

# 지능형 유비쿼터스 웹을 위한 온톨로지 통합 및 확장

## Ontology Integration and Extension for Intelligent Ubiquitous Web

이지윤, 양진혁, 정인정

고려대학교 전산학과

충남연기군 조치원읍 서창리 208 고려대학교 자연과학대학 202호

Tel: +82-41-860-1432, Fax: +82-41-864-0014, E-mail: {jjoon81, grjinh, chung}@korea.ac.kr

### Abstract

언제 어디서나 자유롭게 인터넷을 사용할 수 있는 다가오는 유비쿼터스 환경에서 원하는 정보를 빠르고 정확하게 찾고자 하는 사용자들의 요구는 중요한 이슈가 되었다. 대량의 정보에 대한 정확하고 의미 있는 검색의 해결책으로 제시된 한 분야가 바로 시맨틱 웹이다. 우리는 본 논문에서 시맨틱 웹의 핵심 기술인 차세대 데이터베이스라 불리는 온톨로지를 통합하고 이를 확장해 나아가는 방법에 대해 제안한다. 본 논문에서 온톨로지 통합에 사용되는 기법은 맵핑을 이용한 온톨로지 병합(merging) 기법이고, 이를 바탕으로 하는 상위 온톨로지 확장은 통합된 온톨로지를 메타레벨과 언어레벨로 나누는 작업을 통해 이루어진다. 제안된 방법은 중복되는 온톨로지들을 최소화하여 재사용성을 높이는 데 기여한다. 또한 온톨로지들 사이의 개념들 간의 관계를 효율적으로 만들고, 시맨틱 웹뿐만 아니라 나아가 유비쿼터스 웹 환경에서 사용자들이 원하는 정보의 공유를 용이하게 한다.

### Keywords:

Semantic Web, Ontology, Ontology Reuse, Ontology Integration, Upper Ontology

### Introduction

정보기술의 발달과 인터넷 사용 환경의 편리성 등으로 웹의 사용이 점차 일반화되고, 사용자간의 정보 공유에 대한 욕구 증가에 따라 웹 데이터의 양은 매우 방대해졌다. Tim Berners-Lee 는 1998 년 이에 대한 해결책의 하나로 시맨틱 웹을 제안하였다[1]. 이전까지의 웹이 단지 사람이 이해하기 편하고 사람이 사용하기 편리한 형태로

발전해왔다면, 시맨틱 웹은 웹 상의 데이터를 사람이 아닌 기계가 이해하고 처리할 수 있는 형태로 표현하는 것을 말한다. 다시 말해 인간과 기계가 상호작용하는 지능화된 웹 환경을 말한다. 언제 어디서나 자유롭게 인터넷을 할 수 있는 유비쿼터스 환경이 다가오에 따라 이러한 시맨틱 웹의 중요성 역시 증가하게 되었고, 이에 따라 시맨틱 웹의 핵심 기술이자 중추적 역할을 담당하는 온톨로지의 중요성 역시 증가하였다.

온톨로지는 일반적으로 특정한 도메인의 공유된 개념에 대한 명시적인 명세로 정의할 수 있다. 온톨로지를 이용하여 우리는 주어진 도메인에 대한 중요한 개념들의 관계와 속성을 추론을 통해 새로운 개념을 유추해 나아갈 수 있다. 이러한 온톨로지에 관한 연구는 현재 온톨로지의 생성 및 설계, 통합, 표현 등의 세 가지 방향으로 진행되고 있다[2]. 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)[3]이나 맵핑(mapping)[4], 정렬(alignment)[5] 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[6]. 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 중요성

에도 불구하고 이용이 쉽지 않다. 또한 이들 사이에 정식 맵핑 방법이 없어 상위 온톨로지를 기술하는 방법들에 대한 비교 및 연구 분석이 어렵다.

우리는 본 논문에서 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[7]를 사용해 두 온톨로지를 병합 또는 정렬한다. 이 결과를 가지고 개념과 관계를 바탕으로 메타레벨과 언어레벨로 온톨로지를 분류한다. 이는 온톨로지들 사이에 맵핑을 통해 개념들 간의 관계를 효율적으로 만들고, 이와 더불어 해당 도메인에 대한 일반화된 온톨로지는 추후 사용자들의 공유를 용이하게 한다[8].

