

Description Logic을 이용한 전자카타로그 온톨로지 모델링

(Ontological Modeling of E-Catalogs using Description Logic)

이 현 자 [†] 심 준 호 ^{**}
(Hyunja Lee) (Junho Shim)

요 약 전자카타로그는 상품에 관한 풍부한 의미를 포함하며 온톨로지가 적용될 여지가 높은 영역이다. 온톨로지는 사물의 본질과 사물간의 관계를 나타내는데, 전자카타로그를 형식적(formal)인 온톨로지로서 표현하는 것은 전자상거래 영역에서 중요하다. DL(Description Logic)은 현재 사용되는 여러 온톨로지 언어의 이론적 기반을 제공한다. 본 논문은 DL 언어를 사용해 전자카타로그를 형식적으로 온톨로지 모델링한다. 기본적인 모델링 구성요소 집합은 확장 개체 관계(Extended Entity Relationship)로 개념적으로 나타내고, DL 언어로 대응해서 변환한다. EER로 나타낼 수 없는 추가적인 의미적 지식은 직접 DL언어로 표현한다. 논문에서 제시하는 모델링 언어는 언어의 표현력과 복잡도를 고려할 때, 현실적으로 추론에 적합하다고 알려져 있는 SHIQ(d) 언어 범위에 기반하고 있다. 모델링한 온톨로지 e-카타로그의 활용 이해를 돕기 위해 시나리오를 제시하고, 실제 DL 추론 도구를 통해 그 시험적 검증을 한다.

키워드 : 전자카타로그, 전자상거래, 온톨로지, Description Logic, EER

Abstract Electronic catalog contains rich semantics associated with products, and serves as a challenging practical domain for ontology application. Ontology is concerned with the nature and relations of being. It can play a crucial role in e-commerce as a formalization of e-Catalogs. Description Logics provide a theoretical core for most of the current ontology languages. In this paper, we present an ontological model of e-Catalogs in DL. We take an Extended Entity Relationship approach for conceptual modeling method, and present the fundamental set of modeling constructs and corresponding description language representation for each construct. Additional semantic knowledge can be represented directly in DL. Our modeling language stands within SHIQ(d) which is known reasonably practical with regard to its expressiveness and complexity. We illustrate sample scenarios to show how our approach may be utilized in modeling e-Catalogs, and also implement the scenarios through a DL inference tool to see the practical feasibility.

Key words : e-Catalog, e-Commerce, Ontology, EER, Description Logic

1. 서 론

1.1 배경

온톨로지는 [1]에 의해 제안된 시멘틱 웹(Semantic Web)의 비전(vision)을 실현하는 필수불가결한 요소 중의 하나이다. 도메인에 알맞은 온톨로지를 구축하는 것은 응용 도메인에서 사용자의 해석과 필요에 따라 모호해질 수 있는 방대하고 이질적인 정보를 의미적 지식으

로 명확히 정의하여 공유할 수 있다는 이점을 제공해 준다[2-4]. 공급망사슬관리(supply chain management), e-마켓플레이스(e-marketplace), 전자조달(e-procurement)과 같은 전자상거래 분야에서 관련 정보의 풍부한 의미가 시스템적으로 처리된다면 얻을 수 있는 효과는 상당하다[5-7].

e-카타로그는 전자상거래의 중요한 연구 주제이고 또한 전자상거래 시스템의 핵심이 되는 시스템 컴포넌트이기도 하다. e-카타로그는 가격, 상품의 특성, 용어 등 거래파트너 간에 요청하고 제공되는 상품과 서비스에 관한 정보를 담고 있다. 전자상거래에서 이루어지는 비즈니스 트랜잭션(business transaction)은 이러한 정보를 서로 교환하고 처리하는데, 의미적으로 풍부하고도

· 본 연구는 숙명여자대학교 2003학년도 교내연구비 지원에 의해 수행되었음

[†] 비 회 원 : 숙명여자대학교 컴퓨터학과
hyunjalee@sookmyung.ac.kr

^{**} 정 회 원 : 숙명여자대학교 정보과학부 교수
jshim@sookmyung.ac.kr

논문접수 : 2004년 6월 23일

심사완료 : 2004년 12월 27일

명확한 정보가 잘 처리되게 되면 그만큼 트랜잭션의 효과와 질적인 향상을 가져올 수 있다. 이러한 측면에서 e-카타로그 영역에 온톨로지가 활용될 수 있는 것이다.

e-카타로그에 온톨로지를 사용한다는 것 자체는 새로운 발상은 아니더라도, 문헌상으로 나타난 기존의 연구에서 e-카타로그에 온톨로지를 형식적(formal)하게 모델링하고 그 이점을 구체적으로 보여준 사례는 아직까지 없었다. e-카타로그 온톨로지 모델링하기 위해서는, 먼저 e-카타로그 도메인 내용에 대한 전문적 기술과 경험이 필요하다. 또한 그 도메인에 필요한 온톨로지의 표현과 복잡도를 기술적으로 분석해 낼 수 있어야 한다. 뿐만 아니라, DL 언어에 대한 확실한 이해가 바탕이 되어야 하는데, 이는 DAML+OIL[3]과 OWL[8]과 같은 이용할 수 있는 온톨로지 언어가 DL(Description Logic)기반의 언어이기 때문이다.

본 논문에서는 e-카타로그를 온톨로지로 표현하기 위해서 DL 언어를 사용한다. 제시되는 모델은 카타로그 개별 아이템의 지식 표현뿐 아니라, 아이템 간의 의미적 관계에 대한 지식을 명시적이면서도 형식적으로 표현해 준다. DL 언어의 기술적인 난이도는 DL 만을 통한 전체 온톨로지 모델의 개발을 어렵게 하므로, 개념적(conceptual)이고 가시적(visible)인 모델링 기법의 가미는 전체 온톨로지 모델 개발을 용이하게 하여준다. 따라서 본 논문에서는 확장 개체 관계(Extended Entity Relationship)[9,10] 모델링 기법을 사용하여 온톨로지의 기본적인 구성요소 집합을 제시한다. 기본 구성요소로는 클래스, 속성, is-a관계, 부분-전체관계, 객체(object)를 들 수 있다. e-카타로그의 풍부한 의미를 모두 EER로 표현하기는 불가능하므로 추가적인 의미들은 직접 DL로 표현한다. DL은 EER모델에서 표현할 수 없는 불의 전이적 관계, 역 관계, 롤의 계층관계를 표현할 수 있다. EER로 표현된 기본 구성요소도 기계적으로 DL언어 영역으로 번역될 수 있으므로, 결과적으로 모델은 모두 DL로 표현되어 진다.

