자라기보다는 가시적이고 언어적인 설명력이 있는 개념 노드와 그것들 간의 의미적 연관성을 물리적 연결로 표현함으로써 고도화되고 있는 디지털 방법론에 인간적 개입을 더욱 체계화시키는 기제이다. 지식그래프의 내용은 그것이 지식인지 아닌지의 구별이 요구되기에 언어적 지식으로서의 인간적 가치 판단이 들어갈 여지가 크다. 오늘날과 같이 대용량 데이터로부터 정보를 추려내는 기법의 발달이 언어 연구자의 분석 결과를 활용할 부문에 대한 기대치를 축소시키고 인문학자의 역할에 대한 위기를 불러일으키는 때에, 지식그래프는 지식을 구조화하고 연관 지식을 연결해주는 언어 지능(linguistic artificial intelligence)을 표현하는 기제라는 점에서 언어 자료를 구성하여 어휘부로 구조화하는 기법의 연장선상에서 연구해갈 필요가 있겠다.

이 글에서는 지식그래프의 적용을 정서 장애 분야에서의 전문용어와 연관된 개념, 그리고 그 맥락 의미를 담은 텍스트에서 추출한 개념어를 연결하는 수준에 그쳤지만, 보다 언어학적 지식 분야를 다룬다면 말뭉치 언어학적 의의가 매우 클 것이다.

9) 마치 사람이 지문을 읽고 이해한 내용을 답하는 일(Reading Comprehension)을 하는 것과 같은 목적으로 가지고 있다.

참고문헌

- 김일환(2019), “신문 빅데이터 기반의 언어 계량과 시각화”, 언어와 정보 사회 38, 서강대 언어정보연구소, 335-354.
- 김학래(2010), “시맨틱 웹과 의미적 연결성 : 웹 사이언스를 위한 출발점”. 정보과학회지 제28권 제3호. 한국정보과학회. 31-40.
- 김학래(2017), 지식그래프. 데이터사이언스 총서. 커뮤니케이션북스.
- 남길임(2018), “웹 말뭉치를 활용한 언어 연구의 현황과 쟁점”. 한국어의미학 60, 한국어의미학회. 23-49.
- 송도규(2006), “한국어 부정표현의 형식화와 온톨로지 변환”. 한국어의미학 19. 한국어의미학회. 85-104.
- 양성권·김흥기(2019), “개체 링크를 위한 RDF 지식그래프 기반의 포괄적 상호의존성 짝 연결 접근법”. 정보처리학회논문지 소프트웨어 및 데이터공학 8(3) 한국정보처리학회. 129-136.
- 옥철영(2007), “On the Evaluation of Korean WordNet”, LECTURE NOTES IN ARTIFICIAL INTELLIGENCE 4629, 123-130.
- 윤애선(2018), “세분화된 한국어 감정 온톨로지 KoFiGEmOnto의 구축과 응용”. 한국사전학 32. 한국사전학회. 61-94.
- 이수경(2005), “효율적인 암정보 제공을 위한 온톨로지 기반 대화시스템 개발”, 연세대학교 대학원 석사학위논문.
- 이숙의(2019), “온톨로지 지식과 개체명 세분화를 통한 맥락의미 자동인식의 향상”. 한국언어문학 109. 한국언어문학회. 37-59.
- 임지희·최호섭·옥철영(2008), “어휘망(U-WIN)의 구문관계 자동구축”. 정보과학회논문지 : 소프트웨어 및 응용 35(10), 627-635.
- 조은경(2000), “전문용어와 전문어 말뭉치”. 전문용어연구 2. 전문용어언어공학연구센터. 201-229.
- 조은경(2016), “세종전자사전 활용을 위한 자연어 생성적 접근—어휘문법적 미시 구조의 격자 형식화를 통해—”. 한글 311. 한글학회. 197-226.
- 조은경(2019ㄱ), “단어 임베딩으로 보는 한국어 젠더의 의미 변화”, 언어와 문화 15-3, 한국언어문화교육학회, 221-241.
- 조은경(2019ㄴ), “언어 정보의 계량화와 시각화: 감성 어휘의 분포 의미 표현”, 언어와 정보 사회 38, 서강대 언어정보연구소, 413-436.
- 최호섭·임지희·배영준·최수일·옥철영(2006), “온톨로지 구축 방법과 사례”. 정보과학회지 제24권 제4호. 한국정보과학회. 31-44.
- 한영균·이두행(2014), “웹 문서와 인터넷을 이용한 현대 한국어 참조 코퍼스의 구축”. 어문학 124. 한국어문학회. 129-166.

