

# ERP 시스템의 데이터 정확성 유지를 위한 물품 온톨로지의 설계 및 구현

(Design & Implementation of an Item Ontology for  
maintaining the accuracy of data managed by  
ERP system)

정 현 숙\*

(Hyun-Sook Chung)

**요 약** ERP 시스템은 기업의 기간 시스템으로서 기업이 e-business에 참여하기 위해서는 시스템의 성공적인 운영이 필수적인 요소이다. 그러나 대부분의 ERP 도입 기업들은 ERP 운영에 많은 문제점과 어려움을 가지고 있다. 이는 ERP 시스템에 의해 관리되는 데이터가 부정확하기 때문이다. 본 논문에서는 기업의 핵심 데이터인 물품마스터의 정확성 유지와 물품과 관련된 지식 자원을 통합 관리하기 위해 물품 온톨로지의 모델 및 구축 프로세스를 제시하고 이 온톨로지를 효율적으로 관리할 수 있는 온톨로지 관리 시스템을 설계 및 구현하였다. 제안된 모델과 구축 프로세스에 따라 물품 온톨로지를 구축하고 이를 관리하는 온톨로지 관리 시스템을 기업의 기간 시스템과 연계함으로써 중소기업의 e-business로의 전이가 용이하리라 본다.

**핵심주제어** : ERP 시스템, 물품 온톨로지, 토픽맵, 온톨로지 관리 시스템

**Abstract** ERP system is a fundamental system that is commonly adopted by most companies that intend to take part in e-business. However, few companies are running ERP systems successfully because of inaccuracy in data managed by ERP system. In this paper, we propose a model and the building process of item ontology which organize concepts describing all items managed by enterprises. In addition, we design and implement a system to create and manage item ontologies. We interface our system with a commercial ERP system, then install and run these systems on a manufacturing company. Our item ontologies and management system might help enterprises when they try to take part in e-business.

**Key Words** : ERP system, Item Ontology, Topic Maps, Ontology Management System

## 1. 서 론

ERP(Enterprise Resource Planning) 즉, 전자사  
원관리 시스템은 기업의 프로세스를 개선하고 프  
로세스 흐름에 따라 단위 업무를 통합함으로써 업

무의 효율을 높이고 낭비되는 비용을 절감하며 기  
업의 이익을 높이기 위한 기반 시스템이다. 또한  
e-Business에 참여하기 위해서는 무엇보다 기업  
내의 데이터를 정확히 유지하고 프로세스를 자동  
화하기 위한 ERP의 구축이 필수적이다. 이는 기업  
내부의 표준화와 개선 없이 e-business에 참여하  
는 기업 사이의 자동화된 전자 거래가 성립될 수

\* 가톨릭대학교 컴퓨터정보공학부 교육전담 전임강사

없기 때문이다[1,2].

그러나 많은 기업들이 ERP를 도입했음에도 불구하고 성공적으로 운영하는 사례를 찾아보기는 쉽지 않다. 특히 중소기업의 경우 시스템 구축에 오랜 기간 투자하였으나 실제 업무에 효과적으로 적용하지 못하고 있는 것이 대부분이다. 여기에는 여러 가지 원인이 있으나 그 중 가장 중요한 문제는 ERP 시스템 내에 관리되는 데이터의 정확도 및 신뢰에 관한 문제이다[1].

이는 대부분 ERP를 도입하려는 기업들 특히 중소기업은 데이터의 표준화가 부족하여 부서 사이에도 동일 품목에 대해 서로 다른 용어 및 코드를 사용하고 있으며 상호 연결된 프로세스에서 데이터의 동기화가 적절히 이루어지지 않아 단절된 데이터들만 보관되고 있기 때문이다.

데이터 부정확의 문제는 보다 엄밀히 말하면 물품 마스터(item master) 정보의 부정확으로 인해 야기되는 문제라고 볼 수 있다. 여기서 물품(item)은 완제품(final product), 반제품(assembly), 원자재(raw material), 부품(component), 부자재(sub-material) 등을 총칭하는 것으로 여기에는 기업의 유형물과 시간, 서비스, 비용 등 무형물을 포함하는 개념을 가진다[1].

