

전자카탈로그에서의 의미적 관계 분석과 모델링

Analysis and Modeling of Semantic Relationships in e-Catalog Domain

이민정(Lee, Minjung), 이현자(Lee, Hyunja), 심준호(Shim, Junho)*

초 록

도메인별로 적합한 온톨로지를 구축하는 것은 시맨틱웹의 한 구현 방법으로서 해당 도메인 응용프로그램의 의미적 정보를 풍부히 해준다. 전자상거래에서 카탈로그는 상품과 서비스에 대한 가격, 특성, 조건 등의 다양한 정보를 저장 관리해주는데, 카탈로그에는 상품 개별 정보뿐 아니라 상품간에도 다양한 정보가 내재되어 존재하게 된다. 따라서 전자카탈로그 영역에 온톨로지를 적용하기 위하여는 먼저 카탈로그 영역에서 존재하는 의미적 관계(semantic relationships)를 분석해 보는 것은 의미 있는 일이다. 본 논문에서는 카탈로그의 의미적 관계를 분류체계를 통해 분석해보고, 각각의 관계가 어떻게 온톨로지 모델화 될 수 있는지를 제시한다. 모델링 기법으로서는 기본적으로 EER(Extended Entity-Relationship)을 사용하는데, 이에 제한되지 않고 궁극적으로 온톨로지 모델을 사용하여 추론을 할 수 있도록 모델을 Description Logics으로 표현한다.

ABSTRACT

Building a domain-suited ontology, as a means to implement the Semantic Web, is widely believed to offer users the benefit of exploiting the semantic knowledge constrained in the application. Electronic Catalog, shortly e-Catalog, manages the information about the goods and services offered to or requested by the trade partners and, such information as pricing, features or conditions play an important role in e-commerce domain. Consequently, semantically enriched yet precise information by the ontology may elaborate the business transactions. In this paper, we analyze the semantic relationships embodied within the catalog domain, as the first step towards the ontological modeling of e-catalog. Exploring ontology should leverage not only the representation of semantic knowledge but also provide the inferencing capability for the model. We employ the EER(extended Entity Relationships) for the basic model. Each modeling construct can be directly translated by DL(Description Logics). Semantic constraints that can be hardly represented in EER are directly modeled in DL.

키워드 : 전자카탈로그 전자상거래, 온톨로지, 확장 개체-관계 모델

Description Logics e-Catalog, e-commerce, Ontology, EER, Description Logic

본 연구는 숙명여자대학교 2004년도 교내연구비 지원에 의해 수행되었음.

* 숙명여자대학교 정보과학부 부교수

1. 서 론

전자상거래에서 전자카탈로그(e-catalog)는 전자상거래 시스템의 시스템 컴포넌트로서 중요한 연구 주제이다. 카탈로그는 상품과 서비스에 대한 가격, 특성, 조건 등에 관한 정보를 저장 관리하고, 거래 파트너간의 협상과 화해(settlement)에 있어서 중요한 역할을 한다. 이러한 비즈니스 트랜잭션(business transactions)에 이용되는 카탈로그에 의미적으로 풍부하고 정확한 정보가 더해질 때 트랜잭션의 효과를 향상시킬 수 있을 것이다. 전자카탈로그 영역에 온톨로지를 적용할 수 있는 모델을 작성하는 것은 이러한 측면에서 매우 의미 있는 일이다.

그런데 정확한 의미적 추론(inference)이나 추리(reasoning)를 하기 위해 먼저 모델이 그에 따른 의미를 명확히 부여할 수 있어야 한다. 그러나 현재 일반적으로 전자카탈로그 모델로서 대표적으로 사용되는 분류체계 모델을 예로 보더라도, 하나 이상의 서로 다른 카테고리에 속하는 의미적 관계들이 단순한 인클루전(inclusion)으로 혼용되어 사용되고 있다. 따라서 우리는 모델에서 표현되어야 할 의미적 관계를 UNSPSC 카탈로그 코드 체계를 분석해 봄으로서 현실세계에서 어떤 의미적 관계가 필요에 의해 은연중에 사용되었는지를 분석하였다. 이과정을 통해 우리는 전자카탈로그영역에 존재하는 다양한 의미적 관계를 분석하고 어떤 관계가 모델에 표현되어야 하는지를 체계화하였다.

전자카탈로그에 온톨로지를 적용하는 아이디어는 비록 새로운 것은 아니지만, 현재까지

의 선행연구를 살펴보면 아직까지 실제적으로 전자카탈로그에 온톨로지를 체계적으로 적용하여 사용된 사례는 없다. 본 논문에서는 이를 위한 모델링 기법으로서 기본적으로 EER(Extended Entity Relationship)을 사용하여 궁극적으로 온톨로지 모델을 이용하여 추론을 할 수 있도록 DL(Description Logics) 시멘틱 모델링 기법을 개발하였다. 또한 DL로서 온톨로지 모델을 표현하도록 우리가 설계한 DLC(Description Language for Catalog)를 이용하였다[14].

다음은 논문의 구조이다. 먼저 2장에서는 의미적 관계에 관한 기존의 연구에 대한 조사를 보이고 우리 모델을 간략히 제시한다. 3장에서는 본 논문의 핵심인 전자 카탈로그에 필요한 의미적 관계가 무엇인지 UNSPSC를 통해 살펴본 우리의 분석 결과를 제시하고 이러한 의미적 관계를 우리 모델에서 어떻게 표현하는지 살펴본다. 4장에서는 분석을 통해 우리가 얻은 경험과 몇 가지 관련연구에 대한 가이드를 제공한다. 마지막으로 5장에서 본 연구의 의의와 연구 내용을 정리해 본다.