본 논문은 다음과 같은 순서로 구성되어 있다. 먼저 2 장에서는 온톨로지 통합과 상위 온톨로지 추출과 관련된 연구를 소개하고 현재 진행 방향에 대해 살펴본다. 3 장에서는 우리가 선택한 맵핑 방법과 그에 따른 도메인 온톨로지 통합 방법을 알아 본다. 또한 맵핑 결과를 바탕으로 상위 온톨로지 추출을 위한 알고리즘을 제안한다. 4 장에서는 3 장에서 제안한 방법을 토대로 시나리오를 구성하고 그에 따른 실험 및 결과를 검증한다. 마지막 5 장에서는 결론을 도출한 후, 향후 과제를 살펴본다.

## Related Work

온톨로지 통합(integration)이란 같은 도메인 내에서 서로 독립적으로 개발된 각각의 온톨로지를 재사용 가능한 새로운 형태의 온톨로지로 만드는 것이다. 기존의 온톨로지를 바탕으로 새로운 온톨로지를 만들기 위해서는 두 온톨로지 사이의 개념의 우선순위를 비교하거나 개념들 간의 의미를 파악한 후 관계를 연결해주는 방법이 필요하다. 이러한 온톨로지 통합에 사용되는 기법들은 병합, 정렬, 맵핑 등이 있다. 먼저 병합은 같은 주제에 대한 서로 다른 온톨로지를 통합된 새로운 하나의 온톨로지로서 합치는 것을 말한다. 정렬은 두 개 또는 그 이상의 온톨로지를 상호동의 하에 가져와 일관되고 일치된 상태로 만드는 것이고, 맵핑은 서로 다른 소스들로부터 동일 관계에 의해 서로 간에 유사한 개념이나 관계를 연관시키는 방법이다. 그림 1[7]은 이러한 기법들을 나타낸 것이다.

### 1. 온톨로지 맵핑 도구들

온톨로지 맵핑에 관한 관련연구를 살펴보면 먼저 두 개의 온톨로지를 합병하여 새로운 하나의 온톨로지를 만드는 도구로서 개념 구조와 그들 사이의 관계로부터 정보를 사용하는 IPROMPT[7]가 있다. 또한 온톨로지 에디터를 바탕으로 하는 병합 도구인 Ontolingua[9]가 있고, 서로 다른 형식으로 만들어

진 온톨로지를 사용하는 Chimaera[10]와 두 개의 온톨로지를 공통의 문법 표현으로 변환한 후 axiom을 정의하여 인스턴스 해석 과정을 거치는 OntoMerge[11]도 있다.

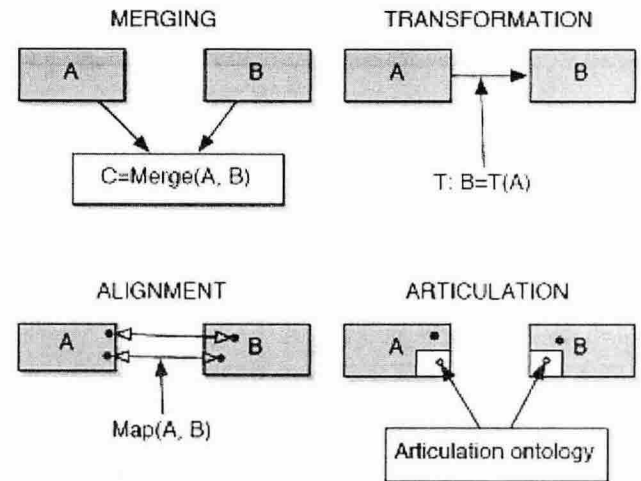


그림 1 - 다양한 온톨로지 통합 기법

연관된 개념들을 찾음으로써 두 온톨로지의 개념들 사이에 맵핑을 정의하는 도구들은 그래프 표현 방식의 온톨로지를 분석하여 소스 온톨로지의 유사한 개념들의 쌍을 합병하여 새로운 쌍을 생성하는 ANCHORPROMPT[7], 맵핑을 찾기 위해 기계학습 기법을 적용한 GLUE[12], 기술 로직(Description Logic)을 사용하여 맵핑에 대한 질의에 답하는 OBSERVER[13] 등이 이에 해당한다.

