

텍스트 내용 지식 기반의 철학 온톨로지 구축 (Implementation a Philosophy Ontology based on Knowledge of Text Contents)

김정민^{*} 최병일^{**} 김형주^{***}

(Jung-Min Kim) (Byoung-Il Choi) (Hyoung-Joo Kim)

요약 시맨틱 웹이나 지식기반 시스템에서 지식 검색의 핵심 컴포넌트는 온톨로지이다. 하지만 현재 실용적인 온톨로지가 구축된 경우를 찾아보기는 어렵다. 실용적인 온톨로지 구축을 위해서는 특정 도메인의 지식 전문가에 의한 지식의 개념화와 온톨로지 설계자에 의한 지식의 명세화가 필요하며 이를 바탕으로 개발 팀원들 사이에 공유할 수 있는 자세한 온톨로지 생성 가이드라인이 필요하다. 그러나 기존의 온톨로지 생성 방법론은 온톨로지 구축 프로세스의 단계들을 상위 수준에서 나열하는 절차 중심의 방법론이다. 본 논문에서는 철학 학문의 지식을 구조화하는 실용적인 온톨로지를 구축하는 방법으로서 텍스트 내용에 들어있는 지식을 온톨로지로 생성하는 토픽맵 기반의 텍스트 온톨로지 방법론을 제시하였다. 이 방법론은 지식 자원으로부터 지식을 분석하고 추출하는 온톨로지 모델링 단계와 형식 언어와 데이터 모델을 이용하여 온톨로지를 생성하는 온톨로지 구현 단계로 구성된다. 방법론과 함께 본 논문에서는 철학 온톨로지 에 기반한 철학 지식 검색의 효율성을 보이기 위해 철학 지식 포털을 구현하였다.

키워드 : 온톨로지 구축 방법론, 철학 온톨로지, 토픽맵

Abstract Ontologies are the core components of the Semantic Web and knowledge-based systems. But it is difficult to find useful ontologies in actual domains. In order to build useful domain ontology, the conceptualization of the domain knowledge by knowledge experts of the specific domain and the specification of conceptualized knowledge with formal languages by ontology designers are required. In addition, structured and detailed guidelines and methods should be provided to be shared by the development team members. However, existing ontology building methodologies define and describe the skeletal structure of the whole building process at the top-layer. We build a useful academic ontology that is based on the conceptual knowledge structure in the domain of philosophy, and propose a detailed methodology to build a text ontology based on Topic Maps. Our methodology consists of two phases, ontology modelling and ontology implementation. We implement a philosophy knowledge portal to support retrieving and navigating of the philosophy knowledge.

Key words : Ontology building methodology, Philosophy ontology, Topic Maps

1. 서론

철학을 비롯한 대부분 인문학의 연구방법은 텍스트의 내용에 기반한 지식의 축적 및 재구성, 그리고 새로운 해석의 유도가 중요하다. 인문학 연구자들은 자신의 연

구주제와 관련된 텍스트의 내용 지식을 습득하기 위해 관련 문헌을 검색한 다음 그 내용을 이해하는 데에 많은 시간과 노력을 기울인다. 연구자들은 문헌을 검색하기 위해 검색 엔진, 전자 도서관, 학술문헌정보 포털 등을 이용함으로써 과거에 비해 짧은 시간에 관련 문헌 및 정보를 찾을 수 있다. 그러나 검색의 결과가 대부분 서지 정보만을 제공하기 때문에 텍스트의 내용 지식 습득은 여전히 어려운 문제이다.

또 하나의 어려움은 텍스트의 내용 지식 표현의 문제이다. 전자 도서관이나 학술 데이터베이스가 제공하는 대부분의 원문의 형태가 HTML 또는 PDF 문서인데, 이는 매체만 달라졌을 뿐 지식 표현에 있어서 책과 동일한 다음과 같은 문제점들을 가진다. 첫째는 연구자들

· 본 연구는 BK21과 ITRC 및 학술진흥재단 기초과학연구 과제에서 지원받았음

* 비회원 : 서울대학교 컴퓨터공학부
jmkim@oopsla.snu.ac.kr

** 비회원 : 서울대학교 철학사상연구소 팀장
tarski@snu.ac.kr

*** 종신회원 : 서울대학교 컴퓨터공학부 교수
hjk@oopsla.snu.ac.kr

논문접수 : 2004년 8월 26일

심사완료 : 2005년 1월 19일

이 원하는 내용만 부분적으로 발췌하고자 할 경우에도 원문을 출력해서 전체를 읽어야 하는 경우가 많다. 둘째는 문헌의 주요 개념들과 각 개념의 서술 구조를 파악하기 어렵다. 셋째는 문헌들 사이에 있어서 내용의 동일, 대립, 보완 등 여러 가지 연관성에 대해 각각의 문헌을 읽고 비교하기 전에는 파악할 수 없다. 넷째는 문헌내의 특정 용어 또는 개념에 대해 관련된 인터넷 상의 정보가 연결되어 있지 않다.

이러한 문제점들은 현재의 문헌 정보화 서비스가 도서관 서고에 보관되는 출력물을 단순히 컴퓨터에 저장 가능한 전자문서로 변환하여 데이터베이스화해 놓은 것에 불과하기 때문이다. 위와 같은 문제점들을 해결하기 위해서는 학문별 텍스트 온톨로지(text ontologies)의 구축과 함께 텍스트 온톨로지에 기초한 학문별 지식 포털을 구축하는 것이 필요하다.

본 논문에서는 토픽맵 기반의 텍스트 온톨로지 구축 방법론을 제시하고 구체적인 철학 분야의 텍스트 온톨로지의 구축을 통한 철학 지식의 분류 및 검색의 효율성을 보이는 철학 지식 포털을 소개한다. 온톨로지는 특정 도메인의 지식을 표준적이고 공유할 수 있는 용어들로 정의하고 용어들 사이에 계층 및 연관 관계를 설정함으로써 그 도메인의 주요 개념들을 형상화해 놓은 것이다[1]. 대부분의 철학 텍스트들은 시대적으로, 지역적으로, 그리고 분야별로 광범위하고 복잡한 개념 및 지식 구조를 가진다. 또한 철학자나 철학 학파에서 주장하는 개념에 대해 여러 가지 다양한 해석, 반박, 지지 등의 서술이 존재하며 초기 개념에서 파생되어 보다 세분화된 개념들과 이들 상호간에 의미적으로 중요한 연관성이 존재한다. 그러므로 복잡한 개념 구조의 철학 텍스트의 내용 지식을 온톨로지 모 델링하는 것은 매우 적합한 접근방법으로서 본 연구의 결과는 철학외의 역사, 언어 및 다른 모든 학문 분야에도 적용이 가능하다.