2. 배경 지식

2.1 의미적 전자 카탈로그 모델

UNSPSC[21], eCI@ss[7], UCC[20]와 같이 상품과 서비스의 정보를 모델링하는 몇몇 시스템은 이미 존재 하며, 이 시스템들은 상품 클래스 계층과 상품 분류와 같은 콘텐츠 표준을 제공한다. 그러나 상품 카테고리는 상

품과 서비스에 대해 기계적으로 운영하기 위해 필요한 많은 정보 집합 중 하나일 뿐이다. 현재의 표준은 eOTD, GDD와 같이 기술사전의 형태이고/거나 분류코드가 너무 한정적이어서 수백의 상품 클래스, 클래스 속성, 속성과 클래스의 의미(혹은 제약) 사이의 복잡한 관계를 표현하기에는 부족하다.

최근의 연구동향을 보면 전자상거래 영역에 온톨로지를 적용하는 것은 잠재적인 장점이 있고[4], 특히 전자카탈로그에서는 그 효과가 큰 영역이라 생각된다. [6]은 상품 정보에 초점을 맞추어 B2B통합 이슈와 상품 정보 구축, 유지, 통합의 측면에서 어려운 점을 고려하였다. [15]는 eCI@ss에 대한 구체적인 모델을 표현하고, 분류 스키마에 대한 이슈를 다루었다. 이상의 연구에서는 전자카탈로그 관리의 다양한 이슈들을 다루었으나 상품 분류의 근본적인 모델을 UNSPSC, eCI@ss, HS 등의 간단한 코드기반 계층 모델로 가정하였다. [12, 23]은 전자카탈로그영역에서 시멘틱 모델이 필요한 이유를 다양한 분류코드 통합의 영역에서 고려하고 있다. 그러나, 의미적 인클루전 관계(semantic inclusion relationship)를 최소화하고, 부모-자식 클래스간의 포함관계를 찾는 데 중점을 두었다. 또한 [23]은 유의어, 반의어와 같은 상품 클래스간의 의미적 관계의 특별한 타입만을 다루었다.

본 논문에서는 실제적인 분류의 예, 즉 UNSPSC를 분석해 봄으로서 전자카탈로그 영역에 존재하는 다양한 의미적 관계를 분석하고 어떤 관계가 모델로 표현되어야 하는지를 체계화하였고, 우리가 [14]에서 제안한 DLC언어를 이용하여 표현하였다. 본 연구에

서 분석한 UNSPSC분류체계는 UNSPSC v6.0315 unified version이며, 저자들의 도메인 지식을 이용하여 다른 세그먼트에 비해 상대적으로 의미적 분석이 용이한 세그먼트를 집중 분석하였다. 분석한 세그먼트는 다음과 같다. :

- Information Technology Broadcasting and Telecommunications(43000000)세그먼트
- Office Equipment and Accessories and Supplies(44000000)세그먼트
- Food Beverage and Tobacco Products(50000000)세그먼트
- Apparel and Luggage and Personal Care Products(53000000)세그먼트

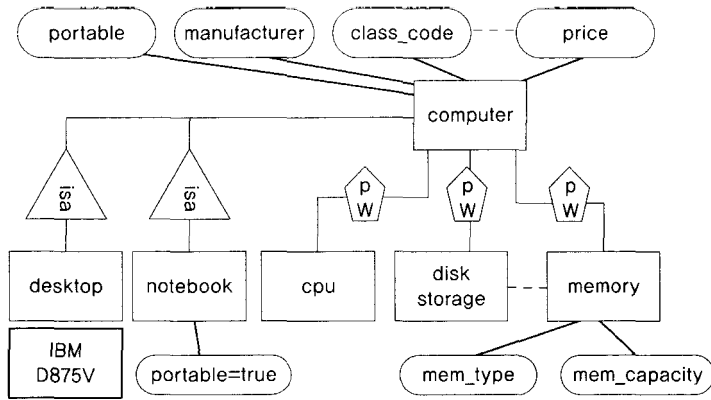
분석을 통하여 얻어진 의미적 관계를 전자카탈로그영역에서 시멘틱 모델로 표현함에 있어서 상품 정보를 개념(concepts)과 공리(axioms)가 일치되는 집합을 이용하여 체계적인 지식을 표현할 수 있도록 한다. 모델링 기법으로서 기본적으로 EER을 사용하며, [3]에서 제안된 스키마를 이용하여 클래스, 속성 및 관계를 DL로 변환하고, 추가적인 의미적 제약은 직접 정의한 DL문법으로 표현하여 확장하였다.

2.2 온톨로지와 DL(Description Logics)

2.2.1 EER(Extended Entity Relationship)

전자카탈로그 기본 모델¹⁾에 대한 개념적인 표현을 위해 EER을 적용하였다. <그림 1>

기본 모델링 원소(element)는 상품 객체(object) 집합인 개체(entity)와 개체간의 관



〈그림 1〉 EER로 표현된 전형적인 전자 카탈로그 단순 모델

제이다. 두 개체간의 관계는 임의의 의미적 관계로 표현한다. 본문의 전자카탈로그 모델은 클래스 인클루전(class inclusion : is-a)과 전체-부분의 관계 인클루전(meronymic inclusion : part-whole)으로 구성된 인클루전 관계에 초점을 맞추고 있다. 전자는 삼각형으로 후자는 오각형으로 표기한다. 개체와 관계는 정수, 실수, 부울린, 스트링과 같이 이미 정의된 값의 속성을 가질 수 있다. 〈그림 1〉은 EER로 표현된 컴퓨터 상품에 대한 전형적인 카탈로그 모델이다. 각 모델의 구성원은 다음 절에서 설명할 언어 DLC(Description Language for Catalog)로 표현한다.

2.2.2 DLC(Description Language for Catalog)

DLC에서 상품 클래스는 아래의 문법 규칙에 따라 정형화 된다. A는 원자적 클래스(atomic class)를 의미하며 미리 정의된 도메인 즉, 정수, 실수, 부울린, 스트링 값을 갖는다. C와 D는 상품 클래스를, R과 S는 원자적 역할(atomic role)을, \top 은 최상위 상품 클래스(universal top product class)²⁾를 각각 의미한다.