이 문제를 좀 더 자세히 살펴보면 다음과 같은 문제점을 찾을 수 있다. 첫째, 물품 마스터가 부정확하다. 즉, 동일 품목에 대해 중복된 다른 코드가 부여되었거나 품목 체계가 잘못 분류된 경우이다. 또한 물품의 기준정보, 자재정보, 공정정보, 회계정보 등의 속성정보들이 잘못 기입된 경우이다.

둘째, 물품 마스터의 확장이 유연하지 못하다. ERP 시스템은 데이터베이스를 저장소로 하기 때문에 모든 데이터가 테이블의 필드 단위로 저장된다. 이때 필드의 추가는 ERP 프로그램의 코드 변경을 요구하는 경우가 많기 때문에 사용자 입장에서 부가적인 정보를 추가로 기술하기가 어렵다.

셋째, ERP 시스템 내에 모든 정보를 저장 및 관리하기가 어렵다. 하나의 물품에 대한 자세한 명세서, 설계도, 시방서 등의 문서나 이미지를 ERP 시스템 내에 저장할 수 없으므로 사용자 입장에서는 기업 내의 중요 정보에 쉽게 접근하지 못하는 문제를 가진다.

넷째, 물품 코드 생성을 위한 체계적이고 표준화된 품목 분류 체계가 부족하다. 이는 기업 내부에 품목

분류가 제대로 관리되지 않을 수도 있으며 ERP 시스템 자체가 품목 분류에 의거한 자동 코드 생성 기능을 제공하지 못하기 때문이다. 이러한 문제는 동일 품목에 대해 중복된 코드 생성을 야기 시킨다.

이러한 문제들은 사용자의 운영 능력 부족에서 비롯된 것일 수도 있으나 근본적으로 ERP 시스템의 정적인 데이터 관리에서 찾을 수 있다. 이를 해결하기 위한 접근은 정적 데이터에 유연성을 부여할 수 있는 방법을 찾는 것이다. 본 논문에서는 ERP 시스템과 독립적으로 물품에 대한 온톨로지(ontology)를 구축하고 관리함으로써 물품 마스터에 확장 및 유연성을 제공하고 데이터의 표준화를 유도한다. 또한 ERP 데이터베이스외에 문서, 이미지, 동영상, 웹페이지 등의 다른 구조의 자원들도 물품과 연결함으로써 하나의 물품 지식베이스를 구축한다.

온톨로지는 현재 시맨틱웹(semantic web), 지식 기반 서비스, 전자상거래, 데이터 통합 등 다양한 분야에서 지식을 표현하고 지식들 사이의 의미적 계층 및 연관 관계를 설정함으로써 지식 색인과 검색을 위한 핵심 요소로 사용되고 있다. 여기서 물품 또한 기업이 내재하고 있는 하나의 중요한 지식이므로 물품의 분류 체계를 규정하고 의미 관계를 설정하는 것은 문제를 해결하는 올바른 접근법이라 볼 수 있다.

본 논문에서는 ERP의 지식 처리를 위해서 먼저 ERP 상위에 물품의 지식층(knowledge layer)을 설계하고 이 지식층과 ERP 데이터베이스내의 데이터, 문서, 도면 이미지 등을 포함하는 자원층(resource layer)을 연결하는 의미망(semantic network) 구조의 물품 온톨로지(item ontology)를 구축한다. 여기서 지식맵을 구축하기 위한 데이터 모델로 ISO 국제기구에서 표준(ISO/IEC 13250)으로 정한 토픽맵(topic maps)[3]을 사용한다.

토픽맵은 RDF(Resource Description Framework)[4]와 함께 온톨로지의 데이터 모델로 사용되고 있으며 콘텐츠 관리 시스템, 지식 관리 시스템, 데이터 통합 시스템 등의 지식 색인 모델로서 사용되고 있다. 웹상의 자원들을 의미적으로 연결하는 RDF와 달리 토픽맵은 웹 뿐만 아니라 추상적인 개념들의 연결도 가능하기 때문에 ERP의 지식 표현 및 구조화에 더 적합하다고 볼 수 있다[5].