대부분의 중요한 철학 지식은 철학 고전 텍스트에 존재하므로 본 논문에서는 이러한 철학 지식을 온톨로지 모 델링하기 위해 문헌 분석에서부터 지식 추출, 용어 정의 및 계층 구조와 연관 구조를 설정하는 과정에 대한 효과적인 온톨로지 구축 방법론을 정의하였다.

또한 확장성을 가지는 대용량 온톨로지를 구축하기 위해 컴포넌트 기반의 분산 토픽맵(Topic Maps)[2,3] 구조를 제안하였으며 온톨로지 관리 시스템 K-Box[4]를 개발하여 철학 온톨로지 기반의 철학 지식 포털을 구축함으로써 효율적인 지식 관리와 검색이 가능함을 보인다. 본 연구의 철학 온톨로지와 이를 기반으로 철학 지식을 서비스하는 철학 지식 포털을 철학온톨로지(PHILosophy MAP)이라 하였다.

K-Box 시스템은 온톨로지의 데이터모델로 토픽맵을

사용하기 때문에 토픽맵을 생성하고 변경, 저장하며 키워드 검색 및 주제 검색등을 지원하기 위한 여러 컴포넌트들로 구성된다. K-Box의 컴포넌트들은 토픽맵 객체들의 단일 인터페이스 제공을 위한 토픽맵 오브젝트 래퍼(Topic Map Object Wrappers), 토픽맵 객체들을 저장하기 위한 스토리지 래퍼(Storage Wrappers), 사용자가 토픽맵에 접근할 수 있게 하는 토픽맵 제공자(Topic Map Provider), 대용량의 토픽맵의 효율적인 검색을 위한 토픽맵 캐쉬 관리자(Topic Map Cache Manager), 토픽맵 객체들을 생성하는 토픽맵 생성자(Topic Map Factory), 토픽맵 관리를 위한 토픽맵 관리자(Topic Map Manager), 토픽맵의 가져오기 및 내보내기등 여러 유용한 기능들을 제공하는 토픽맵 도구(Topic Map Utilities)등이 있다.

K-Box 시스템은 클라이언트-서버 구조로 사용자의 토픽맵 접근을 지원하는 프론트-엔드(front-end)와 토픽맵 엔진 역할의 백-엔드(back-end)로 나누어지며 K-Box API를 제공함으로써 지식 관리 시스템, 콘텐츠 관리 시스템 등의 내부 온톨로지 관리 모듈로 활용될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 다른 분야의 온톨로지들을 살펴보고 3장에서는 철학 텍스트 분석에서부터 온톨로지 생성까지의 구축 방법론을 자세히 서술한다. 4장은 시스템 내에 토픽맵을 기반으로 철학 온톨로지를 구축하기 위한 분산 토픽맵 구조를 소개하고 5장은 K-Box 시스템을 이용한 철학 온톨로지 기반의 지식 검색의 효율성을 보인다. 마지막 6장 결론에서는 철학 온톨로지의 의의와 다른 학문 분야로의 파급 효과 그리고 향후 연구 계획에 대해 서술한다.

2. 관련연구

온톨로지를 효율적으로 구축하기 위한 연구로서 온톨로지 공학적 측면의 방법론이 연구되어 왔으며 그 접근법은 크게 순공학(Forward engineering)과 역공학(Reverse engineering)으로 나누어진다.

순공학적 접근법은 전문가에 의한 지식 분석 및 추출로부터 온톨로지를 구축하고 온톨로지와 지식 자원을 연결하는 절차로서 대표적 온톨로지 구축 방법론으로 엔터프라이즈 온톨로지 방법론(Enterprise Methodology)[5], TOVE(TORonto Virtual Enterprise) 방법론[6], METHONTOLOGY[7] 방법론, On-To-Knowledge[1] 및 Development 101[8]이 있다. 이들 방법론은 대부분 기업, 화학공정 등 특정 도메인에서의 온톨로지 구축 경험을 기반으로 작성된 방법론으로서 상위 수준의 관점에서 온톨로지 생성 절차를 단계별로 정의하고 있다[9].

엔터프라이즈 온톨로지 방법론은 온톨로지 생성 절차를 4단계로 정리하고 있는데, 목적 설정, 온톨로지 구축, 평가, 문서화로 제시하고 있다. TOVE 방법론은 competency question을 정의한 다음 온톨로지를 통하여 질문에 대한 답을 구할 수 있는지에 따라 온톨로지를 평가한다는 점에서 다른 방법론과 차별성을 가진다. 본 논문의 철학온톨로지 온톨로지도 철학 전문가에 의해 competency question을 미리 정의한 다음 철학 온톨로지의 정당성을 평가하기 위해 사용하였다. METHONTOLOGY 방법론은 소프트웨어 개발 생명 주기를 기반으로 온톨로지 생성 단계를 정의하였으며 On-To-Knowledge는 지식 관리 프로세스와 지식 메타 프로세스 기반으로 5개의 주요 단계와 13개의 세부 과정으로 이루어진 온톨로지 방법론을 정의하였다. Development 101은 클래스와 속성을 분류하는 방법, 클래스들의 계층구조를 정의하는 방법 등 앞서 언급한 방법론들에 비해 보다 구체적인 세부 프로세스와 각 프로세스에서 수행해야 할 과제에 대해 언급하고 있다.

이들 방법론이 상위 수준에서 온톨로지 생성 절차를 나열한 반면 철학온톨로지는 지식 추출과 지식 구조 정의를 위한 세부 과정과 주요 산출물을 구체적으로 정의하였다는 것에 있다. 또한 철학 전반의 공통 지식과 함께 철학 텍스트의 내용적 지식에 대한 온톨로지를 구축함으로써 학술 연구를 지원하는 텍스트 지식 온톨로지를 체계적이고 효율적으로 생성할 수 있는 지침을 제시하였다.

