

온톨로지 통합기반의 상위 온톨로지 추출에 관한 연구

이지윤, 양진혁, 정인정

고려대학교 전산학과

e-mail: {jjoon81, grjinh, chung}@korea.ac.kr

A Study on Upper Ontology Extraction Based on Ontology Integration

Ji-Yoon Lee, Jin-Hyuk Yang, In-Joeng Chung
Dept. of Computer Science, Korea University

요 약

같은 도메인 내에 서로 다른 관점을 다루고 중복되는 정보를 포함하는 온톨로지들이 증가함에 따라, 이들 사이에 재사용 및 공유를 위한 통합이 중요한 과제로 떠오르게 되었다. 또한 도메인 온톨로지로부터 상위 온톨로지를 추출하는 방법에 관한 연구도 온톨로지의 재사용과 공유를 목적으로 진행되고 있다. 본 논문에서 우리는 맵핑을 이용한 온톨로지 통합을 바탕으로 상위 온톨로지 생성 방법을 제안하고, 이를 정형화된 형태의 일반적인 목적을 가지는 온톨로지(*general-purpose ontology*)로 확장시켜 나가는 방안에 대해 살펴본다. 본 논문에서 제안된 온톨로지들 사이의 맵핑은 개념들 간의 관계를 효율적으로 만들고, 해당 도메인에 대한 일반화된 온톨로지는 추후 사용자들의 공유를 용이하게 한다.

1. 서론

차세대 지능형 웹인 시맨틱 웹을 구성하고 있는 온톨로지에 관한 연구는 현재 온톨로지의 생성 및 설계, 통합, 표현 등의 세 가지 방향으로 진행되고 있다[1]. 그 중에서도 같은 도메인 내에 서로 다른 관점을 다루고 중복되는 정보를 포함하는 온톨로지들이 증가함에 따라, 이들 사이에 재사용 및 공유를 위한 통합이 중요한 과제로 떠오르게 되었다. 이에 따라 최근 몇 년 동안 병합(*merging*)[2]이나 맵핑(*mapping*)[3], 정렬(*alignment*)[4] 등의 방법을 이용하여 온톨로지를 통합하는 방법에 대한 연구가 활발히 이루어졌다.

온톨로지의 재사용 및 공유를 위한 또 다른 연구 방향으로서 도메인 온톨로지로부터 상위 온톨로지를 유도해내는 방법이 있다. 특정 도메인에 독립적인 일반 목적을 가지는 온톨로지(*general-purpose ontology*)를 이용하여 해당 도메인으로부터 정형화된 형태의 온톨로지로서 확장시켜 나아갈 수 있다[5].

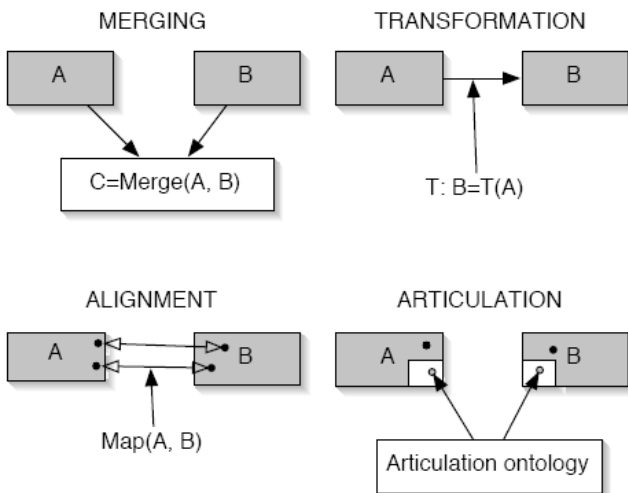
이를 통해 사용자들은 특정한 도메인에 대한 상위 온톨로지를 공유하는 동시에 확장하고 재사용할 수 있다. 그러나 같은 도메인에 대한 다양한 관점을 수용하기 어렵고, 상위 온톨로지를 기술하는데 사용되는 표현 형식과 용어들이 서로 달라서 그 중요성에도 불구하고 이용이 쉽지 않다. 또한 이들 사이에 정식 맵핑 방법이 없어 상위 온톨로지를 기술하는 방법들에 대한 비교 및 연구 분석이 어렵다.