### 2. 스키마 통합

데이터베이스 스키마는 지식의 구조적 표현이라는 점에서 온톨로지과 유사하다. 데이터베이스 스키마 연구에서 스키마 통합에 대한 공통적인 접근 방법들은 둘 사이의 상관관계를 찾기보다는 중개자(mediator)를 개발하는 것이었다. 최근의 연구는 스키마 통합에 대한 스키마 매칭을 온톨로지 맵핑과 유사하다는 관점에서 바라본다. 두 방법 모두 관련된 개념을 찾는 것이 필요하고, 동시에 대부분의 스키마 통합 기법은 관계형 데이터베이스 스키마나 개념 트리를 포함하는 스키마를 고려한다. ICOM[14]은 스키마 통합과 온톨로지 통합 도메인을 합쳐놓은 프로젝트의 한 예로 스키마 통합을 위한 개념적 모델을 제시하고 그에 대한 추론 메커니즘을 제공한다.

### 3. 상위 온톨로지

상위 온톨로지란 특정 도메인에 대한 재사용 및 확장을 목적으로 만들어진 도메인 독립적인 온톨로지를 말한다. SUMO(the Suggested Upper

Merged Ontology)[6]의 가장 상위 단계에서는 일반적인 개념을 나타내는 엔티티를 루트 노드로 하는 단일 계층 형태로 구성된다. 이것을 실제 존재하는 부분(physical)과 그렇지 않은 부분(abstract)으로 나누고, 각각의 하위 단계를 구별한다. SUMO의 단점은 개념들과 자연 언어와의 상호 연결이 어렵다는 점이지만 이는 자동적 자연어 처리를 위한 어휘 사전인 WordNet[15]을 연결시킴으로써 일부분 해결이 가능하다. 그림 2는 SUMO의 최상위 단계의 개념들을 나타낸 계층도이다.

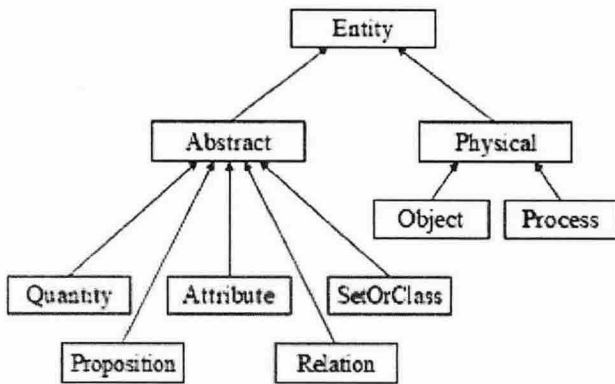


그림 2 - SUMO의 상위 단계 개념들

## Approach and Methods

### 1. 도메인 온톨로지 통합

우리는 이 장에서 상위 온톨로지를 추출하기 위한 방법에 대해 제안한다. 이 작업을 수행하기 위해서는 상위 온톨로지를 추출하기 이전에 두 개의 도메인 온톨로지를 통합하는 작업이 선행되어야 한다. 먼저 같은 도메인에서 두 개의 온톨로지를 선택한다. 같은 도메인내의 온톨로지를 사용해야 하는 까닭은 두 개의 온톨로지를 비교하는 위해서는 비교대상이 동일한 범주 내에 있어야 하기 때문이다. 현재 알려진 방법들로는 여러 개의 온톨로지를 동시에 비교 분석하는 것이 불가능하기 때문에 한번의 작업에 두 개의 온톨로지를 사용하는 방법을 선택한다.

그 다음, PROMPT를 플러그-인 한 Protege를 사용하여 두 개의 온톨로지의 병합 작업을 수행한다. PROMPT는 온톨로지 통합에 사용되는 다른 툴들과 비교해 비교적 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.

### 2. 상위 온톨로지 추출 방법론

PROMPT를 통해 생성된 결과 온톨로지는 계층적 트리의 클래스 형태이다. 결과 클래스의 인스턴스들과 내부 속성들은 외부에서는 보이지 않고, 단지 클래스로만 표현된 형태이다. 클래스로 표현된 이 결과 온톨로지를 메타레벨 클래스와 언어레벨 클래스들로 나눈다. 클래스 이름들을 하나의 개념으로 보고 내부 속성을 배제한 채 온톨로지 표현언어의 하나인 RDFS의 메타데이터[16]와 같이 상위 온톨로지를 구성한다. 계층적 형태를 가진 이 클래스에서 하위 클래스를 제거한 상위 클래스들만을 상위 온톨로지의 개념으로 보고 추출한 뒤 RDFS의 표현방법과 같이 방향이 있는 그래프 형태로 나타낸다. 위에서 제시한 방법들을 조합하여 순서도로 나타내면 그림 3과 같다.

