

UML 클래스 모델 기반 온톨로지 통합

서진원*, 김영태*, 임재현*, 김치수*

*공주대학교 컴퓨터공학과

e-mail : sjwon@kongju.ac.kr

An Integration of the Ontology Based on UML Class Models

Jin-Won Seo*, Young-Tae Kim*, Jae-Hyun Lim*, Chi-Su Kim*

*Dept. of Computer Engineering, Kong-Ju National University

요 약

데이터 통합은 사용자에게 상이한 소스의 데이터에 대한 단일 접근 점을 제공할 목적으로 데이터를 내부적으로 운영하기 위한 효과적인 방법이다. 1)데이터의 풍부한 시멘틱스는 이종의 데이터 소스들 사이의 충돌을 해결하기 위한 주요한 요인이다. UML 클래스 모델은 데이터의 스키마기반 시멘틱만을 표현하기 때문에 온톨로지와 같은 대체 방법은 추가적인 시멘틱을 표현하는데 유용한 방법이다.

본 논문에서는 각각의 데이터 소스의 시멘틱을 기술하기 위해 온톨로지를 사용하고, 데이터들의 유사성과 차이점을 결정하기 위해 온톨로지를 분석하고, 비교한다. 비교의 결과를 이용해 통합 정보에 대한 질의를 사용 가능하게 할 수 있는 통합 온톨로지를 구축한다.

1. 서론

사용자들은 원하는 정보가 수많은 정보시스템에 존재하는 방대한 양의 데이터로부터 양질의 정보가 정확하고 일관성 있게 제공되기를 원한다. 이를 위해 서로 다른 정보시스템에 있는 데이터들을 결합하여 통합된 데이터 시스템을 구축함으로써 일관되고 정확한 양질의 데이터를 제공할 수 있다. 그러나 스키마의 이종성, 시멘틱의 이종성 같은 문제가 이종자료의 완벽한 통합에 걸림돌이 되고 있다.

현재 많은 연구에서 이런 문제를 해결하기 위해 전역 스키마, 지역 스키마, 전역 스키마와 지역 스키마의 매핑 구조를 제시하고 있다.

그러나 이런 방식에서도 서로 다른 이름으로 동일 데이터를 표현하는 이름 충돌, 서로 다른 데이터 타입으로 동일 데이터를 표현하는 구조 충돌, 시멘틱 또는 데이터 레벨 충돌 등이 있을 수 있다[1]. 또한 이러한 표현방식은 매핑의 어려움으로 인한 비용과 시간의 문제, 부정확한 매핑의 문제가 있다.

본 논문에서는 이러한 매핑의 문제를 해결하기 위해 UML 클래스 모델을 사용하여 스키마 충돌을 해결하고, 시멘틱 또는 데이터 레벨의 충돌을 해결하기 위하여 추가적인 시멘틱을 사용하는 온톨로지 개념을 사용하였다[2].

본 논문에서는 OMG에서 제안한 DSTC 온톨로지 정의 모델을 이용해서 데이터 소스 스키마를 온톨로지 기반 스키마로 변환한다[3]. 이 온톨로지 기반 스키마는 추가적인 데이터 시멘틱에 의해 확장된다.

OBSERVER, FCA-MERGE, PROMPT, Chimaera 등의 연구에서 이기종의 데이터 통합을 위한 기술을 제안하고 있지만 비구조적이거나 반구조적인 데이터를 통합하는 경향 때문에 알고리즘이 복잡해지게 된다.

본 논문에서는 구조적인 두 UML 클래스 모델을 통합하므로 온톨로지는 보다 더 구조적이고, 보다 덜 복잡한 알고리즘을 사용하게 된다.

2. 관련 연구

2.1 온톨로지 데이터 통합

기존의 온톨로지 데이터 통합 방법은 세 가지 방

이 논문은 2006년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(NO. R01-2006-000-10555-0)

법으로 나눌 수 있다[4].

a) 단일 온톨로지 접근방법

전역 온톨로지에 모든 소스의 데이터를 생성 하고, 전역 온톨로지를 통하여 통합데이터 기반 쿼리를 요청하는 구조로 모든 소스의 데이터에 대하여 잘 아는 전문가를 필요로 한다.

b) 다중 온톨로지 접근방법

데이터 쿼리 요청 시 각 지역 온톨로지의 정리된 매핑 자료의 조합으로 데이터가 제공된다. 사용자는 적당한 매개자가 되는 온톨로지를 선택해야 한다.

c) 하이브리드 온톨로지 접근방법

공유 온톨로지가 지역 온톨로지 사이의 공유된 표현 형식에 의해 구축되어 쿼리 요청을 해결한다. 온톨로지간의 쿼리 과다 발생으로 오버헤드가 발생하는 문제를 갖는다.