역공학적 접근법[10,11]은 구조적, 반구조적 및 자유형식의 지식자원으로부터 자연어처리에 의하여 명사형 및 동사형 용어들을 추출한 다음 기계 학습 알고리즘으로 온톨로지를 구축하는 반자동적인 절차이다. 그러나 역공학적 접근법을 적용할 수 있는 도메인은 지식자원의 형식이 온톨로지화 변환하기에 적합한 구조를 가지거나 아니면 상품정보, 용어사전 등과 같이 명사형 용어들을 쉽게 추출할 수 있으며 의미적으로 명확한 경우이다.

본 연구와 관련된 철학 디지털 지식 자원들의 형태를 살펴보면 크게 다음의 4가지 유형으로 분류할 수 있다. 철학 디지털 텍스트 사이트[12]는 주로 철학 텍스트들의 원전 또는 번역본을 직접 또는 하이퍼링크를 통해 제공하는 웹사이트들이다. 이와 같은 웹사이트들은 철학 연구자들이 1차 문헌을 웹을 통해 직접 읽거나 인용할 수 있도록 하기 위한 것이다.

철학 디지털 저널 사이트[13]는 주로 철학 논문들을 직접 또는 하이퍼링크를 통해 제공하는 웹사이트들이다. 이와 같은 웹사이트들은 철학자들의 연구 활동의 핵심이 되는 논문 출판과 논문 인용을 웹에서 이루어 질 수 있도록 지원하기 위한 것이다.

철학 디지털 참고문헌 사이트[14]는 주로 철학 디지털 사전 또는 철학 디지털 백과사전을 제공하는 웹사이트들이다. 이와 같은 웹사이트들은 철학 연구자들이 주로 사용하거나 참고하는 용어 및 철학자, 철학 문헌, 철학 개념 등에 대한 지식을 웹에서 찾을 수 있도록 하기 위한 것이다.

철학 메타 사이트[15]들은 철학 관련 사이트들에 대한 정보 및 디렉토리를 제공하는 웹사이트들이다. 이와 같은 웹사이트들은 철학 텍스트, 철학 저널, 철학 단체, 철학자 개인 뿐 아니라 철학 토픽, 철학 분야 등에 대한 철학 지식을 제공하는 철학 웹사이트들에 대한 메타 정보를 제공하기 위한 것이다.

본 연구의 철학 온톨로지는 원문 제공을 목적으로 하는 것이 아니라 원문의 내용을 분석하여 중요한 개념들을 추출하고 각 개념들에 적합한 용어를 정의하며 이들 용어사이에 연관 관계를 정의하는 의미적 지식망을 구축하는 것이다.

3. 철학 온톨로지 생성 방법론

철학 온톨로지는 크게 철학 분야의 포괄적 지식과 철학 텍스트들의 내용적 지식으로 구성된다. 철학 분야의 포괄적 지식은 철학자, 철학문헌, 철학사, 철학분야, 철학이론, 철학학과, 철학용어를 중심으로 각 주제의 세부 지식들로 이루어져 있으며 철학 텍스트의 내용적 지식은 철학 고전 텍스트를 분석하여 주요 개념들을 추출하고 그 개념을 설명하기 위한 상세 개념들을 파악하여 계층 및 연관 관계에 따라 개념들을 연결한 지식 구조로 이루어져 있다.

다른 온톨로지들과 비교하여 철학 온톨로지가 가지는 의의는 철학 텍스트 속에 들어있는 지식을 텍스트 외부에 형상화함으로써 텍스트를 읽지 않더라도 주요 개념들을 이해하고 나아가 다른 철학 지식들과의 계층적 및 의미적 연관성에 따라 연결된 지식 구조를 끊임없이 탐색이 가능하다는 것이다.

본 장에서는 철학 온톨로지의 생성 과정을 소개하고 각 과정에서의 중요한 고려 사항들에 대해 서술한다. 특히 온톨로지 구축 측면에서 철학 텍스트를 분석하고 개념을 추출하는 방법에 대해 자세하게 정의한다.

철학 온톨로지 생성 과정은 그림 1에서 볼 수 있듯이 크게 철학 분야의 포괄적 지식을 정의하는 과정과 철학 텍스트의 지식을 정의하는 과정, 그리고 토픽맵을 기반으로 철학 온톨로지를 구현하는 과정으로 나누어진다.

목적 및 범위 설정은 온톨로지를 구축하는 이유와 사용 목적, 그리고 구축 범위를 결정하는 과정으로 온톨로지에 포함되는 지식의 수준을 정하는 기준을 제시한다. 본 연구의 철학 온톨로지는 동양과 서양을 포함하여 고

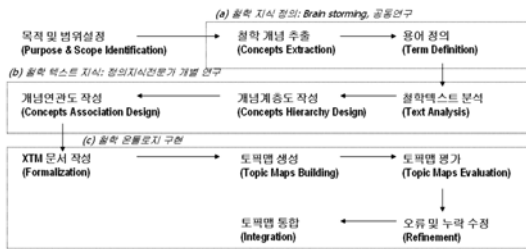


그림 1 철학 온톨로지 생성 프로세스

대에서부터 현대에 이르기까지 모든 철학 영역을 대상으로 하고 있으며 주요 철학 고전 300 여권의 철학 텍스트 지식을 온톨로지로 구축할 계획이다. 철학 온톨로지의 목적은 철학 연구자들에게 철학 텍스트 원문 및 인터넷에서 접근 가능한 철학 자료를 제공하는 것과 철학 지식 지도(Philosophy Knowledge Map)를 통한 지식 검색의 효율성을 높이는 것이다.

3.1 상위 수준의 철학 온톨로지 정의

온톨로지를 생성하기 위해 특정 도메인의 지식을 분석하고 추출하는 기법으로 하향식 기법(top-down approach), 상향식 기법(bottom-up approach), 그리고 상하향식 기법(middle-out approach)[9]이 있다. 이 기법들 중에서 최상위 수준의 개념으로부터 출발하여 상세 개념들로 분류해 나가는 기법이 하향식 기법이며 철학 온톨로지 에서 철학 전반의 포괄적 지식 구조를 생성하기 위해 사용된다.