- _____, https://en.wikipedia.org/wiki/Knowledge_Graph, Retrieved 10 Feb. 2020.
- _____(2019). “Knowledge Graph Generation for Financial Databases”.
<https://blog.bitext.com/knowledge-graph-generation-for-finance-databases>
- Berners-Lee, T., James Hendler, Ora Lassila(2001). *The Semantic Web*, Scientific American.
- Bonatti, P.A., Decker, S., Polleres, A. and Presutti, V.(2019). “Knowledge graphs: new directions for knowledge representation on the semantic web” (*Dagstuhl seminar* 18371). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- Bowman, Samuel R.(2016). *Modeling natural language semantics in learned representations*. Stanford University Ph.D. Dissertation.
- Chemla, E. and Spector, B.(2011). “Experimental evidence for embedded scalar implicatures”. *Journal of semantics*, 28(3), 359-400.
- Ehrlinger, Lisa; Wöß, Wolfram(2016). “Towards a Definition of Knowledge Graphs”.<https://towardsdatascience.com/auto-generated-knowledge-graphs-92ca99a81121>. Retrieved 10 Feb. 2020.
- Gesese, G.A., Biswas, R. and Sack, H.(2019). “A comprehensive survey of knowledge graph embeddings with literals: Techniques and applications”. In Workshop on *Deep Learning for Knowledge Graphs*.
- Hirst, Graeme(2010). “Ontology and the Lexicon”. *Handbook on ontologies*. Springer Science & Business Media.
- James, P. McCusker(1992). “Knowledge Graphs”. *Linguistic Instruments in Knowledge Engineering*, Elsevier Science Publishers B.V., 97 - 117.
- Minsky, Marvin(1985). *The society of Mind*. Simon & Schuster Paperbacks.
- Nathani, D., Chauhan, J., Sharma, C., & Kaul, M.(2019). “Learning attention-based embeddings for relation prediction in knowledge graphs”. *ACL Proceedings*. arXiv preprint arXiv:1906.01195.
- Ray Jackendoff(2015). *User’s Guide to Thought and Meaning*, Oxford University Press.
- Singhal, Amit(2012). “Introducing the Knowledge Graph: things, not strings”.
Official Google Blog.
<https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html> Retrieved 10 Feb. 2020.
- Staab, S., & Studer, R. (eds.)(2010). *Handbook on ontologies*. Springer Science & Business Media.
- Stevens, R., Lord, P., Malone, J., & Matentzoglu, N.(2019). “Measuring expert performance at manually classifying domain entities under upper ontology

- classes”. *Journal of Web Semantics*, 57, 100469.
- Suchanek, Fabian M., Gjergji Kasneci, and Gerhard Weikum(2007). “Yago: a core of semantic knowledge.” In Proceedings of the 16th international conference on *World Wide Web*, pp. 697–706.
- Trouillon, T., Dance, C. R., Gaussier, É., Welbl, J., Riedel, S., & Bouchard, G. (2017). “Knowledge graph completion via complex tensor factorization”. *The Journal of Machine Learning Research*, 18(1), 4735–4772.
- Udo Hahn and Stefan Schulz(2010). “Building a Very Large Ontology from Medical Thesauri”. *Handbook on ontologies*. Springer Science & Business Media.

강승식(교신 저자)
국민대학교(교수)
서울시 성북구 정릉로 77
국민대학교 소프트웨어융합대학
02707
전화 번호 : 02-920-4800
전자 우편 : sskang@kookmin.ac.kr

조은경(제1 저자)
서강대학교(조교수)
서울시 마포구 백범로 35
서강대학교 지식융합미디어학부 글로벌한국학전공
04107
전화 번호 : 02-705-4742
전자 우편 : jek.cl.nlp@daum.net

원고 접수일 : 2020. 02. 20.
원고 수정일 : 2020. 03. 12.
게재 확정일 : 2020. 03. 17.