본 논문에서는 단일 온톨로지 접근방법과 하이브리드 접근방법을 혼합 사용하여 온톨로지 기반의 데이터 통합을 하고자한다.

2.2 UML 클래스 모델에서 온톨로지로의 변환

온톨로지는 UML 클래스 모델과 같은 구조적 데이터를 표현한다. 온톨로지 개념을 사용하면 지식은 클래스, 속성, 속성에 대한 제약사항으로 기술된다.

많은 연구에서 UML 클래스 모델로부터 온톨로지로의 변환을 다룬다. [5]에서는 UML, DL, ER와 온톨로지 사이의 매핑을 제안했다. Java2OWL은 자바 클래스와 사례를 카디널리티 제약과 XML 스키마 데이터타입을 사용하여 OWL로 변환한다[6].

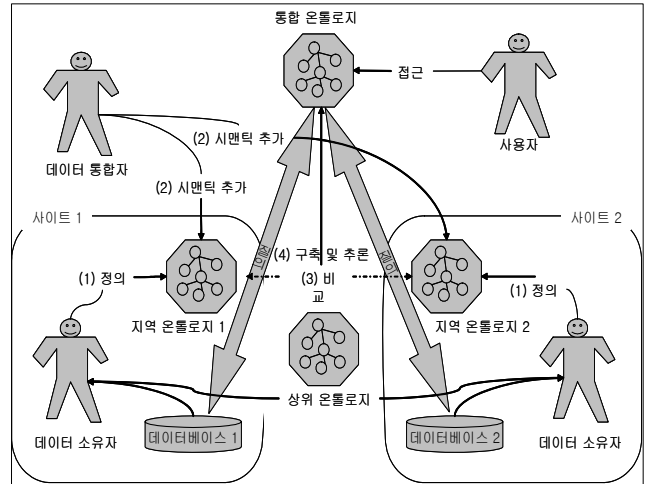
본 논문에서는 OMG에서 발표한 DSTC 온톨로지 정의의 메타 모델에 따라 온톨로지의 시멘틱을 표현하고, UML을 사용하여 온톨로지를 기술하여 UML 클래스 모델에서 온톨로지로 변환한다[3].

3. UML 클래스 기반 온톨로지 통합

본 논문의 데이터 통합 방법은 두 개의 UML 클래스 모델의 구조적 정보를 기술하는 온톨로지를 통합한다. 데이터 통합은 [그림 1]과 같이 정의, 시멘틱 추가, 비교, 구축 및 추론의 4 단계로 구성된다.

3.1 지역 온톨로지의 정의

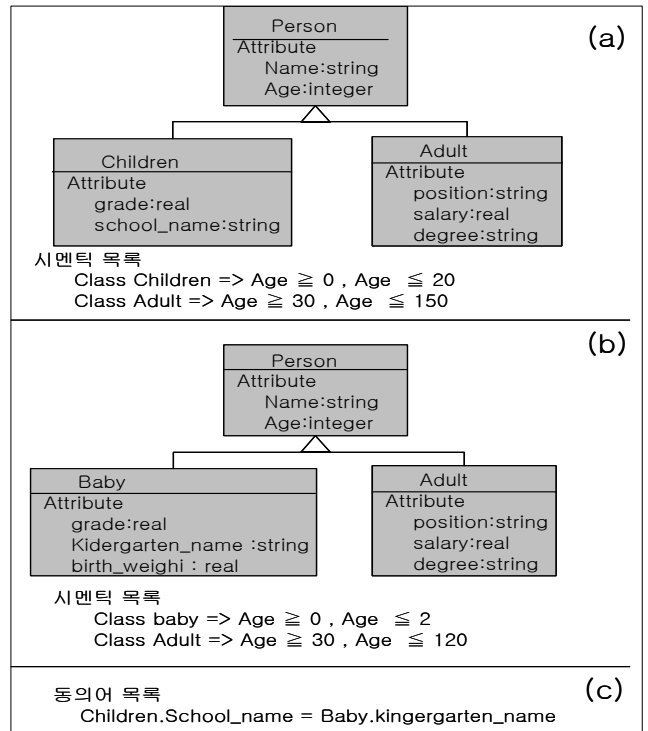
[그림 1]의 상위 온톨로지는 두 데이터 모델의 지역 온톨로지에 대한 메타 온톨로지이다. 지역 온톨로지의 데이터 소유자와 데이터 통합자는 상위 온톨로지를 이용하여 시멘틱을 정의하거나 추가한다.