철학 온톨로지의 최상위 개념은 철학(philosophy)이다. 철학이라는 개념은 몇 문장으로 정의할 수 없는 개념이므로 반드시 그 개념과 'is-a' 또는 'part-of' 관계를 가지는 하위의 상세 개념들을 분류해야 된다. 이에 따라 철학 개념과 'part-of' 관계를 가지는 것으로 철학자, 철학문헌, 철학분야, 철학사, 철학학과, 철학이론, 철학용어의 7가지 세부 개념들을 분류하였다. 그리고 각각의 세부 개념들도 보다 상세한 세부 개념으로 분류하였다. 철학자의 경우 지역적 및 시대적 기준에 따라 철학자와 'is-a' 관계를 가지는 한국철학자, 중국철학자, 인도철학자, 서양고대철학자, 서양중세철학자, 서양근대철학자, 서양현대철학자로 분류하였으며 철학문헌, 철학사 등 다른 세부 개념들도 이와 동일한 기준으로 분류하였다.

온톨로지에서 하나의 개념은 클래스와 동일하게 자신을 설명하는 속성을 가진다. 개념의 속성은 내적 속성(internal attribute)와 외적 속성(external attribute)으로 구분되는데, 내적 속성은 URI 주소로 접근할 수 없는 지식 자원으로서 온톨로지 내부에 정의되는 텍스트 데이터, 즉 리터럴(literal)이다. 외적 속성은 URI 주소로 접근이 가능한 지식 자원으로서 인터넷 또는 인터넷

철학온톨로지 CID 생성 규칙	
1. 명문 소문자뿐만 작성한다.	
2. 반관, 특수기호는 사용하지 않는다.	
3. 복합명사의 명사와 명사 사이의 구분은 언더바(_)를 사용한다.	
4. 한글(ko), 중국(ch), 인도(in), 서양고대(anc), 서양중세(mod), 서양근대(mod), 서양현대(con) 등의 키워드를 사용한다.	
5. 철학자(philosophers), 철학문헌(phil_texts), 철학사(phil_history), 철학분야(phil_branch), 철학학위(phil_schools), 철학이론(phil_doctrines), 철학용어(phil_terms) 등의 키워드를 사용한다.	
6. 개별 철학문헌의 경우 철학자_학위키워드 형식으로 생성한다.	
예) 칸트 실천이성어판: kant_practical	

그림 2 철학온톨로지의 CID 생성 규칙의 일부

넷에 존재하는 웹페이지, 파일, 데이터베이스 등 모든 자원들과의 연결을 정의한다.

온톨로지에서 개념들 사이의 관계는 위에서 언급한 'is-a'와 'part-of'의 수직적 계층 관계 외에 의미정보를 기반으로 연결되는 수평적 연관 관계로 나누어진다. 수평적 연관 관계는 개념들 사이의 의미적 연관성을 표현하는 관계로서 철학 전문가가 가지는 지식에 의해 구체화된다. 예를 들어, 철학자 개념과 철학문헌 개념 사이의 연관 관계는 '저자' 또는 '주요 저서', '대표 저서' 등이 존재하고 철학자와 철학이론과의 연관 관계는 '주장이론', '반대이론' 등의 관계가 존재한다. 철학자 사이에는 '스승과 제자', '영향을 받은 철학자', '영향을 끼친 철학자' 등의 관계가 성립될 수 있다.

개념은 하나 이상의 개념명을 가진다. 개념명의 형식은 '철학', '철학자', '형이상학' 등과 같은 단일명사와 '철학문헌', '철학학과' 등의 복합명사 그리고 '도덕적 소외', '관념론적 역사관' 등의 명사구를 가진다. 상위 수준의 온톨로지에서는 대부분의 개념들이 단일명사 또는 복합명사의 개념명을 가진다. 그러나 철학 텍스트 온톨로지에서는 단일명사, 복합명사 외에 대부분 명사구 형식으로 개념명을 부여한다.

철학 온톨로지에서 하나의 개념은 하나의 CID(concept identifier)를 가진다. CID는 유일한 값으로 개념들 사이의 참조에 사용된다. 개념명으로는 철학 학문 분야의 표준 용어를 사용하지만 개념의 CID를 정의하기 위해서는 그림 2와 같은 생성 규칙을 정의하여야 한다.

철학 온톨로지 생성 프로세스의 개념 분석 및 추출과정의 산출물은 개념사전(concept dictionary)이다. 개념사전은 철학 온톨로지의 개념들의 집합이며 하나의 개념은 그림 3과 같이 정의된다.

레벨(level)은 계층 구조에서의 위치를 가리키는 것으로서 최상위 개념 철학의 레벨이 1이고 그 하위의 철학자는 철학의 하위 개념들 중의 첫 번째 개념이므로 레벨이 1.1로 표기된다. 개념 타입(concept type)은 철학 온톨로지서 개념의 성격을 가리키는 것으로 철학자는 그 하위에 다른 개념들을 포함하며 직접 지식 객체를 가리키는 것이 아니므로 클래스가 된다. 반면, 지식 객체에 대한 내적 또는 외적 참조가 있으면서 하위 개념



그림 3 K-Box 온톨로지 편집기의 개념 정의 화면

을 가지지 않는 온톨로지에서 최하위 개념들의 개념 유형은 인스턴스가 된다.

슈퍼 클래스는 철학자 개념을 내포하는 상위 개념의 CID를 가리키고 서브 클래스는 철학자가 내포하는 하위의 개념들의 CID를 가리킨다. 이때 슈퍼 클래스와 서브 클래스 값은 0개 이상 여러 개가 존재한다. 속성은 철학자 개념을 설명하는 항목으로서 0개 이상 여러 개가 존재하며 속성명, 내적/외적 참조, 다중성, 속성 타입을 가진다.

관계성(relationships)은 철학자 개념과 다른 개념들과의 사이에 존재하는 연관 관계에 대해 정의하는 것으로 관계명(name), 기준 개념(origin), 대상 개념(destination), 기준 개념의 역할(ori_role), 대상 개념의 역할(dest_role)을 가진다. 이때 기준 개념과 대상 개념은 모두 CID로 표시되며 속성 타입과 역할도 하나의 개념으로 정의되므로 각각의 CID로 표시된다.