본 논문에서는 제시된 e-카타로그 형식적인 온톨로지 모델링의 활용 이해를 돕기 위해 시나리오를 제시하였고, 모델링 제시 자체에 그치지 않고 그 시험적 검증을 위해 실제 추론 도구를 사용하였다

1.2 개괄 예제(A Motivating Example)

본 절에서는 온톨로지 카타로그 모델링을 개괄적으로 보여주도록 한 예제 시나리오를 소개한다. 제시한 모델링에 대한 설명은 본문에서 자세히 다룰 것이다.

식품 의약품 안전청(Food & Drug Administration)에서 관리하고 있는 모든 상품 목록에서 광우병을 유발할 수 있는 상품을 모두 찾아내고자 한다고 하자. 광우병(mad-cow disease 또는 의학 용어로 BSE)을 유발

하는 상품이란, 광우병에 감염된 소의 뇌(cow brain), 뼈(cow bone), 내장(cow tripe)을 포함하는 식품으로서 잠재적으로 사람에게 광우병을 옮길 수 있는 품목이다. 소의 뇌, 뼈, 내장은 그 자체로 상품이 될 수도 있고, 다른 식품의 재료로 사용될 수도 있다. 따라서 카타로그에는 상품과 재료의 정보가 있어야 관련 상품을 찾아낼 수가 있다. 예를 들어, 의료용기기중의 수술용 봉합실(suture)의 재료로 소의 내장이 쓰이고, 화장품이나 왜소증 치료를 위한 호르몬제에 소의 뇌가 쓰이므로 이러한 정보 표현이 필요하다.

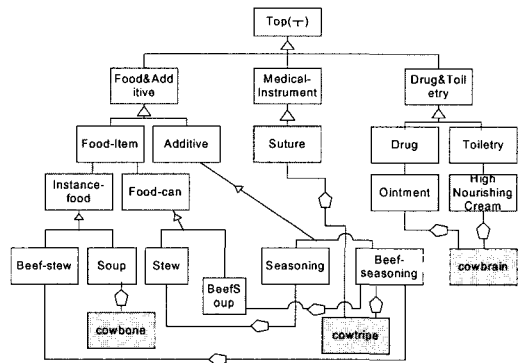


그림 1 식품 의약품 안전청의 카타로그 모델: 광우병 유발 상품 찾기 위한 온톨로지적 표현이 요구됨

그림 1의 EER 모델에서는, 소의 내장(cowtripe)은 쇠고기양념(Beef-seasoning)의 재료이고, 쇠고기양념은 다시 쇠고기스프(Beef-soup) 또는 쇠고기스튜(Beef-stew)의 재료가 되는 식품과 그 재료사이의 관계를 보여주고 있으며, 모델에서 오각형의 표시가 바로 식품과 그 재료간의 관계를 나타내고 있다. 즉, Beef-seasoning (쇠고기양념)은 소의 내장(cowtripe)으로 만들어진(made-of) 양념(Additive)이므로, 이를 DL 언어로 표현하면, $Beef-seasoning \subseteq Additive \cap \exists made-of.^1) cowtripe$ 과 같다. $\exists made-of$ 의 의미는 쇠고기양념(Beef-seasoning)과 재료관계(made-of)가 있는 클래스의 적어도 하나는 소의 내장이라는 것이다.

'광우병에 감염된 소의 뇌, 뼈, 내장, 혹은 그것을 재료로 한 식품이 광우병을 유발할 수 있는 품목이다'라는 지식은 규칙(rule)으로서 DL표현으로는, $FoodAdditive \cap (\exists made-of.Cowbrain \cup \exists made-of.Cowtripe \cup \exists made-of.Cowbone) \subseteq Bse$ 과 같이 모델내에서 표현될 질 수 있다. 따라서, 광우병을 유발 할 수 있는 식품 클래스를 모두 찾기 위해 그림 1의 모델에 대해 질의(concept-descendants of Bse)를 하면, {Soup,

1) $A \equiv \exists R.B$ 가 의미하는 것은. $a \in A, b \in B$ 라고 할때, 모든 a에 대해서, a가 적어도 하나이상의 b와 R의 관계임이다.

Beef-stew, Beef-soup, Beef-seasoning)을 추론할 수 있게 된다.

2. 연구 배경 및 관련 연구

2.1 온톨로지와 DL(Ontology and Description Logics)

웹에서의 정보는 잘 정의되어진 의미로 통합되어, 컴퓨터와 인간 사이에 상호작용이 가능한 방향으로 확장되고 있다[10]. 향후 시멘틱 웹으로 가기 위해서는 온톨로지를 구축하고 활용해야 한다. 온톨로지는 형이상학의 한 분야로 사물의 본질과 사물간의 관계에 관련한 것이다. 일반적으로 컴퓨터 분야에서의 온톨로지는 개념적으로 해석되는 부분에 대한 설명을 명시적인 형태로 형식화하여 공유할 수 있도록 약속한 것이다[2]. 이것은 보다 구체적으로 클래스의 정의, 클래스의 분류 계층, 클래스간의 포함(subsumption)관계, 그리고 객체와 그에 관한 의미적 지식을 의미한다.

OIL(Ontology Inference Layer), DAML(DARPA Agent Markup Languages)+OIL[3], OWL(Web Ontology Language)[8]등 XML기반의 마크업 언어는 웹에서 온톨로지를 실현하기 위해 개발된 언어들이며 전자상거래 시스템과 같은 웹 응용 분야에서 상당한 영향을 끼치고 있다.