DLC의 의미는 다음과 같이 정의될 수 있다. I 는 공집합이 아닌 집합, A^I (I 의 추상 도메인)을 구성하는 변환함수(interpretation function)이다. 즉, 집합 A 는 $A^I \subseteq A^I$ 로 이

- $C \rightarrow A \mid \top \mid \neg A \mid C \cap D \mid C \cup D \mid \forall R.C \mid \exists R.C \mid (\geq n.R.C) \mid (\leq n.R.C)$
- $C \subseteq D$ *concept inclusion*
- R *role inverse*
- $R \in S$ *transitive*
- $R \subseteq S$ *role hierarchy*

진 관계(binary relation) R 은 $R^I \subseteq \mathcal{A}^I \times \mathcal{A}^I$ 로 각각 할당된다.

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

변환함수 I 는 ' $C \subseteq D$ 이면 $C^I \subseteq D^I$ 이다.'의 필요충분조건을 만족한다. 마찬가지로 I 는 역할계층 R 에 대해 ' $R \subseteq S$ 인 모든 경우에 대해 $R^I \subseteq S^I$ 이다.'를 만족한다. DLC는 ABOX에서의 개별적인 상품을 다음의 두 종류의 전제(assertion) $C(a)$ 또는 $R(a,b)$ 로 나타낸다. 우리는 ABOX의 의미를 열린 세계(open world)와 유일 이름 전제(unique name assumption)로 간주한다. 즉, I 가 평가 함수(valuation function)일 때, I 는 다음과 같다. ' $I = (\mathcal{A}^I, \cdot)$. I 는 $a^I \subseteq C^I$ 일 때 $C(a)$ 를 만족하고 이것은 $(a^I, b^I) \in R^I$ 일 때 $R(a,b)$ 를 만족한다.' 정수, 실수, 스트링, 부울린과 같은 기본적인 도메인을 지원하기 위해서 술어 표현을 위한 변환을 고려해야 하며, DLC의 의미는 구체적인 도메인 I, \mathcal{A}^I . 즉, $I = (\mathcal{A}^I, \mathcal{A}^I$

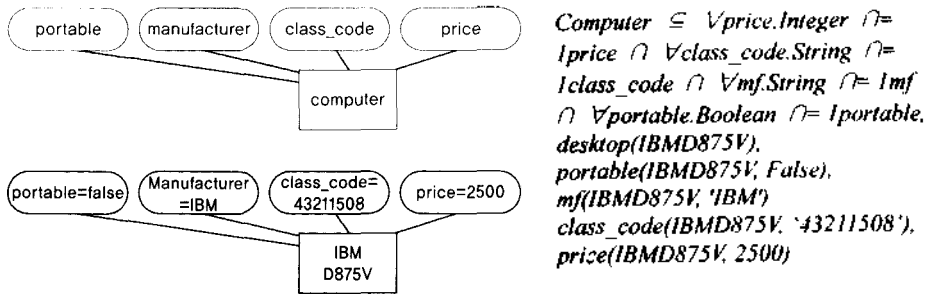
$D, \cdot)$ 를 포함하는 것으로 확장된다.

2.2.3 DLC에서 클래스와 객체

상품 클래스는 같은 속성 집합을 갖는 상품 객체(혹은 아이템)의 집합이다. DLC에서 각 클래스는 개념 인클루전(concept inclusion : TBOX)으로 표현한다. 각 속성은 원자적이며, 미리 정의된 도메인의 값 중 하나의 값을 갖는다. 만일 클래스 C 가 속성 A_1, A_2, \dots, A_n 와 같은 단일 값을 갖고 각 속성은 각각 D_1, D_2, \dots, D_n 를 갖는다면, DLC에서의 표현은 ' $C \subseteq \forall A_1.D_1 \cap = 1A_1 \forall A_2.D_2 \cap = 1A_2 \cap \dots \cap \forall A_n.D_n \cap = 1A_n$ '이다. 수제한(number restriction) 표현, ' $= 1A_i$ '는 A_i 가 단일 값을 의미한다. 만약 클래스가 다중 값 속성이라면, 속성의 최대 카디널리티(cardinality)는 ' $\max A_i$ '로 표현한다. 미리 정의된 도메인의 속성은 $\leq d, \geq d, > d, < d, \geq d \cap \geq d, = d$ 의 구체적 형태로 표현한다. DLC에서 최상위 클래스를 \neg 라 하며, \neg 은 모든 카탈로그 아이টে 나타나는 공통 속성 집합을 갖고 있다. 예를 들어, 모든 상품이 분류코드와 가격에 대한 값을 갖고 있다면, \neg 은 $\{\text{class_code, price}\}$ 를 갖는다. DLC에서 각 개별 상품 객체는 ABOX로 표현된다.

<그림 2>는 <그림 1>의 일부분으로 우리의 모델에 대한 한 예를 보여준다. 컴퓨터 클래스는 속성으로 휴대용, 제조사, 분류코드, 가격을 갖고, 그 값은 구체적 도메인의 자료 형태를 갖는다. 컴퓨터 상품 객체인 *IBMD875V*는 개념과 역할 전제를 사용하여 모델화된다. 굵은 실선은 EER 모델에서의 객체를 나타낸다.

- 1) 본 연구의 모델링의 전통적인 페이지 기반(paper-based) 카탈로그에 적용될 수도 있다. 본 논문에서는 두 용어 페이지 기반(paper-based) 카탈로그와 전자카탈로그를 혼용한다.
- 2) DL 도메인에서 클래스는 컨셉(concept)이라 부른다.