3.2 철학 텍스트 온톨로지 정의

철학 온톨로지의 상위 수준의 온톨로지는 철학 전반의 포괄적인 지식을 정의하고 분류하는 온톨로지인 반면 철학 텍스트 온톨로지는 텍스트 내의 내용상에 존재하는 지식을 정의하고 분류하는 내용 기반 온톨로지이다. 그림 4는 철학 온톨로지의 두 계층 사이의 연결을 보이고 있다. 칸트의 실천이성비판 문헌을 검색하고자 할 경우 최상위 철학에서부터 철학문헌, 서양근대철학문헌으로 탐색하면 실천이성비판 개념 노드를 찾게 된다.

실천이성비판 텍스트의 하위 개념들은 그 텍스트의 내용상에 존재하는 주요 개념들로서 이성, 자유, 도덕법칙, 의지 등이 있다. 한 텍스트에는 많은 개념들이 추출될 수 있으므로 이 개념들의 CID는 각각 c1, c2, c3, c4로 부여되며 이성 개념은 그 하위에 이성의 정의와 이성의 구분이라는 개념들로 세분화된다. 이때 세부 개념



그림 4 철학 온톨로지의 계층간 연결구조

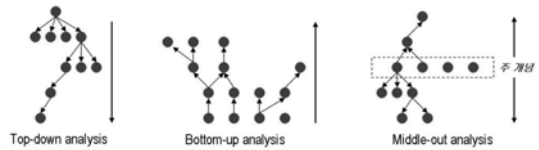


그림 5 철학 텍스트의 내용상의 지식 분석 및 추출 기법

의 CID는 c1.1과 c1.2로 부여된다.

철학 텍스트 온톨로지를 구축하기 위해 먼저 각 철학 분야의 철학전문가들이 대상 철학 고전 텍스트를 선정한 다음 텍스트의 내용 분석을 하였다. 내용상의 지식 추출을 위한 텍스트 내용 분석 기법은 상하향식 기법을 사용하였다. 텍스트 분석을 위해 하향식 기법을 사용할 경우의 문제점은 추출된 개념 집합 속에 텍스트 내용상의 개념보다 상위 수준의 일반적인 개념들이 더 많은 부분을 차지할 수 있다는 것이다.

그리고 상향식 기법을 사용할 경우의 문제점은 최하위 인스턴스 타입의 개념에서 상위의 클래스 타입의 개념들을 추출하여 올라갈 경우 상위의 개념들이 너무 넓게 파악될 수 있다는 문제점을 가진다. 이에 반해 상하향식 기법은 텍스트의 주요 개념에서 출발한다. 모든 철학 텍스트는 저자가 강조하고 설명하는 주요 개념들이 존재하므로 이 개념들을 시작점으로 두고 아래로는 하향식 기법을 적용하고 위로는 상향식 기법을 적용하여 계층 관계의 개념들을 추출한다. 그림 5는 이 세 가지 기법의 처리 형식을 보여준다.

텍스트에서 추출되는 개념의 이름은 단일명사, 복합명사, 명사구 형태를 가진다. 이 경우 개념명이 명사구 형식일 경우에는 개념명을 아래 규칙에 따라 분석함으로써 새로운 개념 및 계층 구조를 파악할 수 있다. 본 논문에서는 대표적인 명사구 유형 4가지를 제시하고 각각의 경우에 어떤 규칙으로 분석이 가능한지 보인다.

규칙 1. 개념명 형식이 'A적 B(A-al B)'인 경우

규칙 1의 예시는 '법적 관점', '자연주의적 오류', '도덕

적 소외', '관념론적 역사관' 등으로 볼 수 있으며 이 경우 B가 문서내의 중요 개념일 경우 반드시 B를 상위 개념으로 분류하고 그 하위에 'A적 B' 개념을 두어야 한다. 예를 들어, '역사관' 개념을 별도로 분류하고 그 하위에 '관념론적 역사관'을 연결하여야 한다.

그러나 '자연주의적 오류'와 같이 오류 자체가 문서 내에서 중요하게 서술되는 개념이 아닐 경우 오류를 별도로 분류하지 않고 '자연주의' 하위에 '자연주의적 오류'로만 분류한다.

규칙 2. 개념명 형식이 'A의 B(B of A)'인 경우

규칙 2의 예시는 '물질의 증여', '도덕법칙의 원천', '행위의 자유', '의지의 자유' 등으로 볼 수 있으며 이 경우 A가 중요한 개념이라면 A를 상위 개념으로 분류하고 그 하위에 'A의 B'를 두며 반대로 B가 중요한 개념이라면 B를 상위 개념으로 분류하고 그 하위에 'A의 B'를 둔다. 만일 A와 B가 각각 중심 개념으로 내용 서술을 가진다면 A와 B를 상위 개념으로 두고 그 하위에 'A의 B'를 둔다. 이때 'A의 B'는 클래스의 다중상속처럼 여러 상위 개념을 가진다.

규칙 3. 개념명 형식이 'A와(과) B', 'A와(과)B의 C', 'A의 B와(과) C'인 경우

규칙 3의 예시는 '기도와 책임', '선과 양심', '자유와 초탈의 복음', '의지의 주관성과 객관성' 등으로 볼 수 있으며 'A와(과) B'의 경우 A와 B를 각각 상위 개념으로 분류하고 그 하위에 A와 B의 관계에 대해서 설명하는 자세히 분류된 여러 개의 하위 개념들을 생성하여 계층 구조로 연결한다. 'A와(과) B의 C'의 경우에는 C를 상위 개념으로 분류하고 그 하위 개념으로 'A의 C'와 'B의 C'를 둔다. 예를 들어, '자유와 초탈의 복음'의 경우 '복음'이라는 넓은 개념 속에 '자유의 복음'과 '초탈의 복음'이라는 좁은 개념들이 존재하는 것으로 정의할 수 있다. 또한 'A의 C'와 'B의 C' 사이에 의미적 관계 설정을 위해 연관 관계를 정의한다. 'A의 B와(과)C'의 경우에는 A를 상위 개념으로 분류하고 그 하위 개념으로 'A의 B'와 'A의 C'를 연결한다. 이때 'A의 B'와 'A의 C' 사이에 의미적 관계 설정을 위해 연관 관계를 정의한다.