본 논문에서 우리는 맵핑을 통한 온톨로지 통합을 기반으로 상위 온톨로지 생성 방법에 대해 제안한다. 먼저 두 개의 온톨로지를 선택한 후, 온톨로지 사이의 중재를 위해 효율적이고 비교적 정확도가 높다고 알려진 PROMPT[6]를 사용해 두 온톨로지를 병합 또는 정렬한다. 이 결과를 가지고 개념과 관계를 바탕으로 메타 레벨과 언어 레벨로 온톨로지를 분류한다. 이는 온톨로지들 사이에 맵핑을 통해 개념들 간의 관계를 효율적으로 만들고, 이와 더불어 해당 도메인에 대한 일반화된 온톨로지는 추후 사용자들의 공유를 용이하게 한다[7].

이 논문은 다음과 같은 순서로 구성되어 있다. 먼저 2장에서 온톨로지 통합에 대한 관련 연구를 살펴보고, 3장에서 상위 온톨로지를 추출하기 위한 방법론 및 알고리즘을 기술한다. 4장에서 이에 대한 실험 및 검증 과정을 거쳐, 마지막 5장에서는 결론을 도출한 후 향후 과제에 대해 살펴본다.

2. 관련연구

온톨로지 통합(integration)이란 온톨로지를 재사용하여 이용 가능한 새로운 온톨로지를 만드는 것이다. 온톨로지를 통합을 위해 사용되는 기법들로는 병합, 정렬, 맵핑 등이 있다. 병합은 같은 주제에 대한 서로 다른 온톨로지를 통합된 새로운 하나의 온톨로지로서 합치는 것이고, 정렬은 두 개 또는 그 이상의 온톨로지를 상호동의하에 가져와 일관되고 일치된 상태로 만드는 것이다. 또한 맵핑은 서로 다른 소스들로부터 동일 관계에 의해 서로 간에 유사한 개념이나 관계를 연관시키는 것이다. (그림 1)[6]은 이러한 기법들을 나타낸 것이다.



(그림 1) 다양한 온톨로지 통합 기법

2.1 온톨로지 맵핑 도구들

먼저 두 개의 온톨로지를 합병하여 새로운 하나의 온톨로지를 만드는 도구로서 개념 구조와 그들 사이의 관계로부터 정보를 사용하는 IPROMPT[6]가 있다. 또한 Ontolingua[8] 온톨로지 에디터를 바탕으로 하는 병합 도구로서 사용자가 서로 다른 형식으로 만들어진 온톨로지를 사용하는 Chimaera[9]와 두 개의 온톨로지를 공통의 문법 표현으로 변환한 후 axiom을 정의하여 인스턴스 해석 과정을 거치는

OntoMerge[10]가 있다.

그리고 연관된 개념들을 찾음으로써 두 온톨로지의 개념들 사이에 맵핑을 정의하는 도구들도 있다. 그래프 표현의 온톨로지를 분석하여 소스 온톨로지 에서 유사한 개념들의 쌍을 합병하여 새로운 쌍을 생성하는 ANCHORPROMPT[6], 맵핑을 찾기 위해 기계학습 기법을 적용한 GLUE[11], 기술 로직 (Description Logic)을 사용하여 맵핑에 대한 질의에 답하는 OBSERVER[12] 등이 그것이다.

2.2 스키마 통합

스키마 통합에 대한 스키마 매칭은 온톨로지 맵핑과 유사하다. 두 가지 모두 관련된 개념을 찾는 것이 필요하다. 동시에, 대부분의 스키마 통합 기법은 관계형 데이터베이스 스키마나 개념 트리를 포함하는 스키마를 고려한다. ICOM[13]은 스키마 통합 방법과 온톨로지 통합 도메인을 합쳐놓은 프로젝트의 한 예로 개념적 모델을 제시하고 스키마 통합을 위한 추론 메커니즘을 제공한다.