〈그림 2〉 DLC에서의 클래스와 객체 모델

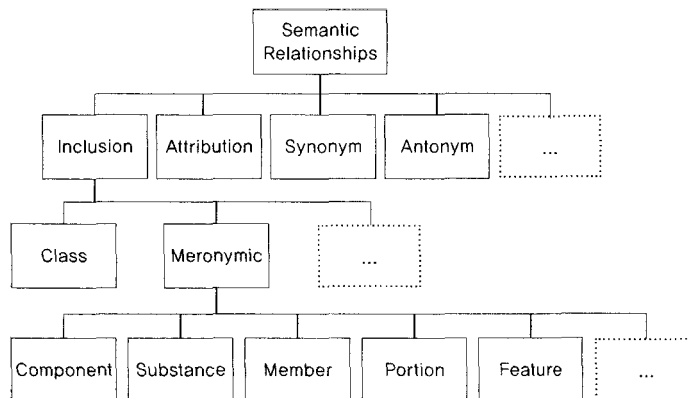
3. 전자카탈로그 도메인에서 의미적 관계

3.1 의미적 관계의 분류

어떤 의미적 모델링을 응용 도메인에 충족시키기 위해 의미적 관계가 도메인에 바람직한 것인지 조사하고 그것을 모델에서 획득할 수 있도록 해야 한다. 의미적 관계에 대한 일반적 분석은 인식과학, 언어학, 철학, 논리학, 데이터베이스 등 다양한 학문의 영역에서 연구되어 왔다[22, 16, 19]. [22]의 분류 스키마

는 의미적 관계를 분류하는 다양한 관점 중 하나이며 일반적으로 여러 도메인에서 채택하여 사용된다[16, 19]. 우리는 [16, 19]를 기반으로 UNSPSC와 같이 현존하는 전자카탈로그를 분석하여 의미적 관계를 분석하였고 그 결과, 〈그림 3〉으로 표현되는 도메인에서 가장 유용한 의미적 관계들을 찾았다.

상위 레벨에서 인클루전(inclusion), 속성(attribution), 유의어(synonym), 반의어(antonym)와 같이 4개의 서로 다른 의미적 관계로 분류할 수 있다. 인클루전 관계는 개체 타입(즉, 상품 클래스)이 다른 개체 타입



〈그림 3〉 의미적 관계의 분류

을 구성하거나 포함하는 경우를 의미한다. 이것은 클래스 인클루전과 부분-전체 인클루전(meronymic inclusion)으로 분류된다. 전자는 서브-슈퍼타입 관계로 표현되며 데이터베이스 모델링에서 자주 나타난다. 후자는 전체와 부분 관계를 나타낸다. 전자카탈로그영역에서 각각의 의미적 제약은 다음의 세부 절에서 분석하였다. 속성관계는 개체 타입이 또 다른 개체 타입의 속성이나 특성을 설명하는 상황을 나타낸다. 유의어와 반의어 관계는 개체 타입이 다른 개체 타입에 대해서 유사하거나 상대적인 의미를 지니는 것을 나타내며 이는 속성, 개체 혹은 관계에 나타날 수 있다. 우리 모델에서는 속성과 개체에서 유의어 경우를 고려한다. 즉, 유의어 관계에 있는 두 개체 C 와 D 는 동일한 상품 클래스를 의미하지만 다른 이름으로 불리는 경우 $C \equiv D$ 즉, DLC에서는 ' $C \subseteq D \wedge C \supseteq D$ '로 나타낸다. 예를 들어, *laptop*이 *notebook*과 동일한 상품이라면 위의 방법대로 $laptop \equiv notebook$ 으로 정의한다. 유의어 관계의 상품 클래스는 같은 속성 관계의 집합을 갖는다. 소유(*possession*)와 같은 의미적 관계의 추가적인 타입은 <그림 3>의 점선으로 표시된 박스로 표현하였다[22]. 이 타입은 비록 일반적인 영역에서 자주 나타나지만 실제로 거의 사용하지 않으므로 본 연구에서도 제외하였다.

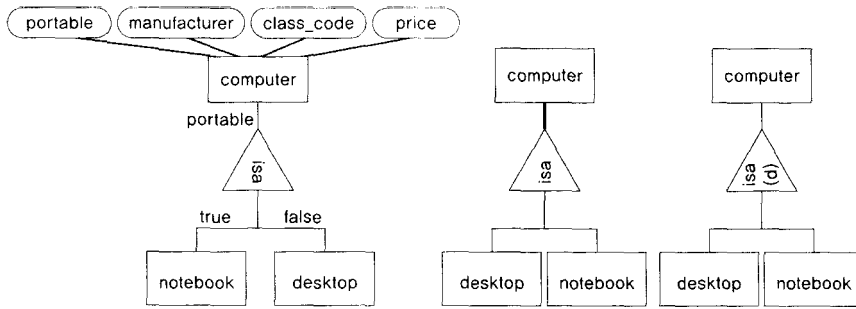
3.2 클래스 인클루전

C 가 D 의 서브 타입임을 표현하는 클래스 인클루전 관계는 ISA(is-a)가 적힌 삼각형으로 나타내고, DLC에서 ' $C \subseteq D$ '로 표현한

다. C 가 D 의 한 종류임을 의미하며 D 의 속성은 C 에게 상속된다. 만약 C 가 D 의 상세화 된 클래스라면, C 는 자신만 갖는 속성을 가질 수 있다. 서브 클래스가 슈퍼 클래스의 몇몇 속성의 값에 대한 조건에 의해서 분류되는 경우를 술어-정의(predicate-defined) 서브 클래스라 하고, 모든 서브 클래스가 슈퍼 클래스와 같은 속성을 갖는 경우를 속성-정의(attribute-defined) 서브 클래스라 한다. 예를 들어, *notebook*은 컴퓨터의 휴대용(*portable*) 속성의 값으로 상세화 된다(그림 4-a). 속성을 식별자(*portable*, *discriminator*라 부름)로 하여 속성 값(*true/false*)을 술어 조건으로 인클루전 관계가 상세화 되고, DLC에서는 ' $notebook \subseteq computer \cap \forall portable = true$ '로 표현한다. 술어-정의와 속성-정의 인클루전 관계는 전자 카탈로그 영역에서 매우 빈번히 나타난다[12].