온톨로지를 사용하려면 추론이 가능하고 잘 정의된 온톨로지 언어가 필요하다. 온톨로지 언어는 관련 개념을 충분히 정의할 수 있을 정도의 표현이 가능해야 하면서, 상대적으로 추론이 결정 가능(decidable)해야 한다. 이러한 조건을 만족하는 것이 DL이며, 이 점이 앞서 언급했던 웹 온톨로지 언어의 이론적 배경으로 DL을 사용하는 이유이다[11,12].

DL은 구조화되고 형식화된 표현 방식이므로 응용 도메인의 지식을 표현했을 때 이해하기 좋은 언어중의 하나이다[13]. DL은 다소 형식적인 논리에 기반한 의미를 사용하고 있으며, 의미적 네트워크(semantic network)나 프레임워크(frame work)와는 구별된다. DL은 컨셉(concept)과 롤(role)을 구성요소로 갖는다. 컨셉은 클래스, 롤은 클래스간의 관계를 표현할 수 있다. 가장 기본적인 DL 언어 중의 하나인 *ALC*[11]는 \sqcap (conjunction), \sqcup (disjunction), \neg (negation), 존재제한구성요소(existential restriction constructor) $\exists R.C$, 값제한구성요소(value restriction constructor) $\forall R.C$ 의 표현이 가능하다.

*ALC*의 확장 언어인 *SHIQ(d)*는 한정수제한(qualified number restriction), 역역할(inverse roles), 이행적 롤(transitive roles), 롤 계층(role hierarchy), 그리고 구체적인 도메인(concrete domain)을 표현할 수 있으며, 언어의 표현과 추론의 복잡도 사이의 균형이 적당

한 DL언어로 알려져 있다[4,12,13]. 본 논문의 모델링 언어는 *SHIQ(d)*에 기반하고 있으며, FaCT[14], RACER[4]과 같은 추론 도구들은 본 논문의 모델의 실행 가능성을 테스트하기 위해 사용되었다.

2.2 전자카타로그와 의미적 모델링

최근의 연구를 살펴보면, 전자상거래에서 온톨로지는 잠재적인 이점이 있고, 상호 운영에서 많이 발생하는 문제들을 해결할 수 있는 열쇠이다[5,7,15]. 인터넷 환경에서의 전자상거래가 지닌 주요 문제 가운데 하나는, 비슷한 상품이 헤아릴 수 없이 많으며 이들 상품에 대한 정보가 다양하고 방대하다는 점이다. 이들 상품 정보로부터 원하는 정보를 알아내기 위해서는 먼저 정보 자체가 명확히 잘 표현되고 정리되어 있어야 하고, 전자상거래 시스템간의 상품 정보의 상호운영이 원활히 이루어져야 한다[14]. 따라서 전자상거래 시스템의 주요 컴포넌트인 e-카타로그는 온톨로지를 활용했을 때 그 효과를 볼 수 있는 분야로 여겨질 수 있다.

온톨로지가 e-카타로그외의 다른 도메인에서 적용 사례로서는, 보험회사인 Swiss Life²⁾의 온톨로지 구축 사례와 Medical Thesauri[16]의 의학용어 온톨로지 구축 사례를 대표적인 예로 들 수 있다. 특히 후자의 사례에서는 잘 정비된 기존의 의학 용어들의 정의와 그 의미적 관계를 형식적인 온톨로지화 하는 경험적 방법론을 제시하고 있는데, 다루고 있는 의미적 관계의 대상이 의학적 도메인에 한정하고 있는 한계가 있다.

앞선 연구들은 e-카타로그 관리에 대한 다양한 이슈를 훌륭히 다루고 있으나 상품 분류 모델은 단순히 UNSPSC와 eCI@ss와 같은 코드 기반 계층 모델에 기반을 두고 있다. [6,7,17]는 e-카타로그가 의미적 모델을 필요로 하는 이슈를 다양한 분류코드 통합 영역에서 설명하고 있고, 상품을 보다 유연하게 보는 것이 필요하고, 의미는 잘 파악되어야 하며 기계가 읽을 수 있는 형태로 표현되어야 한다고 소개하고 있으나, e-카타로그 모델은 부모-자식 클래스 사이의 포함 관계를 찾는 것에 국한하고 있다.

본 연구에서는 확장 개체 관계를 상품 클래스 기본 모델링에 사용한다. [10]는 EER모델에서의 클래스, 속성, 관계를 DL로의 변환가능성을 잘 정리하고 있다.

3. 전자카타로그 모델링

3.1 확장 개체 관계 (Extended Entity-Relationship) e-카타로그의 기본 모델³⁾에 대한 개념적인 표현을 위

2) <http://www.swisslife.com/slcom/en/home.html>

3) 본 연구의 모델링은 전통적인 페이지기반(paper-based)카타로그에 적용될 수도 있다. 논문에서는 두 용어, 페이지기반(paper-based)카타로그와 전자카타로그를 혼용한다.

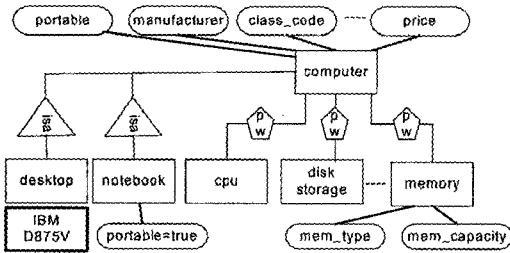


그림 2 EER로 표현된 전형적인 e-카탈로그 단순 모델

해서, 확장 개체 관계를 적용한다(그림 2).

기본 모델링 원소는 상품 객체의 집합인 개체와 개체들 간의 관계이다. 두 개체 사이의 관계는 임의의 의미적 관계로 표현한다. 본문의 e-카탈로그 모델은 클래스 인클루전(is-a)과 전체-부분의 관계 인클루전(part-whole)으로 구성된 인클루전 관계에 초점을 맞추고 있다. 전자는 삼각형으로, 후자는 오각형으로 표기한다. 개체와 관계는 이미 정의된 값, 즉 정수, 실수, 부울린(boolean), 스트링(string)과 같은 속성을 값으로 가질 수 있다. 그림 2는 EER로 표현된 컴퓨터 상품에 대한 전형적인 카탈로그 모델이다. 각 모델의 구성요소는 본문에서 제시하는 DLC(Description Language for Catalog)로 표현한다.