3. 상위 온톨로지 추출 방법론

우리는 두 개의 도메인 온톨로지를 바탕으로 상위 온톨로지를 추출하기 위한 방법에 대해 제안한다.

Step.1 같은 도메인에서 두 개의 온톨로지 선택

같은 도메인내의 온톨로지를 사용해야 하는 까닭은 두 개의 온톨로지를 비교하는 위해서는 비교대상이 동일한 범위 내에 있어야 하기 때문이다. 현재 기술로는 여러 개의 온톨로지를 동시에 비교 분석하는 것이 불가능하기 때문에 두 개의 온톨로지를 선택한다.

Step.2 PROMPT를 플러그-인 한 Protege를 사용하여 두 개의 온톨로지 병합

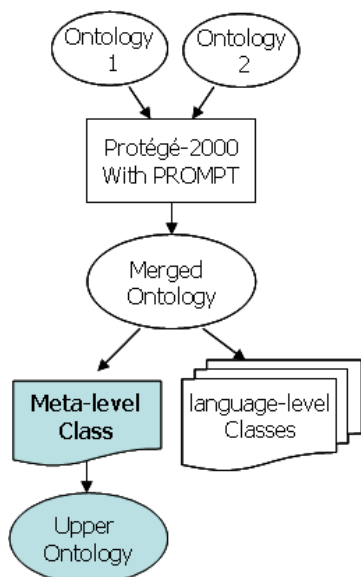
우리는 선택한 두 개의 온톨로지를 비교하기 위해 PROMPT 툴을 사용하였다. PROMPT는 다른 온톨로지 통합 툴들과 비교하여 단순하고 정확하며, 사용자가 사용하기에 편리한 인터페이스를 가지고 있다. 또한 통합을 위한 기법으로 병합을 선택한 까닭은 두 온톨로지 사이에 단순한 비교 및 관련 개념을 연결하는데 그치지 않고, 새로운 온톨로지를 생성하고, 생성된 온톨로지를 바탕으로 확장이 가능하기 때문이다.

Step.3 결과 온톨로지의 계층적 클래스 표현

PROMPT를 통해 나온 결과는 계층적 트리 형태의 클래스이다. 내부의 슬롯과 인스턴스들은 외부에서 보이지 않고 클래스들로만 표현된 형태인데, 클래스 이름들을 하나의 개념으로 보고 내부 속성을 배제한 채 RDFS의 메타데이터[14]와 같이 상위 클래스들을 추출하여 상위 온톨로지를 구성한다.

Step.4 상위 온톨로지 표현

해당하는 상위 클래스를 RDFS의 표현방법과 같이 방향이 있는 그래프 형태로 나타낸다. 위에서 제시한 방법을 순서도로 나타내면 (그림 2)와 같다.



(그림 2) 상위 온톨로지 추출 순서도

우리는 다음 장에서 앞에서 제시한 방법론을 바탕으로 간단한 예를 들고 이를 테스트한다.