규칙 4. 개념명 형식이 'A(으)로서의 B'인 경우

규칙 4의 예시는 '법의 지반으로서의 자유의지', '자유로서의 법' 등으로 볼 수 있으며 이 경우 B를 상위 개념으로 분류하고 그 하위 개념으로 'A(으)로서의 B'를 연결한다.

이들 규칙에서의 원칙은 조사에 의해 제한이 되는 용어를 상위 개념으로 분류하고 하위 개념과의 사이에 계층 구조를 정의한다는 것이다. 또한 대등관계의 용어들 사이에는 어떠한 의미적 연관성이 있으므로 그 연관성

해결 "법철학" 개념 계층도		
CID	Concepts Hierarchy	Resource ID
c1	의지	
c11	의지의 세계기	P1.1.2.1
c12	의지의 세형태	P1.1.3.1
c121	자연적 의지	P1.1.3.3
c122	자의	P1.1.3.5
c123	축자 대자적 의지	P1.1.3.8
c13	의지의 주관성	P1.1.4.1
c131	의지의 주관성의 의미	P1.1.4.3
c14	의지의 객관성	P1.1.5.1
c141	의지의 객관성의 의미	P1.1.5.2
c2	자유	
c21	자유외지	
c211	법의 지반으로서의 자유외지	P1.1.1.4
c212	자유로서의 법	P1.1.1.9
c3	소유	
c31	소유의 재규정	
c311	검유취득	P1.2.2.1.1
c3111	검유취득의 방법	P1.2.2.1.4
c31112	육체적 획득	P1.2.2.1.5
c31113	형식의 부여	P1.2.2.1.7
c31114	표지	P1.2.2.1.10

그림 6 철학 텍스트의 개념 추출 과정의 산출물인 개념 계층도

에 대해 또 다른 하위 개념들을 정의하거나 또는 연관 관계를 정의하여야 한다.

철학온톨로지에서 상위 수준의 온톨로지에 속하는 개념들에 비하여 철학 텍스트 온톨로지의 개념들은 비교적 간단한 속성을 가진다. 철학 텍스트 온톨로지의 개념들은 텍스트 내용에 의해 추출된 개념이므로 그 개념이 가지는 지식은 텍스트의 특정 문단(paragraph)들에 존재한다. 그러므로 철학온톨로지에서 철학 텍스트 온톨로지의 개념 속성은 해설(explanation)과 인용(quotation)으로만 정의하였으며 해설과 인용 속성 값은 내적 데이터와 외적 링크 둘 다를 가지도록 하였다.

철학 텍스트 온톨로지는 개념 사전 대신에 그림 6과 7에서 볼 수 있는 개념계층도와 개념연관도를 작성하였다. 하나의 텍스트에서 추출되는 개념의 수가 최소 300 개 이상이므로 동일한 형식을 가지는 개념들에 대해 개념 사전을 개별적으로 작성하는 것은 비효율적이다.

그림 6의 개념계층도는 분석된 개념의 개념명과 CID, 계층 구조 그리고 개념의 철학 지식이 존재하는 텍스트의 위치를 내용으로 작성된다. 개념의 CID는 상위 수준의 개념과 달리 의미적인 이름을 부여하기 어렵기 때문에 레벨과 일련번호 중심으로 부여된다. 즉 CID 자체에 온톨로지의 계층 구조에서의 위치 정보를 가지도록 생성 규칙을 정하였다.

개념의 철학 지식의 존재 위치를 가리키는 resource ID는 문단일 경우 시작 문자를 p로 하였으며 인용문일 경우에는 시작 문자를 q로 하였다. p나 q 뒤의 일련번호는 텍스트의 구조를 가리킨다. 즉 1부, 1장, 1절, 1문단인 경우 p1.1.1.1로 명명된다. 개념에 의해 참조되는 텍스트의 범위는 최소 문단 단위로 하였으며 인용문인 경우에는 문단내의 한 인용문을 최소 단위로 하였다.

그림 7의 개념연관도는 철학 텍스트의 개념들 사이의 의미적 연관성을 보이는 산출물로서 텍스트를 분석하는

Association	Concept1	Concept2
동일관계	사치주의 이론	공산주의 이론
동일관계	역사적 유물론	유물론적 역사관
대립관계	역사적 유물론	관념론적 역사관
주장과 근거관계	철학적 생산활동의 중요성	사회적 삶의 토대
동일관계	과학주의	실증주의
규정과 피규장관계	토대	사회구조
상호작용관계	생산력	생산관계

그림 7 철학 텍스트의 개념 추출 과정의 산출물인 개념 연관도

철학전문가의 지식에 의존하여 분석된다. 텍스트 내에서 개념들 사이의 연관성은 그 자체가 하나의 중요한 사상이고 개념이 된다. 예를 들어, 칸트의 실천이성비판에서 이성과 도덕법칙의 관계는 한 단어로 설명되지 않는 그 자체가 여러 페이지에 나누어 설명되는 중요한 개념인 것이다. 이 경우 이성과 도덕법칙 사이에는 일반적 연관성을 규정하기 어려우므로 연관 관계로 표현하지 않고 두 개념 사이의 관계를 설명하는 여러 하위 개념들을 추출한 다음 개념들 사이에 계층 구조를 정의해야 한다.

4. 토픽맵 기반 철학온톨로지 구현

철학전문가에 의해 생성된 철학 온톨로지를 토픽맵으로 구현하기 위해 먼저 철학 PSI(Published Subject Indicator)와 철학 스키마를 정의하였다. 철학 PSI는 철학 온톨로지의 주요 용어에 대한 정의를 포함하는 것으로 철학, 철학자, 철학문헌, 철학이론 등의 개념에 주체성(identity)을 부여한다. PSI는 여러 개의 토픽맵들을 하나의 토픽맵으로 통합할 때 동일한 주체성에 속하는 토픽들을 그룹화하기 위해 사용된다. 예를 들어, 칸트 토픽맵과 헤겔 토픽맵, 플라톤 토픽맵을 통합할 경우 칸트 토픽과 헤겔 토픽은 서양근대철학자 PSI에 속하므로 이 두 토픽은 서양근대철학자 토픽 하위에 연결되고 플라톤 토픽은 서양고대철학자 PSI에 속하므로 서양고대철학자 토픽 하위에 연결된다.