3.2 카탈로그를 위한 설명 언어(DLC)

DL 언어에 카탈로그를 위한 특징을 반영하면, 상품집합 표현을 위한 클래스⁴(C), 상품의 속성, 상품집합사이의 단순한 관계를 표현할 수 있는 롤(R), 카탈로그의 상품집합 사이의 분류계층을 표현하기 위한 인클루전 관계(⊆)를 생각해 볼 수 있다.

DLC에서, 상품 클래스 설명은 아래의 문법 규칙에 따라 정형화 된다. 대문자 A는 원자적 클래스(atomic class)이며 미리 정의된 도메인, 즉, 정수, 실수, 스트링, 부울린 값을 갖는다. 대문자 C 와 D는 상품 클래스 설명이며, 대문자 R과 S는 원자적 롤(atomic role), ⊥는 분류계층의 top과 마찬가지로, 가장 상위의 클래스를 의미한다.

- $C \rightarrow A \mid \perp \mid \neg A \mid CND \mid CUD \mid \forall R.C \mid \exists R.C \mid (\geq n.R.C) \mid (\leq n.R.C)$
- $C \subseteq D$ concept inclusion R^{-} role inverse
- $R \in S$. transitive $R \subseteq S$ role hierarchy

DLC의 의미는 다음과 같이 정의될 수 있다. I는 공집합이 아닌 집합, Δ^I (abstract domain of I)을 구성하는 변환(interpretation) 함수이고, 집합 A 는 $A^I \subseteq \Delta^I$ 이고, 이진 관계인 R에 대해서는 $R^I \subseteq \Delta^I \times \Delta^I$ 로 할당된다.

$$\begin{aligned} \perp^I &= \Delta^I \\ (CND)^I &= C^I \cap D^I \\ (\forall R.C)^I &= \{a \in \Delta^I \mid \forall b, (a,b) \in R^I \rightarrow b \in C^I\} \\ (\geq nR.C)^I &= \{a \in \Delta^I \mid \#\{b, (a,b) \in R^I \wedge b \in C^I\} \geq n\} \\ (R)^I &= \{(a,b) \mid a, b \in \Delta^I \wedge (b,a) \in R^I\} \\ (\neg A)^I &= \Delta^I \setminus A \\ (CUD)^I &= C^I \cup D^I \\ (\exists R.C)^I &= \{a \in \Delta^I \mid \exists b, (a,b) \in R^I \wedge b \in C^I\} \\ (\leq nR.C)^I &= \{a \in \Delta^I \mid \#\{b, (a,b) \in R^I \wedge b \in C^I\} \leq n\} \\ (R \in S)^I &= \{a, c \in \Delta^I \mid \forall b, (a,b), (b,c) \in R^I \rightarrow (a,c) \in R^I\} \end{aligned}$$

변환함수 I는 각각 다음을 만족한다.

$$C \subseteq D \text{ iff } C^I \subseteq D^I$$

role hierarchy R iff $R^I \subseteq S^I$ for each $R \subseteq S$ in R.

DLC에서, 개별적인 상품(클래스의 인스턴스)은 다음의 두 종류의 확인(Assertion), ABox: $C(a) \mid R(a,b)$ 로 설명할 수 있다. 롤 필링 구성요소(role filling constructor)의 의미는 $R^I a$ (또는 $\exists R.(a) \rightarrow \{c \in \Delta^I \mid (c, a^I) \in R^I\}$)이다. ABOX의 의미는 열린 세계(open world)의 의미로 간주하고, 유일 이름 전체(unique name assumption)로 간주한다. 즉,

I 가 가치 함수(valuation function) 일 때 변환 $I = (\Delta^I, I^I)$ 이며,

$a^I \in C^I$ 이면, I 는 $C(a)$ 을 만족하고,

$(a^I, b^I) \in R^I$ 이면, $R(a,b)$ 을 만족한다.

기본적인 도메인 즉 정수, 실수, 스트링, 부울린을 지원하기 위해서는 술어 표현을 위한 명백한 변환을 고려해야 하며, DLC의 의미는 다음과 같이 확장된다. I의 구체적 도메인, Δ^D , 즉, $I = (\Delta^I, \Delta^D, I^I)$.

3.3 DLC와 클래스, 객체와의 대응

상품 클래스는 같은 속성을 갖는 상품 아이템의 집합이다. 각 클래스는 DLC에서 클래스 인클루전(TBOX)으로 표현한다. 각 속성은 원자적이며, 미리 정의된 도메인의 값 중 하나의 값을 갖는다. 만일 클래스 C가 속성 A_1, \dots, A_n 의 단일 값을 갖고, 각 속성은 각각 미리 정의된 도메인인 D_1, \dots, D_n 를 갖는다면, DLC표현은, $C \subseteq \forall A_1.D_1 \cap \dots \cap \forall A_n.D_n \cap \exists A_n$ 이다.(i.e. $Computer \subseteq \forall portable.Boolean \cap \exists portable \cap \forall mfr.String \cap \exists mfr \cap \forall class_code.String \cap \exists class_code \cap \forall price.Integer \cap \exists price$)

수 제한(number restriction) = IA_i 은 A_i 을 의미하며, 단일한 값을 갖는다. 만일 클래스가 다양한 속성 값을 갖는다면, 속성의 최대 카디널리티(cardinality)를 $max A_i$ 로 표현할 수 있다. 미리 정의된 도메인의 속성은 구체적인 형태로 $\leq d, \geq d, > d, < d, \leq d \cap \geq d, = d$ 을 표현한다. DLC에서 최상위 클래스를 \perp 라고 한다. \perp 는

4) DL(description logic) 도메인에서는 클래스를 컨셉(Cecept)이란 용어로 부른다.