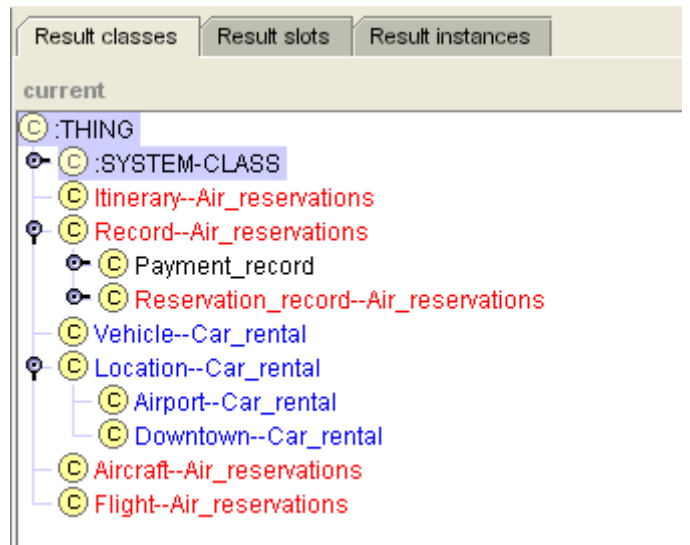
4. 실험 및 검증

위에서 제안한 방법을 실험하기 위해 먼저 윈도우 XP에서 Protege-2000의 가장 최신 버전인 Protege 3.1.1을 다운로드 받아 실행한다[15]. 여기에 PROMPT를 plug-in[16]하여 맵핑 및 정렬을 위한 적절한 환경을 만든다. 실험에서 사용된 온톨로지로는 PROMPT에서 제공하는 테스트 온톨로지인 항공기예약 온톨로지(Air_reservation)와 자동차대여 온톨로지(Car_rental)를 사용한다[16]. 본 논문에서 사용되는 PROMPT의 주요 기능들은 <표 1>과 같다.

<표 1> PROMPT의 주요 기능

기능	설명
Merging Class	같은 이름의 클래스 합병
Copy Class	클래스 복사
Merging Slot	슬롯의 합병
Removing Conflicts	충돌의 제거
Removing Suggestion	불필요한 암시 제거
Removing suffixes	접미사 제거

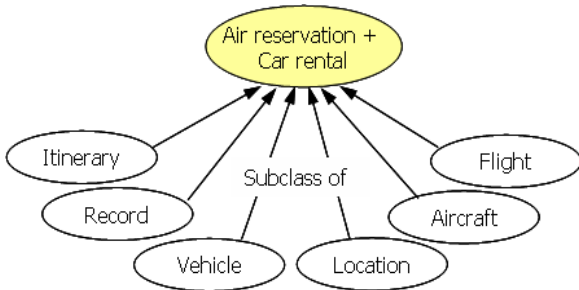
먼저 Protege에서 병합 작업을 선택한 후 해당하는 두 개의 온톨로지를 로드하면 작업창에는 병합해야 하는 두 온톨로지의 클래스 목록을 사용자에게 보여준다. 그 다음 같은 이름을 가진 클래스를 선택하여 병합한다. 병합한 결과는 클래스 이름 옆에 붙은 'm'을 통해서 확인할 수 있다. 병합 과정이 진행될수록 'm' 표시는 증가한다. <표 1>의 주요 기능 외에 PROMPT의 다양한 기능을 이용하여 두 온톨로지를 병합한 결과는 (그림 3)과 같다.



(그림 3) 병합된 온톨로지

클래스 이름 뒤에 적힌 것을 'suffix'라 하는데, 최종 병합 결과에서 'suffix'를 제거한다. 제시된 결과에서 상위 클래스를 추출하여 RDFS의 방향이 있는 그래프 형태로 나타낸다[17]. 여기서 한 가지 유의할 점은 합병된 온톨로지가 두 개의 온톨로지를 합친 상태이기 때문에 상위 클래스를 포괄할 수 있는 최상위 클래스 이름이 정의되어야 한다는 점이다. 이 작업을 위해서는 Wordnet[18] 등의 어휘사전

을 통해 의미 해석 및 추론 과정을 거쳐 비행기예약과 자동차대여를 포괄하는 범주의 용어를 선택해야 한다. 단어 추론 단계를 고려하지 않은 최종 추출된 상위 온톨로지는 (그림 4)와 같다.



(그림 4) 추출된 '여행 예약' 상위 온톨로지

5. 결론 및 향후 과제

본 논문에서 우리는 같은 도메인 내에 있는 상이한 온톨로지를 PROMPT에서 제공하는 병합 기법을 통해 통합된 하나의 온톨로지로 만들고, 이를 바탕으로 누구나 접근할 수 있는 일반적인 목적을 지닌 온톨로지를 추출해 보았다.