슈퍼 클래스와 관련된 제약 또는 역할은 서브 클래스로 상속된다. DLC에서 상속은 이미 드의 의미에 내포되어 있다. 따라서 서브 클래스로 상속된 속성, 제약 또는 역할을 다시 구분할 필요는 없다.

클래스 인클루전은 커버(*Covering*)와 디스조인트(*Disjoint*)와 같은 추가적인 제약도 가질 수 있다. 커버 제약은 슈퍼 클래스의 모든 개체가 인클루전과정에서 최소한 하나의 서브 클래스에 포함되어야 하는 것을 의미하며, 서브 클래스 D_1, \dots, D_n 을 C 가 커버하는 커버 제약은 ' $C \subseteq D_1 \cup D_2 \dots \cup D_n$ '로 표현한다. 디스조인트 제약은 서브 클래스의 어떤 개체도 다른 서브 클래스의 개체가 될 수 없음을 의미하며, 서브 클래스 D_1, \dots, D_n 의 디스



(그림 4) (a) 컴퓨터 클래스 인클루전

(b) 커버/디스조인트 제약

조인트 제약은 '어떤 서브 타입 D_i, D_j 에 대해서 $D_i \subseteq \neg D_j$ '로 표현한다. EER기수법에서, 커버/디스조인트 제약은 두꺼운 선으로 표현되거나 삼각형에 (d)를 써준다(그림 4-b).

3.3 부분-전체 관계

클래스 C 와 D 의 부분-전체 인클루전 (meronymic inclusion)은 ' C 가 D 의 부분(D 가 C 의 전체)'을 의미한다. 예를 들어, HDD와 CPU는 컴퓨터의 부품이고 쇠고기 스투(beef-stew)는 쇠고기, 마늘, 양파를 재료로 사용한다. 각기 다른 의미에 따라 관련되어지는 각각의 부분-전체 관계에서의 속성 또한 다르다.

실제로 이용자 질의에서의 부분관계가 직접적인 부분인지, 직/간접적인 부분인지 모르

기 때문에 부분-전체의 의미적 차이는 추론을 복잡하게 한다. 따라서 부분-전체 관계의 정확한 집합을 명백히 밝힐 필요가 있다. 각각의 부분-전체 관계가 의미하고 있는 것과 그 특성들은 다음과 같다[16]. (pw 는 부분-전체 관계임)

클래스 인클루전 관계와 달리 \supseteq 은 포함(subsumption) 제약을 내포하고 있다. 즉, 어떤 부분-전체 관계가 다른 부분-전체 관계와 서로 다른 속성을 가질 수 있는 것을 의미하므로 각각의 부분-전체 관계의 속성은 모델에서 명백하게 표현된다. 우리는 부분-전체 관계의 타입을 구분하기 위해 Winston[22, 19]에 의해 처음 제안되었던 기준을 이용하여 전자 카탈로그 영역에 적합한 다음의 다섯 개의 부분-전체 관계 타입을 찾았고, 우리 모델에 적용하였다. 각각은 부품, 재료, 멤버, 일부, 특

Transitive : $x pw y, y pw z \Rightarrow x pw z.$

Inverse : $x pw_1 y \Rightarrow y pw_2 x, \text{ where } ((pw_1^{-1})^{-1})=pw_2$

Asymmetric : $x pw y \rightarrow \neg (y pw x)$

Identical : $x pw l_1, l_2, \dots, l_p \ \& \ y pw m_1, m_2, \dots, m_q \ \& \ l_1=m_1, l_2=m_2, \dots, l_p=m_q \Rightarrow x \equiv y$

Atomic : $x_1 pw x_2, \dots, x_i pw x_{i+1}, \dots, x_{N-1} pw x_N$ 을 만족하는 N 이 항상 존재함.

〈표 1〉 부분-전체 관계의 타입과 분류 기준

Types	Examples in catalog domain	Classification criteri		
		functional	separable	homogeneous
component-of	HDD component-of computer	Yes	Yes	
substance-of	beef (garlic and onion) substance-of beef-stew			
member-of	umbrella (hat or belt) member-of clothing accessory		Yes	
portion-of	cheese cake slice portion-of cheese cak		Yes	Yes
feature-of	paying feature-of shopping	Yes		

정이다. 앞의 세 개가 가장 자주 나타나므로 다음 장에서 더 자세히 다루도록 하겠다. 〈표 1〉은 Winston의 것에 기반을 둔 우리의 기준이며, 각각의 기준은 다음을 의미한다:

- Functional : 부분이 전체에 대해서 함수적 역할을 수행하는지 여부
- Separable : 부분이 물리적으로 전체와 분리될 수 있는지 여부
- Homogeneous : 부분과 전체가 서로 유사한 정도

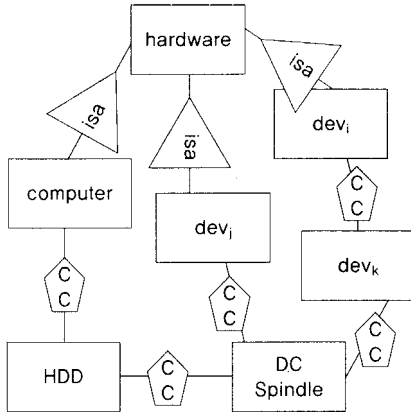
3.3.1 CC : component of / composed-of

CC(*component-of/composed-of*)는 '부분 C가 전체 D의 부품이다 (혹은 D는 C로 구성되었다)'를 의미한다. 이것은 데이터 모델링에서 집합 추상 개념으로 잘 알려져 있다. 〈그림 5〉는 컴퓨터와 관련된 간단한 하드웨어 모델이다. DC 스핀들 모터는 HDD의 부품이고, HDD는 컴퓨터의 부품이다. DC스핀들 모터는 또한 다른 하드웨어 장치의 부품이 된다.