철학 스키마는 클래스 토픽을 정의하고 토픽들 사이의 계층 구조를 정의하는 것으로 모든 인스턴스 토픽들이 공유할 수 있도록 토픽 타입, 어커런스 타입, 연관관계 타입을 정의한다. 또한 철학 스키마는 인스턴스 토픽맵의 템플릿을 제공한다. 즉, 철학자 템플릿은 칸트, 헤겔, 데카르트 등 모든 철학자가 공유할 수 있는 템플릿이고 철학문헌 템플릿은 실천이성비판, 법철학, 독일

```

Hegel.xtm
<?xml version="1.0" encoding="UTF-8" ?>
<topic id="Hegel">
  <instanceOf <topicRef xlink:href="philosophy.xtm#mod_philosophy" /> </instanceOf>
  <baseName <baseNameString>헤겔 </baseNameString> </baseName>
  <occurrence>
    <instanceOf <topicRef xlink:href="philosophy.xtm#teacher" /> </instanceOf>
    <resourceData>헤겔은 1788년에 튀빙겐 신학 대학에 신학과 학생으로 입학했다. ... </resourceData>
  </occurrence>
  ...
  <association>
    <instanceOf <topicRef xlink:href="philosophy.xtm#teacher_of" /> </instanceOf>
    <member>
      <roleSpec <topicRef xlink:href="philosophy.xtm#teacher" /> </roleSpec>
      <topicRef xlink:href="kant.stm#kant" />
    </member>
  </association>
</topic>

Kant.xtm
<?xml version="1.0" encoding="UTF-8" ?>
<topic id="Kant">
  <instanceOf <topicRef xlink:href="philosophy.xtm#mod_philosophy" /> </instanceOf>
  <baseName <baseNameString>칸트 </baseNameString> </baseName>
  ...
</topic>
    
```

그림 8 헤겔 철학자와 칸트 철학자의 정의 및 헤겔과 칸트 사이의 연관 관계의 정의

이데올로기 등 모든 철학 문헌들이 공유할 수 있는 템플릿이다.

그림 8은 XTM 구문으로 작성된 헤겔 철학자 토픽맵과 칸트 철학자 토픽맵의 일부분을 보이고 있다. 각각 헤겔과 칸트 토픽을 정의하고 있으며 헤겔과 칸트 사이에 'teacher_of'의 연관 관계가 존재함을 <association> 엘리먼트에서 정의하고 있다.

5. 철학 온톨로지 기반의 지식 서비스

철학 온톨로지를 기반으로 철학 지식의 검색 및 분류 서비스를 제공하기 위해서는 토픽맵을 생성 및 관리하는 소프트웨어가 필요하다. 본 연구에서는 자체적으로 개발한 온톨로지 관리 시스템 K-Box를 이용하여 철학 지식을 제공하는 철학 지식 포털을 개발하였다.

현재 철학 온톨로지는 철학 고전 텍스트 30권의 분석 결과에 따른 지식 구조로 생성되었다. 분석 대상 텍스트는 서양고전 텍스트 2권, 서양 중세 텍스트 2권, 서양근대 텍스트 11권, 서양현대 텍스트 10권, 동양철학 텍스트 5권이었으며 향후 지속적으로 대상을 확대해 나갈 예정이다.

표 1은 현재 철학 온톨로지의 통계 자료를 보여주는 것으로 철학 온톨로지의 주 카테고리별로 생성된 토픽맵의 수, 토픽들의 수, 그리고 연관 관계의 수를 보여준다. 철학문헌에는 철학 텍스트 자체에 대한 토픽맵과 그 텍스트의 내용 분석에 의한 지식 토픽맵이 포함된다. 연관 관계의 수가 토픽의 수보다 훨씬 큰 이유는 하나의 토픽이 다른 여러 토픽들과 연관 관계를 가지기 때문이다. 예를 들어, 철학자의 경우 철학자와 철학문헌 사이

표 1 철학온톨로지의 통계자료

	철학자	철학문헌	철학사	철학학과	철학분야	철학이론	철학용어
TopicMap	190	68	37	102	53	78	532
Topic	210	13858	42	110	59	80	538
Association	1156	4655	249	621	381	430	277

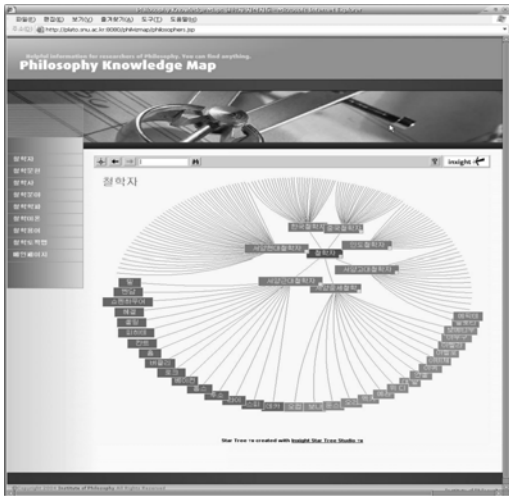


그림 9 철학 지식 서비스를 위한 지식 포털

에 ‘저자’라는 연관 관계가 존재하는데 한 철학자의 저서가 여러 권일 경우 이진(binary) 연관 관계가 여러 개 생성되기 때문이다.

그림 9는 K-Box 시스템을 토픽맵 엔진으로 두고 그 위에 JSP로 개발한 웹 어플리케이션 을 보이고 있다. 이 철학 지식 포털은 철학 온톨로지를 기반으로 철학 지식의 키워드 검색 및 그래프 네비게이션을 지원하고 토픽 타입의 계층 구조를 보임으로써 토픽들 사이의 내포관계를 쉽게 파악할 수 있도록 한다. 또한 철학 고전 텍스트의 해설서를 HTML 및 XML로 작성하여 문헌 저장소에 둬으로써 철학 온톨로지의 내용 기반 토픽맵을 통해 특정 문단이나 인용문을 쉽게 검색할 수 있도록 지원한다.

특히 Inxight사(<http://www.inxight.com>)의 startree SDK를 이용하여 트리 인터페이스를 구현함으로써 사용자는 상위 수준에서 철학 온톨로지의 개념들의 연결 구조를 파악할 수 있으며 링크를 따라감으로써 원하는 개념을 찾을 수 있다.