[그림 1] 데이터 통합 과정

스키마 기반 시멘틱은 이미 UML 클래스 모델에 존재하는 시멘틱으로 [그림 2]의 UML 클래스 다이어그램을 참조하여 클래스와 속성을 나타내는 "Class" 와 "ClassAttribute"를 사용한다. 온톨로지 객체 프로퍼티 "hasClassAttribute"는 속성을 가지고 있는 클래스의 프로퍼티를 기술하고, 온톨로지 데이터 타입 프로퍼티 "hasClassAttributeType"은 속성의 데이터 타입을 기술한다.

추가적인 시멘틱은 데이터 통합자가 지역 온톨로지에 추가할 수 있는 시멘틱으로 [그림 2]에서는 (a), (b)의 시멘틱 목록과 (c)의 동의어 목록으로 표현된다.



[그림 2] UML 클래스 모델과 추가적인 시멘틱

3.2 지역 온톨로지의 비교

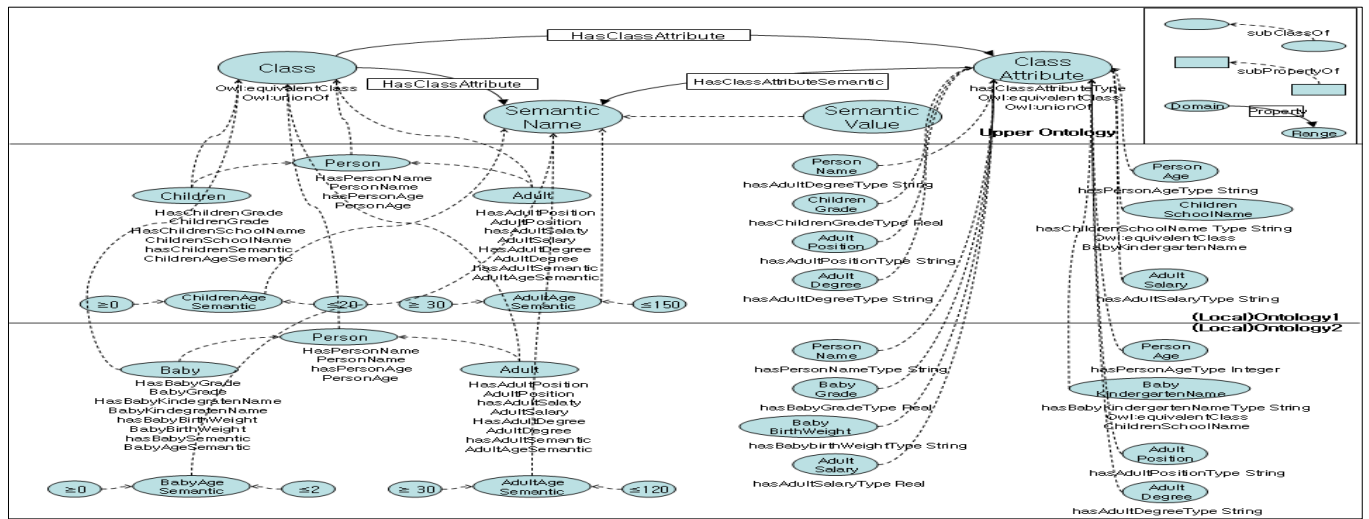
상위 온톨로지의 “Class”의 서브클래스인 두 개의 UML 클래스 사이의 관계를 발견하기 위해 클래스 비교를 한다. 클래스 비교는 시멘틱 비교와 클래스 속성 비교로부터 시작되는 상향식 프로세스이다.

먼저, 두 개의 지역 온톨로지에서 클래스의 각 쌍을 선택한다. 다음으로, 클래스 시멘틱 비교와 클래스 속성 비교를 각각 수행한다. 마지막으로, 클래스 비교에 의해 각각의 클래스 사이의 관계를 발견한다. [그림 3]은 상위 온톨로지와 지역 온톨로지를 나타낸다.