component-of 와 *composed-of* 는 서로 역

관계이다. 즉, $component-of^{-1} \subseteq composed-of$, $component-of \subseteq composed-of^{-1}$ 이다. 일반적으로 CC관계는 집합 및 기능적 의미를 포함하므로 항상 이행적인 것은 아니다. 전체에 대한 어떤 부분이 전체를 구성하는 다른 객체에 대해서 또다시 함수적 역할을 하는 것은 고려할 필요가 없다[16]. 그러나 실제로 사람들은 부품 관계를 사용하는데 있어서 그 의미를 명확히 하지 않는다. 예를 들어, DC스핀들 모터를 부품으로 갖는 모든 하드웨어를 찾으라는 질의는 그것이 DC스핀들 모터를 직접적인 부품으로 갖는 하드웨어인지, 또는 직/간접적인 부품으로 갖는 것인지를 묻는가에 대해 명확하지 않다. 간접적인 부분-전체 관계를 표현하기 위해서는 이행적인 속성이 필요하다. DLC는 모델에서 직접적인 관계만을 표현할 때는 *component-of*를 적용하고 각각의 경우를 따라서 간접적인 관계까지 표현할 때는 *component-of*를 적용한다. (*composed-of*도 유사하게 적용함.)

카탈로그 시스템은 사용자가 모호한 질의를 했을 때 의미를 명확히 하는데 도움을 줄



〈그림 5〉 (a) CC관계의 예

$HDD \subseteq \exists component-of.computer,$
 $HDD \subseteq \exists component-of_i.computer$
 $DCSpindle \subseteq \exists component-of.HDD,$
 $DCSpindle \subseteq \exists component-of_i.HDD$
 $component-of \subseteq composed-of^i,$
 $component-of_i \subseteq composed-of^i,$
 $composed-of \subseteq component-of^i,$
 $composed-of_i \subseteq component-of^i,$
 $component-of_i \in S^+, composed-of_i \in S^+$

(b) DLC 표현

수 있어야 한다. 만약 이전의 질의가 직/간접적인 부품 관계를 모두 얻기 위한 것이었다면, 질의는 다음의 형식을 갖는다. 'hardware $\cap \exists composed-of.DCSpindle?$ ' 그 결과 컴퓨터는 DC스핀들 모터를 부품으로 갖는 상품리스트에 나타날 것이다. 또는 질의어가 직접적인 부품 관계만을 얻기 위한 것이었다면 질의에 *composed-of* 대신에 *composed-of*를 사용하면 된다.

〈표 2〉는 우리가 UNSPSC에서 분석한 CC 관계의 한 예이다. 43222700 전신장비 소분류의 세분류 중에는 텔레그래프 키, 전자석, 음향기 3개의 코모디티(commodity)가 있는데, 이들간의 의미적 관계는 예를 들어, '텔레그

래프 키는 전신장비의 구성 부품이다.'로 해석되는 것이 바람직한 것이다. 현재UNSPSC에서는 의미적으로 CC관계를 나타내는 다음의 관계들을 하나의 클래스 인클루전으로만 표현한다.

3.3.2 SM : substance-of / made-of

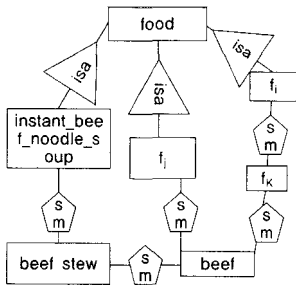
SM(substance-of/made-of)은 '부분 C가 전체 D의 재료이다.(D가 C로 만들어 졌다.)'를 표현한다. *substance-of*와 *made-of*는 서로 역관계이다. 즉, $substance-of^i \subseteq made-of,$ $made-of^i \subseteq substance-of$ 이다. CC와 비교해 볼 때 SM은 이행적이다. 즉, *substance-of,* *made-of*

〈표 2〉 UNSPSC에서의 CC관계의 예

UNSPSC code	Title
43222700	Telegraph equipment
43222701	Telegraph keys
43222702	Telegraph electromagnets
432227 03	Telegraph sounders

〈표 3〉 UNSPSC에서의SM관계의 예

UNSPSC code	Title
50180000	Bread and bakery product
50181700	Baking mixes and supplies
50202300	Non alcoholic beverage
50202311	Powdered drink mix



〈그림 6〉 (a) SM 관계의 예

$beef_stew \subseteq \exists made-of.beef$
 (or $beef \subseteq \exists substance-of.beef_stew$),
 $ibns(instant_beef_noodle_soup) \subseteq \exists made-of.beef_stew$
 $made-of \subseteq substance-of^1, substance-of \subseteq made-of^1$
 $made-of \in S^+, substance-of \in S^+$

(b) DLC 표현

S+이다.

〈그림 6-a〉는 간단한 음식 카탈로그의 모델이다. 쇠고기는 쇠고기 스투의 재료이고 쇠고기 스투는 인스턴트 쇠고기 누들 수프의 재료이다. 〈그림 6-b〉 나타낸 DLC모델로 쇠고기를 재료로 포함하는 모든 음식을 찾는 질의에 응답할 수 있다. 'food ∩ ∃made-of.beef?'

〈표 3〉은 우리가 UNSPSC를 분석하여 찾아낸 SM관계의 한 예이다. SM관계는 주로 음식과 관련된 부분에서 그 예를 찾아 볼 수 있었다.