모든 카타로그 아이템에 나타나야 하는 공통 속성의 집합을 갖는다. 상품 클래스의 객체는 DLC에서 ABOX로 표현된다.

```

Computer ⊆ ∀portable.Boolean ⊃= Iportable ∩
∀mfr.String ⊃= Imfr ∩ ∀class_code.String ⊃=
Iclass_code ∩ ∀price.Integer ⊃= Iprice,
Computer(IBM875V),
portable(IBM875V, False),
mfr(IBM875V, 'IBM')
class_code(IBM875V, '43211508'),
price(IBM875V, 2500)
    
```

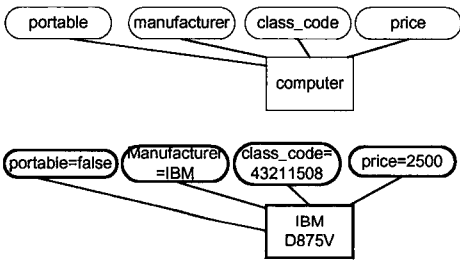


그림 3 클래스와 객체의 DLC 모델링

그림 3은 그림 2의 일부분으로 본 논문 모델의 한 예를 설명해준다. 컴퓨터 클래스는 속성으로 휴대용, 제조사, 분류코드, 가격을 갖고, 그 값은 구체적 도메인에서 정의되는 자료 형태를 갖는다. 컴퓨터 상품 클래스의 객체인 IBM875V는 개념과 롤 확인(assertion)을 사용하여 모델화 된다. EER모델에서 객체는 굵은 사각형으로 표시하고 있다.

4. 의미적 관계(Semantic Relations)

4.1 e-카타로그를 위한 의미적 관계

의미적 관계에 대한 일반적 분석은 인식과학, 언어학, 철학, 논리학, 데이터베이스 등 다양한 학문 영역에서 연구되어 왔다[13,18,19]. [18]에 의한 분류 스키마는 그림 4와 같고 의미적 관계를 분류하는 다양한 관점 중의 하나이며 일반적으로 여러 도메인에서 채택하여 사용한다[13,19]. 이 분류 스키마를 바탕으로 기존의 e-카타로그 코드 시스템인 UNSPSC와 같은 코드를 분석해서, 실제로 절실히 요구되고 있는 의미적 관계가 무엇인가를 보여주고 있다[20]. 그림 4의 의미적 관계가 도메인에서 매우 빈번하게 사용되고 있다.

그림 4를 보면, 상위 레벨에서, 우리는 4개의 다른 의미적 관계 클래스의 분류를 갖고 있다. 인클루전, 속성, 유사어, 그리고 반의어이다. 인클루전 관계는 개체(상품

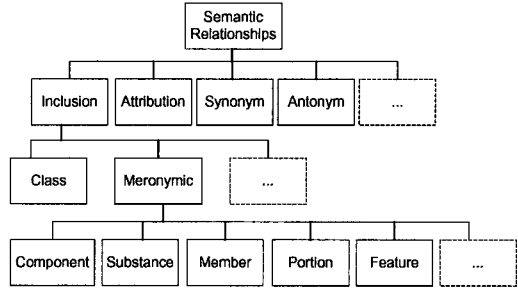
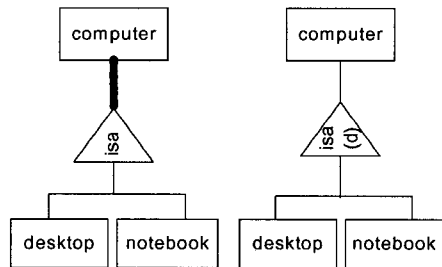


그림 4 의미적 관계의 분류

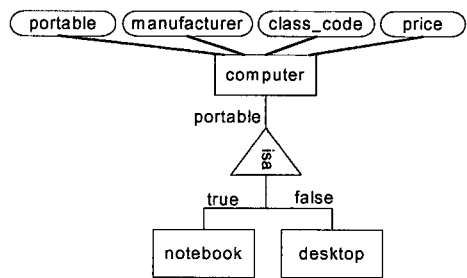
클래스)가 또다른 개체를 포함하거나 조합하는 경우를 의미한다. 이것은 클래스 인클루전 혹은 부분-전체 관계 인클루전으로 분류된다. 전자는 서브-슈퍼 타입 관계로 표현되며 데이터베이스 모델링에 자주 나타난다. 후자는 부분과 전체의 관계를 나타낸다.

4.2 클래스 인클루전(Class Inclusion)

C가 D의 부분임을 표현하는 클래스 인클루전 관계는, DLC에서 $C \subseteq D$ 로 표현하며 is-a 관계를 의미하고 EER에서는 삼각형으로 표기한다. 속성인 portable은 식별요소(discriminator)로 작용한다. DLC로는 $Notebook \subseteq Computer \cap \forall portable.(true)$ 로 표현한다.



(a) 제약의 커버/디스조인트



(b) 컴퓨터 클래스 인클루전

그림 5

클래스 인클루전은 커버(covering)와 디스조인트(disjoint)같은 추가적인 제약을 가질 수 있다.: $C \subseteq D_1 \cup D_2 \cup \dots \cup D_n$ (서브클래스 D_1, \dots, D_n 를 C가 커버한다.)

서브타입인 D_i 와 D_j 에 대해, $D_i \subseteq \neg D_j$. (서브클래스 D_1, \dots, D_n 는 각각 디스조인트한다.)

4.3 부분-전체 관계(part-whole relations)

클래스 C 와 클래스 D 에 대한 전체-부분의 논리 인클루전(meronymic inclusion)은 부분-전체의 관계를 의미한다. 즉, 클래스 C 는 클래스 D 의 부분이고 다시 말해 클래스 D 는 클래스 C 를 인클루전한다. 부분-전체 관계는 의미면에서 몇 가지 구분이 있다[18]. 예를 들어, HDD와 CPU는 컴퓨터의 부품(component)을 의미한다. 양파, 마늘, 쇠고기는 쇠고기스튜의 재료(substance)를 의미하는 부분-전체 관계이다. 부분-전체의 관계가 의미적으로 다르게 인식되므로, 부분-전체를 나타내는 속성이 달라질 수 있다.