6. 결론 및 향후연구

본 논문에서는 철학 온톨로지의 생성을 통하여 문헌으로부터 지식을 추출하고 지식들 사이에 계층적 및 연관적 의미 관계를 정의함으로써 문헌 온톨로지를 생성하는 효율적인 방법론을 제시하였으며 이 방법론은 철학을 비롯한 모든 학문 분야의 지식 온톨로지화에 사용될 수 있으며 기업이나 조직의 지식관리에 있어서도 내용 중심의 문서 관리가 가능하다.

본 논문의 온톨로지 생성 방법론은 학문적 온톨로지를 생성하기 위한 체계적이고 구체적인 방법을 제시한

다는 점에서 기존의 방법론과 차별을 보인다. 기존의 방법론은 특정 도메인의 온톨로지 구축 경험을 정리한 방법론으로서 온톨로지 생성을 위한 상위 수준의 절차를 정의한 반면 철학온톨로지 온톨로지 생성 방법론은 텍스트의 메타 데이터와 텍스트 내용 지식의 연결 방법 및 텍스트 내용 분석 기법, 개념명 분석 규칙, 계층 관계와 연관 관계 추출 방법 등을 구체적으로 제시하였으며 컴포넌트 방식의 분산 온톨로지 구조를 제안함으로써 확장이 용이한 대용량 온톨로지 생성을 위한 프레임워크를 제시하였다.

향후 연구 방향은 온톨로지의 반자동 생성 및 사용자 행동 기반 지식 검색이다. 데이터 마이닝과 텍스트 마이닝에서는 웹 페이지나 텍스트에서 자동적으로 패턴 또는 규칙을 찾아내기 위해 연구하고 있으나 텍스트에서 지식을 추출하는 것은 자동적으로 해결하기 어려운 부분이다. 자연어 처리로 단어를 추출할 수 있으나 가공되지 않은 단어 집합이 지식을 표현하는 것이 아니므로 전문가에 의해 재가공 되어야 한다. 그러므로 지식 추출을 자동적으로 하는 것 보다는 본 논문의 방법론에 따라 전문가가 지식을 추출한 다음 추출된 지식을 온톨로지화하는 과정을 자동으로 처리하는 부분에 초점을 맞추고 있다.

사용자 행동 기반 지식 검색은 분산 온톨로지 환경에서 사용자의 요구에 적합한 온톨로지 뷰(view)를 제공하는 것이다. 예를 들어, 철학 온톨로지는 철학 전반의 지식 구조를 가지는 대용량 온톨로지이므로 철학 연구자들은 온톨로지 전체가 아닌 자신의 연구와 밀접한 특정 분야의 지식만을 검색하게 된다. 이때 연구자의 검색 패턴을 분석하여 그 연구자에게 특화된 뷰를 제공함으로써 지식 검색 및 네비게이션의 효율성을 제공하는 것이다.

참 고 문 헌

- [1] S. Staab, H.-P. Schnurr, R. Studer, and Y. Sure. "Knowledge processes and ontologies," IEEE Intelligent Systems, Special Issue on Knowledge Management, 16(1), January/February 2001.
- [2] Graham M.. "Topic Map technology - the state of the art," XML 2000 Conference & Exposition, Washington, USA, December 2000.
- [3] Michel B., Martin B. and Steve N.. ISO/IEC 13250 TopicMaps.
- [4] 김정민, 박철만, 정준원, 이한준, 민경섭, 김형주, "K-Box: 토픽맵 기반의 온톨로지 관리 시스템", 정보과학회논문지(컴퓨팅의 실제), 10(1), February 2004.
- [5] Uschold, M. and King, M. "Towards A Methodology for Building Ontologies," IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing,

Montreal, August 1995.

- [6] Gruninger, M and Fox, M.S. "Methodology for the Design and Evaluation of Ontologies," IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, August 1995.
- [7] Lopez, M.F., Gomez-Perez, A., and Sierra, J.P. "Building a Chemical Ontology Using Methontology and the Ontology Design Environment," IEEE Intelligent Systems, 14(1), January 1999.
- [8] Noy, N.F. and McGuinness, D.L. "Ontology Development 101: A Guide to Creating Your First Ontology," SMI technical report SMI-2001-0880, 2001.
- [9] Mizoguchi, R. "Tutorial on ontological engineering-Part2: Ontology development, tools and languages," New Generation Computing, OhmSha&Springer, 22(1), 2004.
- [10] S. Staab and R. Studer, "Handbook on Ontologies," Springer, pp. 133-150, 2004.
- [11] S. Staab and R. Studer, "Handbook on Ontologies," Springer, pp.173-190, 2004.
- [12] Perseus Classics Collection, http://www.perseus.tufts.edu/cache/perscoll_Greco-Roman.html
- [13] Online Papers in Philosophy, <http://opp.weatherson.net/>
- [14] Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/>
- [15] Philosophy in Cyberspace, <http://www-personal.monash.edu.au/~dey/phil/>



김형주

1982년 서울대학교 전산학과(학사). 1985년 미국 텍사스 대학교 대학원 전산학(석사). 1988년 미국 텍사스 대학교 대학원 전산학(박사). 1988년 5월~1988년 9월 미국 텍사스 대학교 POST-DOC 1988년 9월~1990년 12월 미국 조지아 공과대학 조교수. 1991년~현재 서울대학교 컴퓨터공학부 교수



김정민

1992년 홍익대학교 전자계산학과 졸업(학사). 1994년 홍익대학교 전자계산학과 졸업(석사). 2002년 서울대학교 전기, 컴퓨터공학부 박사과정 수료. 관심분야는 Semantic Web, Ontology, IR, Database



최병일

연세대학교 철학과 졸업(학사). 연세대학교 대학원 철학과 석사과정 졸업(석사) 미국 UC-Berkeley 논리학 박사과정 졸업(박사). 서울대학교 컴퓨터공학부 박사후연구원. 현재 서울대학교 철학사상연구소 철학문헌정보센터 철학지식지도팀 장관심분야는 시맨틱웹(Semantic Web), 토픽맵(Topic Maps), 온톨로지(Ontology)