

데이터분석기초

Seongmin Mun

30th March 2023



아주대학교
AJOU UNIVERSITY

Outline

온톨로지의 개념

단어 기반 온톨로지를 사용한 말뭉치 분석

단어 기반 온톨로지 사전 Source

온톨로지의 개념

온톨로지란?

- ▶ 정보화의 대상(things)의 개체(individual), 클래스(Class, domain), 속성(attribute, properties), 관계(relation)를 정의한 분류 기준(Gruber, 1992).

온톨로지란?

Books

Title	Author	Publisher	Year Published	Followed By
To Kill a Mockingbird	Harper Lee	J. B. Lippincott Company	1960	Go Set a Watchman
Go Set a Watchman	Harper Lee	HarperCollins, LLC; Heinemann	2015	
The Picture of Dorian Gray	Oscar Wilde	J. B. Lippincott & Co.	1890	
2001: A Space Odyssey	Arthur C. Clarke	New American Library, Hutchinson	1968	

Publishers

Name	City	Country
J. B. Lippincott & Company	Philadelphia	United States
HarperCollins, LLC	New York City	United States
Heinemann	Portsmouth	United States
New American Library	New York City	United States
Hutchinson	London	United Kingdom

Authors

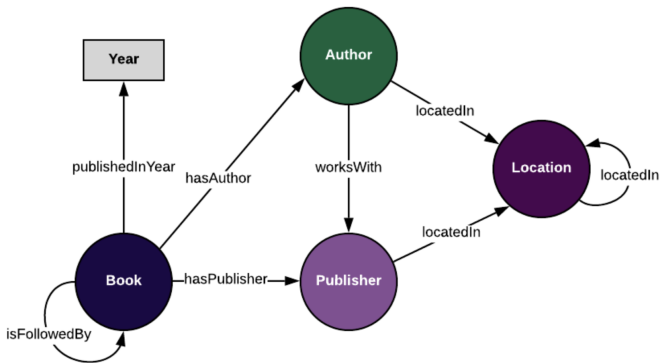
Name	Country of Birth
Harper Lee	United States
Oscar Wilde	Ireland
Arthur C. Clarke	United Kingdom

온톨로지란?

- Book → has author → Author
- Book → has publisher → Publisher
- Book → published on → Publication date
- Book → is followed by → Book
- Author → works with → Publisher
- Publisher → located in → Location
- Location → located in → Location

출처-EnterpriseKnowledge

온톨로지란?



출처-EnterpriseKnowledge

단어 기반 온톨로지를 사용한 말뭉치 분석

단어 기반 온톨로지 적용 사례 - 1

- ▶ 단어 기반 온톨로지를 역사 데이터에 적용한 사례
- ▶ 논문: Bak et al., 2016

단어 기반 온톨로지 적용 사례 - 1

텍스트 마이닝을 통한 조선왕조실록 내 왕의 통치 스타일 분석

박진영, 선우연, 오혜연

한국과학기술원

jy.bak@kaist.ac.kr, yeon.seonwoo@kaist.ac.kr, alice.oh@kaist.edu

Mining the text of the Annals of the Joseon Dynasty to analyze
the king's ruling style

JinYeong Bak, Yeon Seonwoo, Alice Oh
KAIST

Bak et al., 2016

단어 기반 온톨로지 적용 사례 - 2

- ▶ 단어 기반 온톨로지를 음식점 리뷰 분석에 적용한 사례
- ▶ 논문: So and Shin, 2016

단어 기반 온톨로지 적용 사례 - 2



음식점 리뷰 감성분석을 통한 세부 평가항목별 평점 예측

Rating Prediction by Evaluation Item through Sentiment Analysis of Restaurant Review

한국컴퓨터정보학회논문지

약어 : 컴퓨터정보논문지

2020, vol.25, no.6, 통권 195호 pp. 81-89 (9 pages)

DOI : [10.9708/jksci.2020.25.06.081](https://doi.org/10.9708/jksci.2020.25.06.081)

발행기관 : 한국컴퓨터정보학회

연구분야: 공학 > 컴퓨터학

소진수 /SoJinSoo¹, 신평섭 /Shin, Pan-Seop²

¹대전대학교

2대전대학교

So and Shin, 2016

단어 기반 온톨로지 적용 사례 - 3

- ▶ 단어 기반 온톨로지를 영화 리뷰 분석에 적용한 사례
- ▶ 논문: Mun and Lee, 2016

단어 기반 온톨로지 적용 사례 - 3



통계와 시각화를 결합한 데이터 분석: 예측모형 대한 시각화 검증

Data analysis by Integrating statistics and visualization: Visual verification for the prediction model

디자인융복합연구(구.인포디자인이슈)

약어 : SDC

2016, vol.15, no.6, 통권 61호 pp. 197-216 (20 pages)

UCI: [G704-SER000008947.2016.15.6.009](#) ?

발행기관 : 디자인융복합학회(구.한국인포디자인학회)

연구분야: **복합학** > 감성과학











문성민 /Seongmin Mun¹, 이경원 /KyungWon Lee²

¹아주대학교 라이프미디어 협동과정 통합디자인연구실²아주대학교

Mun and Lee, 2016

단어 기반 온톨로지 사전 Source

단어 기반 온톨로지 사전 Source - 1

 park1200656	Update SentiWord_info.json	c95a8a9 on
 KnuSentiLex	dic update	
 data	Update SentiWord_info.json	
 README.md	README second update	
 ReadMe.txt	knu lexicon files	
 SentiWord_Dict.txt	Update SentiWord_Dict.txt	
 knusl.py	knu lexicon files	
 neg_pol_word.txt	Update neg_pol_word.txt	
 obj_unknown_pol_word.txt	knu lexicon files	
 pos_pol_word.txt	Update pos_pol_word.txt	

데이터 링크: KNU 한국어 감성사전 (공부정; Polarity)

단어 기반 온톨로지 사전 Source - 2



한국어 감정 디지털 온톨로지 구축에 관한 연구

The construction of digital ontology for Korean emotions

한국어 의미학

2020, vol.68, pp. 131-162 (32 pages)

DOI : [10.19033/sks.2020.6.68.131](https://doi.org/10.19033/sks.2020.6.68.131)

발행기관 : 한국어의미학회

연구분야 : 인문학 > 한국어와문학

이유미 /Yi Yu-mi¹, 박지영 /Jiyoung Park², 김바로 /Kim Ba-ro³¹중앙대학교

²중앙대학교 인문콘텐츠연구소

³중앙대학교 인문콘텐츠연구소

데이터 링크: Lee et al., 2016

참고문헌

Gruber, Thomas R. (1992). A Translation Approach to Portable Ontology Specifications, Knowledge Systems Laboratory Technical Report KSL, 92-71, Stanford University.

이유미, 박지영 and 김바로. (2020). 한국어 감정 디지털 온톨로지 구축에 관한 연구. *한국어 의미학*, 68, 131-162.

소진수 and 신판섭. (2020). 음식점 리뷰 감성분석을 통한 세부 평가항목별 평점 예측. 한국컴퓨터정보학회논문지, 25(6), 81-89.

문성민 and 이경원. (2016). 통계와 시각화를 결합한 데이터 분석: 예측모형 대한 시각화 검증. 디자인융복합연구(구.인포디자인이슈), 15(6), 197-216.

박진영, 선우연, 오혜연. (2016). 텍스트 마이닝을 통한 조선왕조실록 내 왕의 통치 스타일 분석. 한국정보과학회, 741-743.