부분-전체의 일반적으로 그 특성들은 다음과 같다 [13].

pw 가 part-whole를 나타낼 때:

- *Transitive*: $x pw y \ \& \ y pw z \Rightarrow x pw z$,
- *Inverse*: $x pw_1 y \Rightarrow y pw_2 x$, where $((pw_1^{-1})^{-1}) = pw_2$,
- *Asymmetric*: $x pw y \Rightarrow \neg (y pw x)$,
- *Identical*: $x pw l_1, \dots, l_p \ \& \ y pw m_1, \dots, m_q \ \& \ l_1 = m_1, \dots, l_p = m_q \Rightarrow x = y$
- *Atomic*: $x_1 pw x_2 \ \& \ \dots \ \& \ x_i pw x_{i+1} \ \& \ \dots \ \& \ x_{N-1} pw x_N$ 를 만족하는 정수 N 항상 존재한다.

클래스 인클루전 관계의 가장 중요한 특성중의 하나인 부분-전체 관계를 모델에서 볼 수 있듯이, 클래스의 인클루전 관계가 정확히 포함(subsume) 제약을 의미하는 것은 아니다.

[18,19]에 의해 처음 제안된 표준을 사용하여 부분-전체 관계 형태를 구별하면, e-카타로그 도메인에서 다룰 만한 다섯 가지 형태로 표 1과 같이 간추릴 수 있다 [20].

5. 적용 시나리오의 한 예: 관세청 전자카타로그 모델링

이 장에서는 본 논문의 모델링이 e-카타로그 분야에

적용했을 때 어떻게 사용될 수 있는지 시나리오를 통해 살펴본다. 이 시나리오는 상품 객체(ABOX)와 구체적 도메인 표현을 사용해서 서론에 소개된 시나리오보다 좀더 복잡하면서 실용적이다.

관세청에서 수출금지 품목을 정하고, 그것을 규제한다고 가정해보자. 한 전자 회사인 IBM에서 지름이 0.1 인치 이하인 모터(DC spindle Moter)를 개발한다고 하자. 그런데 이 제품은 강력한 파괴 장치의 핵심 부품이고, 악의적 목적에 사용될 가능성이 있어서 이를 막으려 한다면, 당국은 이 제품과 이것을 포함한 모든 상품을 수출 금지해야 할 것이다. 해당모터와 이것을 부품으로 갖는 모든 제품이 수출금지항목이며, 이러한 항목을 빠짐 없이 찾아낼 수 있도록 모델링한 것이 그림 6과 같다.

그림 6(a) 모델에서 해당 제품, 즉 제조회사(mfr)가 IBM이고, 지름(diameter)가 0.1 이하인 제품은 IBM-SM100이며, 곧 수출금지항목이 된다. 이것을 부품으로 사용하고 있는 제품인 IBMHDD와 HPLBP도 역시 수출금지 항목이며, 마찬가지로 IBMHDD를 부품으로 갖는 SONYVAIOP25도 수출금지 항목이다. 그림 6(b)의 DLC에서 IBMSM100는 모터(DC spindle moter) 클래스의 객체이며 수출 금지 조건을 만족한다.

$$Spindlemotor \cap \forall Mfr.\{IBM\} \cap \forall diameter.\{=0.1\}$$

IBMSM100은 IBMHDD의 부품으로 사용하고, IBM-HDD은 SONYVAIOP25 상품의 부품으로 사용되었다.

$$Hdd(IBMHDD) \subseteq Hardware \cap \exists compose-Of.\{IBMMSM100\}$$

$$Lbp(IBM LBP) \subseteq Printer \cap \exists compose-Of.\{IBMMSM100\}$$

$$Computer(SONYVAIOP25) \subseteq Hardware \cap \exists compose-Of.\{IBM-HDD\}$$

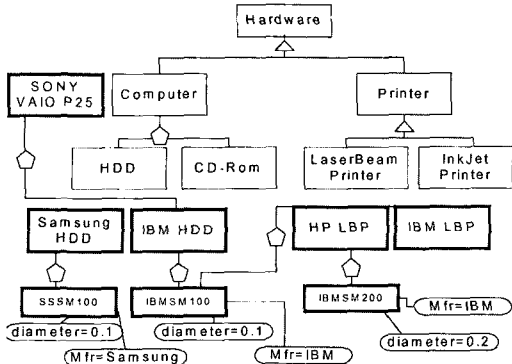
수출금지 항목(Noexport)을 나타내는 클래스는 다음과 같다.

$$(Spindlemotor \cap \forall Mfr.\{IBM\} \cap \forall diameter.\{ \leq 0.1\}) \cup (\exists Compose-Of.(Spindlemotor \cap \forall Mfr.\{IBM\} \cap \forall diameter.\{ \leq 0.1\})) \subseteq Noexport$$

기존의 카타로그로는 컴퓨터, HDD, 스피들 모터 사이의 compose-of 관계와 compose-of 관계가 이행적 특성이 있음을 형식적(formal)으로 표현할 수 없고, 그렇기 때문에 시나리오에서 가정한 문제에 대한 해결 방법은 결코 쉽지 않다. 하지만 우리가 제시하는 온톨로지는

표 1 부분-전체 의미적 관계의 형태와 분류 기준

Types	Examples in catalog domain	Classification criter		
		functional	separable	homogeneous
<i>component-of</i>	HDD <i>component-of</i> computer	yes	yes	no
<i>substance-of</i>	beef (garlic and onion) <i>substance-of</i> beef-stew	no	no	no
<i>member-of</i>	umbrella (hat or belt) <i>member-of</i> clothing accessory	no	yes	no
<i>portion-of</i>	cheese cake slice <i>portion-of</i> cheese cake	no	yes	yes
<i>feature-of</i>	paying <i>feature-of</i> shopping	yes	no	no