3.3.3 MC : member-of / collection-of

MC(member-of/collection-of) 관계는 멤버 집합이 갖는 고유의 권리에 대해서 객체를 고려하는 것을 의미한다. 멤버가 집합에서 특별한 함수를 수행하는 것을 고려하지 않는 부품 관계와 다르다. MC는 전체적인 집합에서 개개의 멤버십을 포함하기 때문에 클래스 인클루전 관계와 쉽게 혼동될 수 있다. 그러나 MC관계는 내부적 특성에 기반을 두어 유사성을 결정하는 클래스 인클루전과 대조적으로 각 멤버에 외부적 특성에 근거를 둔다. 이

〈표 4〉 UNSPSC에서의 MC 인클루전의 예

UNSPSC code	Title
53102500	Clothing accessories
53102501	Belts or suspenders
53102502	Ties or scarves or mufflers
53102503	Hats
53102504	Gloves or mitten
53102505	Umbrella
53102506	Sweat bands
53102507	Clothing hangers
53102508	Armbands

러한 기준은 공간적인 근접 혹은 사회적 관계를 포함할 수 있다.

예를 들어, UNSPSC 531025 클래스 코드(표 4)는 MC관계로 의류 악세서리 클래스와 그것에 포함된 상품간의 관계를 보다 자연스럽게 관련 지을 수 있다. 다른 전자 카탈로그 영역에서도 유사한 경우가 나타난다[18].

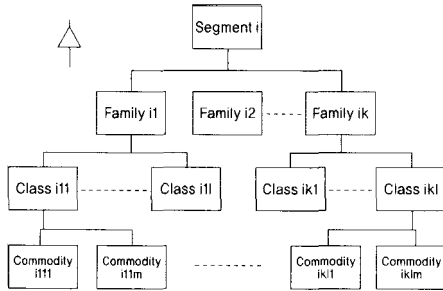
4. 논의 사항

본문의 전자카탈로그 모델인 DLC는 SHIQ(d)의 부분집합이다. SHIQ(d)는 이행적 역할(transitive, S) 역할 계층(role hierarchy, H), 역(inverse, I), 수 제한(number restriction, Q)과 구체적인 도메인(concrete domain, d)으

로 ALC를 확장한 언어이다. SHIQ가 상품간의 부분-전체 관계를 표현하기에 적합하기 때문에 DLC는 SHIQ를 기반으로 하였다. 현재의 DLC의 구문법칙 및 언어능력은 SHIQ(d)와 같다. 그럼에도 불구하고 본 연구에서 SHIQ(d)대신 DLC란 용어를 사용한 것은 추후에 카탈로그 표현에 필요한 언어적 구성이 SHIQ(d)를 넘어서 일반적인 SHIQ와 달라질 수 있고, 이로 인해 불필요한 혼선을 야기할 수 있다는 점을 방지하기위한 것이며, DLC에 기반을 둔 카탈로그 XML언어의 개발 시 SHIQ기반의 OWL과의 차이점을 부각하기 위함이다. 또한 SHIQ는 일반적인 부분-전체 관계의 몇 가지 특성인 원자성(atomicity), 동일성(identity)과 같은 표현뿐만 아니라 부분-전체 관계의 상속에 대한 모

〈표 5〉 DAML+OIL, OWL, DL의 관계

DAML+OIL	OWL	DL syntax
intersectionOf	IntersectionOf	$C1 \cap \dots \cap Cn$
unionOf	UnionOf	$C1 \cup \dots \cup Cn$
complementOf	ComplementOf	$\neg C$
toClass	AllValuesFrom	$\forall R.C$
hasClass	SomeValuesFrom	$\exists R.C$
hasValue	HasValue	$\exists R.\{X\}$
minCardinalityQ	MinCardinality	$(\geq n \ R.C)$
maxCardinalityQ	MaxCardinality	$(\leq n \ R.C)$
inverseOf	InverseOf	R^{-}
transitiveProperty	TransitiveProperty	$R \in S^{+}$
subClassOf	SubClassOf	$C1 \subseteq C2$
sameClassOf	EquivalentClass	$C1 \equiv C2$
samePropertyAs	equivalentProperty	$P1 \equiv P2$
disjointWith	DisjointWith	$C1 \subseteq \neg C2$

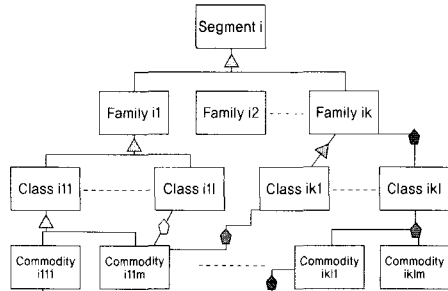


〈그림 7-a〉 인클루전 관계만 표현하는 분류 모델

모델링도 할 수 없다. 역할값지도(role value map)를 모델에 적용하면 부분-전체 관계의 상속을 고려할 수는 있지만, 역할값지도로 모델을 확장하면 만족도 문제를 해결 하지 못할 수도 있다[16].

DAML+OIL이나 OWL은 DL과 많은 공통적 특징을 갖고 있다. 따라서 향후의 연구 방향이 OWL을 기반으로 해야 한다는데 동의한다. 예를 들어, 측정 단위 환산 문제나 같은 의미를 갖지만 다른 용어를 사용하는 경우를 고려해야 한다. 〈표 5〉는 DAML+OIL, OWL, DL이 제공하는 구성자와 공리를 정리한 것이다.

우리의 모델을 UNSPSC, eCI@ss와 같은 표준 상품 코드 시스템을 구축하는 것을 목표로 하는 상품 분류 모델과 비교해 보면, 〈그림 7-a〉와 같은 코드 기반은 모델에서 표현되는 상품 클래스 간의 의미적 관계는 오직 인클루전 한 종류뿐이라서 다양한 의미적 관계나 제약들을 명백히 표시할 수 없다. 그러나 실제로 인클루전의 여러 다른 의미가 존재하고, 분류코드를 분류하는 개발자와 이용자 모두는 그들 나름대로 의미를 해석하여 사용하고



b) 다양한 의미적 관계를 포함하여 확장 시킨 분류 모델

있다. 전자카탈로그시스템이 의미적으로 풍부한 유연성을 제공하기 위해서 모델을 〈그림 7-b〉와 같이 의미적 관계와 제약을 포함하도록 확장해야 한다고 생각한다.