[표 2] 시멘틱의 고려사항과 조건의 관계

Case	Ms	결과	관계	s
1	$\alpha (\alpha = \beta)$	동치	동치	1
2	$\alpha (\alpha < \beta)$	동치	상위클래스	1
3	$\geq \alpha (\alpha = \beta)$	동치	동치	1
4	$\geq \alpha (\alpha < \beta)$	포함	상위클래스	1
5	$\geq \alpha$	강한 집합	형제관계	0.1
6	$< \alpha$	약한 집합	관계없음	0
7	0	관계 없음	관계없음	0

각 쌍의 클래스속성을 비교하여 휴리스틱 값을 기록한다. 휴리스틱 값은 비교되는 두 클래스 사이의 관계 정도를 말한다. C1과 C2는 비교되는 지역 온톨로지 1과 지역 온톨로지 2의 클래스속성이다. θ_{C1} 은



[그림 3] 상위 온톨로지와 지역 온톨로지

3.2.1 시멘틱 비교

비교되는 두 시멘틱 사이의 관계의 정도를 표시하는 수치값을 가지고 비교한다. T1은 지역 온톨로지 1의 조건이고, T2는 지역 온톨로지 2의 조건이다. α 는 T1, β 는 T2의 시멘틱 이름의 수이며 $\alpha \leq \beta$ 이다. 각각의 시멘틱 이름과 시멘틱 값을 비교한 휴리스틱 값은 [표 1]과 같다.

시멘틱 쌍들의 비교 값 $R = \{r_1, r_2, \dots, r_\alpha\}$ 이 반영된 휴리스틱 점수는 $Ms = \sum r_i, r_i \in R, (i=1, \dots, \alpha)$ 로 계산되며, [표 2]는 T1과 T2가 가지는 관계의 종류를 나타낸다.

[표 1] 시멘틱 T1과 시멘틱 T2의 비교 결과

Case	시멘틱 이름	시멘틱 값	r
1	동일	동일	1
2	동일	일부 동일	0.8
3	동일	겹침	0.8
4	동일	연결 없음	0.5
5	다름	전부 다름	0

3.2.2. 클래스속성 비교

C1의 클래스 속성의 수, θ_{C2} 는 C2의 클래스 속성의 수이고, $\alpha = \theta_{C1}, \beta = \theta_{C2}, \alpha \leq \beta$ 이다.

C1의 클래스 속성이 C2의 클래스 속성과 “owl:unionOf” 프로퍼티를 갖는다면 $\alpha = \alpha - 1 + C1$ 의 클래스 속성의 수이고, C2의 클래스 속성이 C1의 클래스 속성과 “owl:unionOf” 프로퍼티를 갖는다면 $\beta = \beta - 1 + C2$ 의 클래스 속성의 수이다.

클래스 속성의 비교 결과 값 $S = \{s_1, s_2, \dots, s_\alpha\}$ 는 $Ms = \sum s_i (s_i \in S)$ 로 계산한다. [표3]은 C1과 C2가 갖는 관계의 종류를 보여준다.

[표 3] 클래스 속성을 고려한 C1과 C2의 관계

Case	Ms	결과	관계
1	$\alpha (\alpha = \beta)$	동치	동치
2	$\alpha (\alpha < \beta)$	포함	상위클래스
3	$\geq \alpha (\alpha = \beta)$	동치	동치
4	$\geq \alpha (\alpha < \beta)$	포함	상위클래스
5	$\geq \alpha$	강한 공통	형제관계
6	$< \alpha$	약한 공통	연결없음
7	0	관계없음	연결없음

3.2.3 클래스 비교

클래스 C1과 C2의 관계는 클래스 시멘틱 관계와 모든 클래스 속성 사이의 관계에 기초한다. 클래스 ‘Children’과 ‘Baby’의 비교 결과는 ‘Children’이 ‘Baby’의 슈퍼클래스인 클래스 시멘틱과 클래스속성 집합을 갖는다. 따라서 ‘Children’은 ‘Baby’의 슈퍼클래스이다. 두 온톨로지의 다른 클래스의 쌍도 같은 방법으로 비교된다.