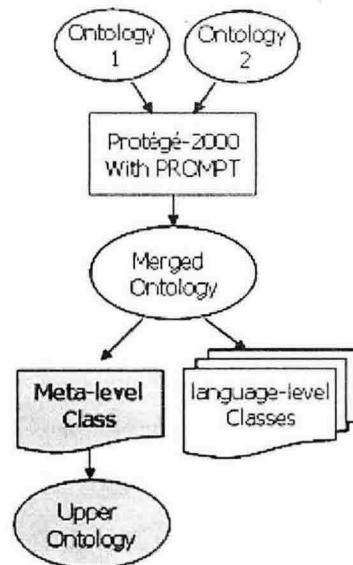


그림 3 - 상위 온톨로지 추출 순서도

이 방법을 통해서 생성된 상위 온톨로지는 기존의 온톨로지를 재사용하는 확장 가능한 형태의 온톨로지가 된다. 우리는 다음 장에서 앞에서 제시한 방법론을 바탕으로 간단한 예를 들고 이를 테스트한다.

## Simulation and Result

위에서 제안한 방법의 실험을 위하여 우리는 윈도우 XP에서 실행 가능한 Protege-2000의 가장 최신 버전인 protégé 3.1.1과 PROMPT 패키지를 다운로드 받는다[17,18]. PROMPT 사용을 위해 protégé 3.1.1에 PROMPT를 plug-in[16]하여 온톨로지를 맵핑 및 정렬하기 위한 적절한 환경을 만든다. 본 실험에서 사용된 온톨로지는 다운로드

받은 PROMPT 패키지에서 제공하는 테스트 온톨로지인 항공예약 온톨로지(Air\_reservation)와 자동차 대여온톨로지 (Car\_rental)를 사용한다[18]. 본 논문에서 사용되는 PROMPT의 주요 기능들은 표 1과 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 작업 창에는 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 1의 주요 기능 외에 PROMPT의 다양한 기능을 이용하여 두 온톨로지를 병합한 결과는 그림 4와 같다.

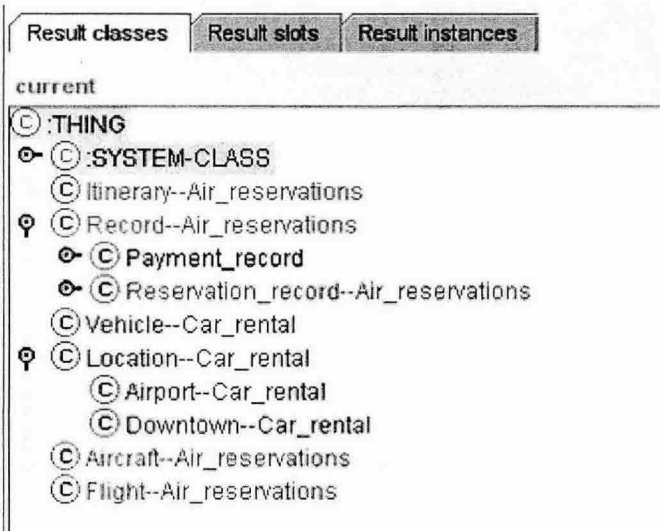


그림 4 - 병합된 온톨로지

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[19]. 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[15] 등의 어휘사전을 통해 의미 해석 및 추론 과정을 거쳐 비행기예약과 자동차대여를 포괄하는 범주의 용어를 선택해야 한다. 단어 추론 단계를 고려하지 않은 최종 추출된 상위 온톨로지는 그림 5와 같다.

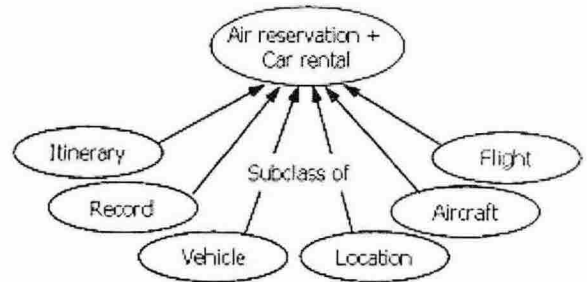


그림 5 - 추출된 '여행 예약' 상위 온톨로지

## Conclusion and Future Work

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

우리의 방법은 온톨로지 통합 도구인 PROMPT를 사용하여 비주얼하고, 직관적인 결과를 얻을 수 있다. 또한 그 결과를 바탕으로 만들어진 상위 온톨로지는 온톨로지 설계 방법에 있어 도메인 온톨로지로부터 상위 온톨로지를 추출해 내는 상황식(Bottom up) 설계 방식을 사용하고 있어 일반적인 온톨로지 및 상위 온톨로지 생성 방법인 하향식(Top down) 설계 방식보다 현실적이고, 온톨로지 공유와 재사용을 용이하게 만드는 장점을 지닌다.