5. 결 론

본 논문에서는 전자카탈로그 영역에서의 온톨로지 모델링을 소개하였다. UNSPSC체계를 분석하여 실제 전자카탈로그영역에 존재하는 다양한 의미적 관계를 체계적으로 정리하였고, 각각의 관계가 어떻게 온톨로지 모델화 될 수 있는지 제시하였다. 모델은 EER을 기반으로 하여 추론이 가능하도록 DL로 표현하였다. 본 연구는 전자카탈로그영역에 온톨로지를 적용하기 위해 전자카탈로그영역에 존재하는 객체들간의 의미적 관계를 예를 통해 분석 체계화 하고, 이를 온톨로지 모델로 표현해 본 데 있다.

본 논문의 DLC언어는 앞으로 또 다른 모델링 작업과 통합하기 위해 확장이 필요할 수도 있으며, 그에 따른 모델링 가능범위와 추

론의 복잡도에 대한 연구 역시 필요할 것이다. 관련 연구들을 통해 모델을 확장하고 DLC언어의 완성도를 높여가는 것이 향후 연구 방향이다.

참 고 문 헌

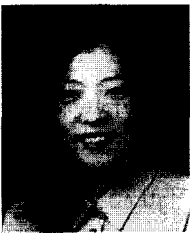
- [1] F. Baader, I. Horrocks, and U. Sattler, "Description Logics as Ontology Languages for the Semantic Web", *Knstliche Intelligenz*, Vol. 16, 2002.
- [2] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web", *Scientific American*, Vol. 284, 2001.
- [3] D. Calvanese, M. Lenzerini, and D. Nardi, "Unifying Class-Based Representation Formalisms", *Journal of Artificial Intelligence Research*, Vol. 11, AI Access Foundation and Morgan Kaufmann Publishers, 1999.
- [4] Z. Cui, D. Jones, and P. O'Brien, "Semantic B2B Integration: Issues in Ontology-based Approaches," *SIGMOD Record*, Vol. 31(1), ACM, 2002
- [5] R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, 4th Edition, Person Addison Wesley, 2003.
- [6] 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.
- [7] eClass - New Standardized Material and Service Classification, <http://www.eclass-online.com/>, Cologne Institute for Business Research.
- [8] V. Haarsler and R. Mller, "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.
- [9] 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.
- [10] I. Horrocks, "Benchmark Analysis with FaCT", *Proc. Of TABLEUX 2000*, Springer-Verlag, 2000.
- [11] I. Horrocks and U. Sattler, "Ontology Reasoning in the

- SHIQ(d) Description Logic”, Proc. Of 17th International Joint Conference on Artificial Intelligence(IJCAI01), American Association for Artificial Intelligence, 2001.
- [12] D. Kim, A Semantic Model for Electronic Catalogs, Ph.D. Dissertation, Seoul National University, 2004.
- [13] 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.
- [14] H. Lee and J. Shim, “When Ontology meets e-Catalog”, e-Business World Conference 2004, Society for e-Business Studies in Korea, 2004.
- [15] J. Leukel, V. Schmitz, and F. Dorloff, “A Modeling Approach for Product Classification Systems,” Proc. of the Database and Expert Systems Applications, 13th International Conference (DEXA 2002), Springer-Verlag, 2002.
- [16] U. Sattler, “Description Logics for the Representation of Aggregated Objects”, Proc. of the 14th European Conference on Artificial Intelligence, IOS Press, 2000.
- [17] M.K. Smith, C. Welty, and D.L. McGuinness, “OWL Web Ontology Language Guide W3C Recommendation”, <http://www.w3c.org/TR/owl-guide/>, 2004.
- [18] P. Spyns, R. Meersman, and M. Jarrar, “Data modeling versus Ontology engineering,” SIGMOD Record, Vol. 31(4), ACM, 2002.
- [19] V. C. Storey, “Understanding Semantic Relationships”, VLDB Journal, Vol. 2, VLDB Endowment, 1993.
- [20] UCCnet, <http://knowledgebase.base.uccnet.org/>, UCCnet.
- [21] UNSPSC, United Nations Standard Products and Services Code, <http://www.unspsc.org/>, UNDP.
- [22] M.E. Winston, R. Chaffin, and D. Herrmann, “A taxonomy of part-whole relations”, Cognitive Science, Vol. 11, Cognitive Science Society, 1987.
- [23] 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.

저 자 소 개



심준호 (E-mail : jshim@sookmyung.ac.kr)
1990. 2. 서울대학교 계산통계학과(학사)
1994. 2. 서울대학교 계산통계학과 전산과학전공(이학석사)
1998.12. Northwestern Univ., Illinois, USA., Electrical & Computer Engineering, 공학박사
1999. 2~1999.11. Computer Associates Int'l, New York, USA, R&D Staff
1999.11~2001. 2. Drexel Univ., Pennsylvania, USA, Assistant Prof.
2001. 3~현재 숙명여자대학교 부교수
관심분야 데이터베이스 데이터웨어하우스 전자상거래, 웹



이현자 (E-mail : hyunjalee@sookmyung.ac.kr)
1996. 2. 숙명여자대학교 전산학과 전산학 전공(학사)
2004. 6. 숙명여자대학교 컴퓨터과학 데이터베이스 전공(석사)
2004. 7~현재 숙명여자대학교 컴퓨터과학 데이터베이스 전공 박사과정



이민정 (E-mail : judy@sookmyung.ac.kr)
2003. 2. 숙명여자대학교 정보과학부 컴퓨터과학, 문헌정보학 전공(학사)
2003. 3~현재 숙명여자대학교 컴퓨터과학 데이터베이스 전공 석사과정
관심분야 데이터베이스 전자상거래, 이미지 데이터베이스