4. 통합 온톨로지 구축

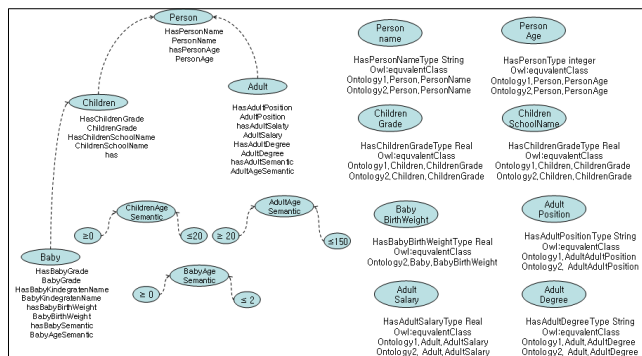
[표 4] 온톨로지 1과 온톨로지 2의 비교 결과

	Ontology1. Person	Ontology1. Children	Ontology1. Adult
Ontology2.p erson	동치	슈퍼클래스	슈퍼클래스
Ontology2. Baby	서브클래스	서브클래스	형제관계
Ontology2. Adult	서브클래스	형제관계	동치

온톨로지 비교 결과는 각 클래스 쌍들을 비교하여 동치, 슈퍼클래스, 서브클래스, 형제관계와 같이 관계를 식별한다. 관계 결과에 의해 새로운 통합 온톨로지를 구축할 수 있다. [표 4]는 클래스 쌍들의 비교 결과이다.

비교한 클래스 각 쌍들의 관계 결과에 따라 통합의 순서가 결정된다. 온톨로지 내의 클래스가 다른 온톨로지 내의 클래스와 통합되기 전에 반드시 클래스와 다른 온톨로지 내의 클래스 사이의 관계를 결정해야 한다. 동치 관계가 가장 높은 우선순위를 갖으며 가장 먼저 통합되고 슈퍼클래스, 서브클래스, 형제관계 순으로 통합이 이루어진다.

[그림 4]은 ‘Person’의 서브클래스인 ‘Children’과 ‘Adult’는 루트 클래스인 ‘Person’으로부터 상속받고, ‘Baby’도 ‘Children’으로부터 상속받는 통합 온톨로지이다.



[그림 4] 통합 온톨로지

통합 온톨로지에 추론 규칙을 추가하면 질의 처리 시에 온톨로지의 능력과 정확성을 증가시켜 줄 것이다. 예를 들어 ‘Children’과 ‘Baby’의 시멘틱에 “2살 이하의 ‘Children’은 또한 ‘Baby’이다.” 라는 사실을 추가할 수 있다.

통합 온톨로지가 만족스럽지 못하면 온톨로지에 더 많은 시멘틱을 추가하기 위해 시멘틱 추가 단계로 되돌아가게 된다.

5. 결론 및 향후 연구

UML 클래스 모델과 같은 구조화된 데이터 모델을 통합하기 위한 온톨로지의 사용은 온톨로지 표현으로부터 얻는 장점이 있고, 좋은 통합 결과를 얻을 수 있다. 질의가 온톨로지의 추론 능력에 의한 시멘틱을 바탕으로 하기 때문에 사용자 또한 온톨로지를 이용해 시멘틱을 기술함으로써 이익을 얻을 수 있다.

본 논문에서는 온톨로지 통합구조의 프로토타입을 제시하였다. 향후 연구로는 온톨로지 통합 접근방법을 정제하고, 속성에 대한 제약조건과 같은 더 많은 종류의 시멘틱을 지역 온톨로지에 포함시킬 것이다. 또한 실무 어플리케이션의 데이터 모델에 적용하고, 다른 연구들과 비교하여 성능을 평가하고자 한다.

[참고문헌]

- [1] Ram, S. and Park, J. Semantic conflict resolution ontology(SCROL): an ontology for detecting and resolving data and schema-level semantic conflicts. Knowledge and Data Engineering, IEEE Transactions on, 16 (2). 189-202.
- [2] Gruber, T.R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition (1042-8143), 5 (2). 199 - 220.
- [3] DSTC. Ontology Definition MetaModel - DSTC Initial Submission, OMG, 2003.
- [4] Wache, H., Vogele, T., et al., Ontology-Based Integration of Information - A Survey of Existing Approaches. in The Seventeenth International Joint Conference on Artificial Intelligence, (Seattle, Washington, USA, 2001).
- [5] Colomb, R.M., Gerber, A., et al., Issues in Mapping Metamodels in the Ontology Development Metamodel Using QVT. in The 1st International Workshop on the Model-Driven Semantic Web (MSDW 2004), (Monterey, California, USA, 2004).
- [6] Dean, M. Java2OWL, DAML.org, 2003.