우리의 방법은 PROMPT 도구를 사용하기 때문에 비주얼하고, 직관적인 결과를 알 수 있다. 그 결과를 바탕으로 만든 상위 온톨로지는 온톨로지 설계 방법에 있어 도메인 온톨로지로부터 상위 온톨로지를 추출해 내는 Bottom-up 방식을 사용한다. 이는 일반적인 온톨로지 및 상위 온톨로지 생성 방법인 Top-down 방식보다 현실적이고, 기존 온톨로지의 공유와 재사용을 용이하게 만드는 장점을 지닌다.

그러나 두 개의 온톨로지를 병합하고 이를 바탕으로 상위 온톨로지를 추출하는 과정에서 최상위 클래스의 이름이 자동으로 지정되지 않아 사람의 개입이 필요하다는 점은 여전히 해결해야 할 과제로 남아있다. 따라서 Wordnet 등을 이용하여 해당하는 온톨로지 이름에 대한 적절한 의미적 추론을 통한 상위 온톨로지 이름의 자동 지정 방법 등에 대한 연구는 향후 해결해야 할 과제이다.

참고문헌

[1] 옥철영 : 한국어 정보처리와 온톨로지.ppt 2004
한국어정보처리연구회 동계 튜토리얼
[2] Pascal Hitzler and Markus Krotzsch : What Is Ontology Merging? AAI (2005)

[3] Marc Ehrig and Steffen Staab: QOM-Quick Ontology Mapping. ISWC (2004)
[4] Jerome Euzenat : An API for Ontology Alignment. ISWC (2004) 698-712
[5] Michal Sev Cenko : Online Presentation of an Upper Ontology
<http://ontology.teknowledge.com/Sevcenko.pdf>
[6] Noy, N.F., Musen, M.A. : The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-computer Studies 59 (2003) 983-1024
[7] Kiyakov, Simov, Dimitrov : OntoMap: Portal for Upper-level Ontologies. Proceedings of the Euroconference Recent Advances in Natural Language Processing(RANLP-2001)
[8] Adam Farquhar, Richard Fikes, James Rice : The Ontolingua server: a tool for collaborative ontology construction. 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (1996)
[9] Deborah McGuinness, Richard Fikes, Jim Rice, and Steve Wilder : An environment for merging and testing large ontologies. 7th International Conference on Principles of knowledge REpresentation and Reasoning, (2000) 483-493
[10] Dou, D., McDermott, D., Qi, P., : Ontology translation by ontology merging and automated reasoning. EKAW'02 workshop on Ontologies for Multi-Agent Systems (2002)
[11] A. Doan, P. Domingos, A. Halevy. Learning to match the schemas of data sources : A multi strategy approach. VLDB Journal, (2003) 279-301
[12] Eduro Mena, Arantza LLarramendi, Vipul Kashyap, Amit Sheth : OBSERVER:An Approach for Query Processing in Global Information Systems based on Information Systems, 2000
[13] Franconi, E., Ng, G., : The i.com tool for intelligent conceptual modelling. 7th International Workshop on Knowledge Representation meets Databases (2000)
[14] <http://www.w3.org/TR/rdf-schema/>
[15] <http://protege.stanford.edu/>
[16] <http://protege.stanford.edu/plugins/prompt/prompt.html>
[17] Jeff Z. Pan, Ian Horrocks : Metamodeling Architecture of Web Ontology Languages. Proceedings of the Semantic Web Working Symposium (2001) 131-149
[18] Miller, G.A., Beckwith, R., Fellbaum, C. Gross, D., Miller, K.J.: Introduction to WordNet: an on-line lexical database. In International Journal of Lexicography 3 (4), (1990) 235 - 244.
[19] Keonsoo Lee, Wonil Kim, Minkoo: 정보 공유를 위한 온톨로지 맵핑 방법. 한국지능정보시스템학회 (2004)