

온톨로지 자동 구축과 온톨로지를 위한 지속적 자기 개선 모델에 대한 연구

김윤덕* · 김기범** · 김종배***

*,*** 숭실대학교 SW특성화대학원

** 숭실대학교 대학원 IT정책경영학과

E-mail : kjb123@ssu.ac.kr

요 약

수동적 온톨로지 구축은 해당 도메인의 지식을 가진 전문가가 필요하고, 시간적인 소모가 크다. 또한 완성된 온톨로지의 수동적인 지속적 개선은 상당한 비용을 초래할 수 있다. 그래서 온톨로지의 자동 구축과 지속적 자기 개선 방법이 하나의 해결책이 될 수 있을 것이다. 따라서, 이 논문에서는 기계 학습을 통한 온톨로지 구축의 자동화 방법과 지속적 자기 개선 모델을 소개하고자 한다.

키워드

지속적 자기 개선 모델, 온톨로지, 온톨로지 자동 구축, 기계 학습

I. 서 론

기존의 수동적 온톨로지 구축은 인간의 손을 거쳐, 도구에 의해 완성되는 구축 방법이다. 하지만, 해당 온톨로지를 구축하기 위한 도메인의 지식을 갖춘 전문가가 필요하고, 해당 도메인을 정의, 분석, 설계, 구현하기 위한 시간이 많이 소요된다. 또한, 이미 완성된 온톨로지의 개선을 위해 지속적으로 관리하기 위한 비용도 고려해야 한다 [1][2].

따라서, 기계 학습을 통한 자동적 온톨로지 구축에 대해 생각해 볼 수 있다[3]. 기계가 해당 도메인에 나타나는 각 지식에 대한 관계를 분석하여, 지식들 간의 상호 연관성을 스스로 도출해 낸다면, 해당 도메인의 지식을 가진 전문가가 개입하지 않아도 될 것이다. 이를 통해 시간적 측면에서 개선이 이뤄질 것이고, 재사용성이 어느 정도 보장되어 있는 온톨로지 자동화 시스템을 구축했을 경우, 비용적인 측면에서도 개선을 기대할 수 있다.

기계 학습은 지속적인 자기 개선 방법에도 적용할 수 있다. 온톨로지가 상주하는 시스템은 주기적으로 새로운 데이터를 받고, 그 데이터를 분석하여 온톨로지를 업데이트 할 수 있다.

II. 온톨로지 자동 구축

온톨로지 자동 구축의 핵심은 기계 학습이다. 기계 학습에서 다루는 유사도 계산은 각 항목들 간의 관련성을 계산하는 것이고, 군집 알고리즘을 통해 공통된 영역에 속하는 데이터를 식별하고 분류하는 것이다[5].

온톨로지의 구조는 지식망으로 구성되어 있고 구조 내의 지식과 지식은 상호 연관성을 가지고 연결되어 있다. 유사도 계산은 지식 간의 상호 연관성을 계산해 줄 것이고, 군집 알고리즘은 타겟 도메인에 해당하는 지식을 구분하고 분류하여 줄 것이다.

(a) 온톨로지 구축에 필요한 데이터 수집 방법

해당 도메인에 해당하는 지식 체계를 만들기 위해서는 그 원천이 되는 자원이 필요하다. 특정 도메인에 해당하는 말뭉치 자료를 수집하거나, 특정 커뮤니티들의 웹 페이지들을 크롤링한다.

(b) 기계 학습을 위한 입력 데이터로 정형화 하기

비정형 상태의 자원들을 기계 학습을 위한 입력으로 만들기 위해, 맵 리듀싱 처리가 필요하다. 해당 자원들에서 어떤 키워드가 나타나는 빈도수

를 키와 값의 형식으로 정형화 한다. 또한, 문단 단위 내에서 인접한 단어들의 키, 값 쌍을 만든다.

(c) 기계 학습 프로세스

기계 학습을 위한 입력 데이터를 학습 데이터로써 추가하고, 임계값을 지정하여, 발생 빈도가 임계값보다 많은 데이터를 해당 도메인에 해당하는 키워드라고 고려한다.

인접한 단어들의 키, 값 쌍이 해당 도메인에 해당하는 키워드인지 검사하고, 맞다면, 유사도 측정을 실행한다. 이를 통해 관련성이 높은 키워드들은 온톨로지 토폴로지에서 인접한 곳에 위치할 것이다.

결과적으로 온톨로지는 군집 알고리즘에 의해 해당 온톨로지로 식별되고 분류될 것이다.

III. 온톨로지의 지속적 자기 개선을 위한 방법

이미 완성되어 있는 온톨로지의 데이터 셋과 새로 유입된 데이터의 정형화된 입력을 토대로 자기 개선을 수행한다.

관련된 데이터가 주기적으로 다시 수집될 수 있도록 시스템을 구축하고, 그 수집 데이터들을 정형화하여 기존의 온톨로지 데이터 셋에 더하여 재분석을 수행하고, 결과를 반영하여 결과적으로 온톨로지에 반영한다.

IV. 결 론

해당 도메인 전문가의 개입이 없고, 구축 시간을 절감할 수 있는 자동 온톨로지 구축 방법과 지속적 자기 개선 방법에 대하여 제안하였다.

해당 방법이 정확한 온톨로지를 구축할 수 있는지 평가할 수 있는 방법이 필요하고, 정확도를 개선할 수 있는 방법과 실제 구현에 대한 연구가 앞으로 필요하다.

참고문헌

[1] Swartout, W., Tate, A., “Ontologies”, Intelligent Systems and their Applications, IEEE Vol.14, Issue.1, (2002)
 [2] van der Vet, P.E., Mars, N.J.I., “Bottom-up construction of ontologies”, Knowledge and Data Engineering, IEEE Transactions on Vol.10, Issue.4, (2002)
 [3] Alexander, Francis J., “Machine Learning [Guest editor’s introduction]”, Computing in

Science & Engineering, Vol.15, Issue.5, (2013)
 [4] A A Krizhanovsky, A V Smirnov, “An approach to automated construction of a general-purpose lexical ontology based on Wiktionary”, Vol.52, Issue.5, (2013)
 [5] Vikas Kumar, Ankur Singh Bist, “CLUSTER ANALYSIS: A SURVEY”, International Journal of Engineering Sciences & Research Technology, Vol.4, Issue.3, (2015)