그러나 두 온톨로지를 병합하고 이를 바탕으로 상위 온톨로지를 추출하는 과정에서 최상위 클래스의 이름이 자동으로 지정되지 않아 사람의 개입이 필요하다는 점은 해결해야 할 과제로 남아있다. 따라서 추후에는 Wordnet 등을 이용하여 해당하는 온톨로지 이름에 대한 적절한 의미적 추론을 통한 상위 온톨로지 이름의 자동 지정 방법 등에 대한 연구가 진행되어야 한다. 또한 통합과 추출이라는 각기 다른 이 두 작업을 자동적으로 연결해주는 방법에 관한 연구 역시 우리 연구의 최종목표인 지능형 웹을 구성하기 위해 반드시 해결해야 할 과제이다.

## References

- [1] Berners-Lee, Tim, Hendler, J, Lassila, O (2001). "The Semantic Web," *Scientific American*, pp.28-37.
- [2] 옥철영 (2004) 한국어 정보처리와 온톨로지.ppt  
한국어정보처리연구회 동계 튜토리얼
- [3] Pascal Hitzler, Markus Krotzsch. (2005). "What Is Ontology Merging?," *AAAI*
- [4] Marc Ehrig and Steffen Staab (2004). "QOM-Quick Ontology Mapping," *ISWC*
- [5] Jerome Euzenat (2004). "An API for Ontology Alignment," *ISWC*, pp. 698-712
- [6] Michal Sev Cenko "Online Presentation of an Upper Ontology"  
<http://ontology.teknowledge.com/Sevcenko.pdf>
- [7] Noy, N.F., Musen, M.A. (2003). "The PROMPT suite: interactive tools for ontology merging and mapping," *International Journal of Human-computer Studies*59, pp. 983-1024
- [8] Kiyakov, Simov, Dimitrov. (2001). "OntoMap: Portal for Upper-level Ontologies," *Proceedings of the Euroconference Recent Advances in Natural Language Processing*
- [9] Adam Farquhar, Richard Fikes, James Rice. (1996) "The Ontolingua server: a tool for collaborative ontology construction," *10th Knowledge Acquisition for Knowledge-Based Systems Workshop*
- [10] Deborah McGuinness, Richard Fikes, Jim Rice, and Steve Wilder. (2000). "An environment for merging and testing large ontologies," *7th International Conference on Principles of knowledge Representation and Reasoning*, pp.483-493
- [11] Dou, D., McDermott, D., Qi, P., (2002). "Ontology translation by ontology merging and automated reasoning," *EKAW workshop on Ontologies for Multi-Agent Systems*
- [12] A. Doan, P. Domingos, and A. Halevy. (2003) "Learning to match the schemas of data sources : A multi-strategy approach," *VLDB Journal*, pp.279-301
- [13] Eduro Mena, ArantzaIL Larramendi, Vipul Kashyap, Amit Sheth. (2000). "OBSERVER:An Approach for Query Processing," *Global Infirmation Systems based on Information Systems*
- [14] Franconi, E., Ng, G., (2000). "The i.com tool for intelligent conceptual modeling," *7th International Workshop on Knowledge Representation meets Databases*
- [15] Miller, G.A., Beckwith, R., Fellbaum, C. Gross, D., Miller, K.J. (1990). "Introduction to WordNet:an on-line lexical database," *In International Journal of Lexicography* 3 (4), pp.235 - 244.
- [16] <http://www.w3.org/TR/rdf-schema/>
- [17] <http://protege.stanford.edu/>
- [18] <http://protege.stanford.edu/plugins/prompt/prompt.html>
- [19] Jeff Z. Pan, Ian Horrocks. (2001). "Metamodeling Architecture of Web Ontology Languages," *Proceedings of the Semantic Web Working Symposium* pp. 131-149
- [20] Keonsoo Lee, Wonil Kim, Minkoo. (2004). "정보공유를 위한 온톨로지 맵핑 방법", *한국지능정보시스템학회*
- [21] Helena Sofia Pinto. (2001). "A Methodology for Ontology Integreation," *Proceedings of the international conference on Knowledge capture*, pp.131-138