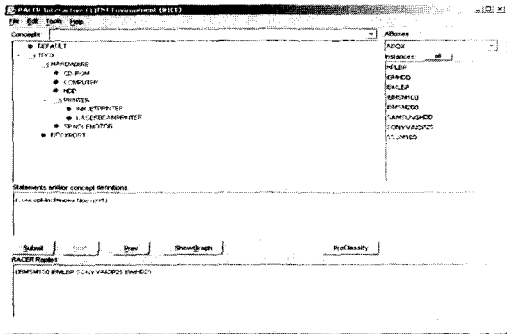


(a) EER로 표현한 관세청의 e-카타로그 모델의 일부분

$Computer \cup Printer \sqsubseteq Hardware, Lbp \cup Inkjetprinter \sqsubseteq Printer,$
 $Computer \sqsubseteq Hardware \cap \exists compose-Of_i.Hdd \cap \exists compose-Of_i.Cdrom,$
 $Hdd \sqsubseteq Hardware \cap \exists compose-Of_i.Spindlemotor,$
 $Lbp \sqsubseteq Hardware \cap \exists compose-Of_i.Spindlemotor,$
 $Spindlemotor \sqsubseteq Hardware \cap \forall Mfr.string \cap =1Mfr \cap \forall diameter.real$
 $\cap =1diameter, compose-Of_i \in S^+,$
 $Hdd(SAMSUNGHDD), Hdd(IBM HDD), Computer(SONYVAIOP25),$
 $Spindlemotor(SSSM100), Spindlemotor(IBM SM100), Lbp(HPLBP),$
 $Lbp(IBM LBP), compose-Of_i(SAMSUNGHDD, SSSM100), compose-Of_i$

(b) DLC의 모델 표현의 일부분

그림 6



(a) Racer의 screen snapshot

(in-knowledge-base Tbox Abox) (signature :atomic-concepts
 (Hardware Computer Hdd Cd-rom Printer Lbp Inkjet Spindlemotor
 Noexport)
 :roles ((compose-Of :transitive t :inverse component-Of))
 :individuals (SONYVAIOP25 SAMSUNGHDD IBM HDD
 SSSM100 IBMSM100 IBMSM200 HPLBP IBMLBP)
 :attributes ((string Mfr) (real diameter)) :objects(100size 200size
 Co SSCo))

(b) Racer에서의 모델 표현의 일부

그림 7

모델링에서는 이들 간의 관계 정보를 DL로 명확히 표현할 수 있고, 이를 다시 OWL이나 DAML-S와 같은 웹 기반 온톨로지 언어로도 변경할 수 있으며, DL 엔진을 사용하면 추론을 통해 문제를 해결할 수 있게 되는 것이다.

그림 7은 그 한 예로서 제시된 모델이 Racer[7] 추론 도구를 사용하여 질의 처리를 하는 것을 보여주고 있다. 그림 7(b)에서는 앞서의 DL로 표현된 모델과 수출 금지 항목에 해당하는 품목, 즉 지름 0.1 인치 이하인 IBM 모터를 가지는 모든 제품을 찾는 질의(*concept-instance Noexport*)이 Racer에서 어떻게 표시되는지를 보여주고 있고, 그림 7(a)에서는 실제로 그 질의처리 결과가 해당하는 제품인 SONYVAIOP250, IBMHDD, IBMLBP, IBMSM100 를 추론해 내고 있음을 보여주고 있다.

6. 논의 사항

본 논문의 e-카타로그 모델인 DLC는 SHIQ(d)에 기반한다. SHIQ(d)는 ALC의 확장으로 이행적 불, 를 계

층관계, 역, 수 제한 그리고 구체적인 도메인을 표현할 수 있는 언어이다. DLC언어가 SHIQ(d) 표현을 기반으로 하게 된 이유 중의 하나는 상품간의 부분-전체 관계를 표현하기에 적합하기 때문이다.

문헌에서 다루는 전자상거래를 포함한 시멘틱 웹 응용의 맥락에서 정수, 부울린, 스트링과 같은 자료 객체는 절실히 필요하다. 구체적인 자료 형태를 갖는 객체는 잘 정의되고 표현될 수 있어야 한다[4,12]. DAML+OIL와 OWL은 모두 XML-Schema를 기반으로 한 자료 형태에 의존하고 있다. 그래서 전자상거래 영역에서 실제로 요구되는 선형적 다항 방정식(linear polynomial equations)과 같은 자료 값의 제약을 제공할 수 없다. 실제 응용에서 그림 8에서처럼 지름에 대한 속성값의 단위를 다르게 사용하기도 한다.

그림 8(a)에서 LGHDD는 부품으로 IBMSM-1100 모터를 갖고, 그 모터는 센티미터로 지름 값이 표현되었다(예 $diameterM = 0.254$). 인치를 센티미터로 변환하는 방정식을, $diameterM = 2.54 * diameter$, 사용하여 LGHDD는 Noexport 품목으로 추론할 수 있다. 그림

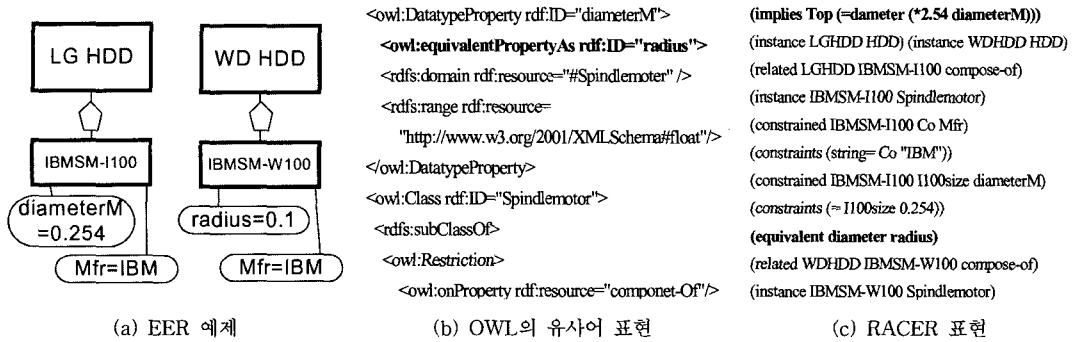


그림 8 자료 값 제약: 측정 단위 변환과 유사적 특성

8(c)는 Racer에서의 모델 표현이다. 서로 다른 단위 문제(미터, 야드, 킬로그램, 관 등)는 e-카타로그 영역에서 빈번하게 다루어야 하는 문제이다. DL의 관점에서 DAML+OIL 이나 OWL은 많은 부분에서 DL과 공통된 특징을 지녔다. 그래서 향후의 연구 방향이 OWL을 기반으로 해야 한다는 주장이 설득력 있다. 앞서 언급한 측정 단위 문제는 e-카타로그 영역에서 흔히 발생하는 경우로 고려되어야 한다. 그림 8(a)에서 WD HDD는 스피너 모터의 속성으로 diameter를 지칭하는 radius란 용어를 사용한다. 두 속성사이의 유사관계는, $radius \equiv diameter$ 라고 표현하고, OWL에서는 *equivalentPropertyAs axiom* 으로 표현될 수 있다(그림 8(b)).

7. 결론

본 논문에서는 e-카타로그를 온톨로지 모델링을 제시하였다. 제시한 방법은 확장 관계(EER) 표현과 DL을 사용한 형식적(formal)인 모델화이다. 본 논문의 공헌은 다음과 같이 요약할 수 있다.

첫째, 전자카타로그 영역에서 논리적이고 형식적(formal)인 DL로 온톨로지 모델링을 최초로 시도하였다.

둘째, 개념적, 시각적으로 온톨로지 모델을 할 수 있도록 기본적인 모델링에는 EER을 사용하였고 EER과 DL과의 관계를 제시하였다.

셋째, 온톨로지 모델의 제시에 그치지 않고 모델의 시험적 검증을 위해 실제 추론 도구를 사용하여 실행해 그 가능성을 검증해 보았다. 또한 온톨로지 모델링의 활용 이해를 돕기 위해 시나리오를 제시하였다.

현재, 우리가 제시하는 모델링 언어의 구성요소는 실제 전자상거래 도메인에서 체험된 요구사항들을 수용하고 있다. 제시한 모델링 언어로 표현할 수 없는 새로운 정보를 카타로그에 추가해야 한다면 언어의 확장이 필요하다. 이를 위해서는 그에 따른 모델링 가능범위와 추론의 복잡도에 대한 분석 연구가 필요하다. 따라서 관련

연구들을 통해 모델을 확장하고 DLC 언어의 완성도를 높여가는 것이 향후 연구 방향이 될 것이다.

참고 문헌

- [1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, Vol. 284, 2001.
- [2] P. Spyns, R. Meersman, and M. Jarrar, "Data modeling versus Ontology engineering," *SIGMOD Record*, Vol. 31(4), ACM, 2002.
- [3] F. van Harmelen, P.F. Patel-Schneider, and I. Horrocks (Editors), Reference description of the DAML+OIL ontology markup language, <http://www.daml.org/2001/03/reference>, 2001.
- [4] V. Haarslev and R. Miller, "Description Logic Systems with Concrete Domains: Applications for the Semantic Web," *Proc. of the 10th International Workshop on Knowledge Representation meets Databases (KRDB 2003)*, CEUR Workshop Proceedings, 2003.
- [5] D. Fensel, Y. Ding, B. Omelayenko, E. Schulten, G. Botquin, M. Brown, and A. Flet, "Product Data Integration in B2B E-Commerce," *IEEE Intelligent Systems*, Vol. 16, IEEE Society, 2001.
- [6] G. Yan, W. K. Ng, and E. Lim, "Product Schema integration for Electronic Commerce - A Synonym Comparison Approach," *IEEE Transaction on Knowledge and Data Engineering*, Vol. 14(3), IEEE Society, 2002.
- [7] D. Kim, J. Kim, and S. Lee, "Catalog Integration for Electronic Commerce through Category-Hierarchy Merging Technique," *Proc. of the 12th International Workshop on Research Issues on Data Engineering (RIDE2002)*, IEEE Society, 2002.
- [8] M.K. Smith, C. Welty, and D.L. McGuinness, "OWL Web Ontology Language Guide W3C Recommendation," <http://www.w3c.org/TR/owl-guide/>, 2004.
- [9] R. Elmasri and S.B. Navathe, *Fundamentals of Database Systems*, 4th Edition, Pearson Addison

- Wesley, 2003.
- [10] D. Calvanese, M. Lenzerini, and D. Nardi, "Unifying Class-Based Representation Formalisms," *Journal of Artificial Intelligence Research*, Vol. 11, AI Access Foundation and Morgan Kaufmann Publishers, 1999.
 - [11] F. Baader, I. Horrocks, and U. Sattler, "Description Logics as Ontology Languages for the Semantic Web," *Knstliche Intelligenz*, Vol. 16, 2002.
 - [12] I. Horrocks and U. Sattler, "Ontology Reasoning in the SHOQ(D) Description Logic," *Proc. of 17th International Joint Conference on Artificial Intelligence(IJCAI01)*, American Association for Artificial Intelligence, 2001.
 - [13] U. Sattler, "Description Logics for the Representation of Aggregated Objects," *Proc. of the 14th European Conference on Artificial Intelligence*, IOS Press, 2000.
 - [14] I. Horrocks, "Benchmark Analysis with FaCT," *Proc. of TABLEAUX 2000*, Springer-Verlag, 2000.
 - [15] Z. Cui, D. Jones, and P. O'Brien, "Semantic B2B Integration: Issues in Ontology-based Approaches," *SIGMOD Record*, Vol. 31(1), ACM, 2002.
 - [16] S. Schulz, U. Hahn: *Building a Very Large Ontology from Medical Thesauri. Handbook on Ontologies 2004*.
 - [17] D. Kim, *A Semantic Model for Electronic Catalogs*, Ph.D. dissertation, Seoul National University, 2004.
 - [18] M.E. Winston, R. Chaffin, and D. Herrmann, "A taxonomy of part-whole relations," *Cognitive Science*, Vol. 11, Cognitive Science Society, 1987.
 - [19] V. C. Storey, "Understanding Semantic Relationships", *VLDB Journal*, Vol. 2, VLDB Endowment, 1993.
 - [20] H. Lee and J. Shim, "When Ontology meets e-Catalog," *e-biz World Conference 2004*, Society for e-Business Studies in Korea, 2004.



이 현 자

1996년 2월 숙명여자대학교 전산학과(학사). 2004년 8월 숙명여자대학교 대학원 컴퓨터과학과(석사). 2004년 9월~현재 숙명여자대학교 대학원 컴퓨터과학과 박사과정. 관심분야는 데이터베이스, 전자상거래, 온톨로지

심 준 호

정보과학회논문지 : 데이터베이스
제 32 권 제 1 호 참조