본 논문에서는 토픽맵을 사용하여 지식 처리를 할 수 있도록 토픽맵 기반 지식 처리 시스템을 설

계 및 구현하였다. 이 시스템의 주요 기능은 토픽 맵 해석 및 생성, 토픽 맵 질의 처리, 그래픽 기반 지식 항해, 키워드 기반 지식 검색, 토픽 맵 저장 등이 있으며 API를 제공하여 ERP, CRM, EDMS 등 기존 시스템의 확장 시스템으로 통합하여 사용될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 기업의 온톨로지 구축과 관련되어 이전 연구들을 살펴보고 3장에서는 물품 온톨로지의 구조를 살펴본다. 4장에서는 물품 온톨로지 구축 프로세스에 대해 논의하고 5장에서는 물품 온톨로지 구축 및 관리를 지원하는 토픽 맵 응용 시스템의 구조 및 주요 모듈들을 살펴본다. 6장에서는 상용 ERP 시스템과의 연계 결과를 보이고 7장에서 결론 및 향후 연구 과제에 대해 소개한다.

## 2. 관련연구

ERP 시스템의 데이터 관리 문제를 제기하고 이에 대한 해결책으로 데이터 리엔지니어링을 소개하는 연구로서 [1]을 들 수 있다. 이 연구에서는 ERP 시스템의 데이터 문제가 발생하는 원인을 물품 마스터 데이터의 부정확에서 찾고 있으며 이를 해결하기 위해 비즈니스 데이터를 상호 인식 가능하도록 재설계하고 구조화하고 이를 시스템이 지원하도록 하고 있다. 이 연구에서는 ERP 시스템 내에 전자 카탈로그 DB를 물품 마스터 DB 위에 둬으로써 품목 분류에 따라 물품 코드를 생성하도록 하여 코드 중복 오류의 문제를 해결하고자 한다. 그러나 이 연구는 데이터 표준화에 대한 이론적인 접근으로서 ERP 시스템 데이터의 비유연성 및 비확장성에 의한 문제를 시스템적으로 해결하지 못하고 있다.

기업의 비즈니스와 데이터를 대상으로 온톨로지를 구축한 연구로서 TOVE[6], ENTERPRISE[7], METHONTOLOGY[8]를 들 수 있다.

TOVE(Toronto Virtual Enterprise)는 기업의 비즈니스 프로세스 모델링을 돕기 위한 온톨로지로서 기업의 조직 구조 및 각 조직의 활동에 관한 개념을 정의하고 개념들 사이의 관계를 정의하고 있다. ENTERPRISE 또한 기업의 비즈니스와 관련된 개념을 정의한 것으로 조직, 활동, 전략, 마케팅의 주제어로 분류된다. 각 주제별로 핵심 용어를 정

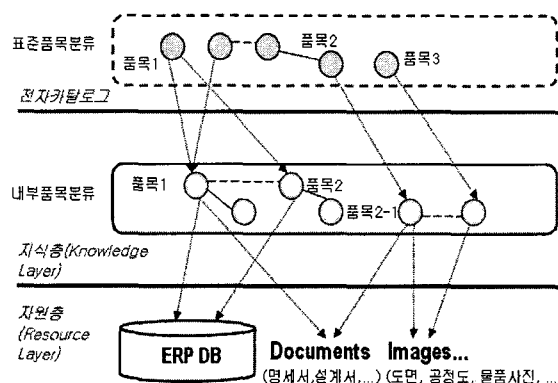
의하고 있으며 용어들 사이에 관계를 설정하고 있다. 예를 들어, 전략 주제에는 목적(purpose), 확정된 목적(hold purpose), 의도된 목적(intended purpose), 전략적 목적(strategic purpose) 등의 용어가 정의되고 이들 사이에 계층구조로 연결되어 있다.

METHONTOLOGY는 화학 분야 기업에서의 비즈니스를 모델링 한 것으로 위 두 온톨로지에 비해 보다 구체적이고 세분화된 개념 표현을 가진다. 예를 들어, 특정 제품의 배합 비율 및 적정 온도와 관련된 규칙, 계산식 등을 별도의 개념으로 분류해 놓았다.

본 연구의 물품 온톨로지는 이들 연구와 유사한 온톨로지 구축 절차를 가지고 있으나 온톨로지 구조상 이들 연구와 달리 보다 상세한 수준의 온톨로지라 할 수 있다. 즉 이전 연구에서 구축된 기업 온톨로지들은 기업의 활동과 관련된 상위 수준의 온톨로지인 반면 본 연구의 온톨로지는 이러한 활동에서 발생하는 데이터 자체를 관리하는 온톨로지이다. 또한 기업의 기간 시스템의 데이터 관리에 유연성과 확장성을 제공하기 위한 온톨로지이다.

## 3. 물품 온톨로지 구조

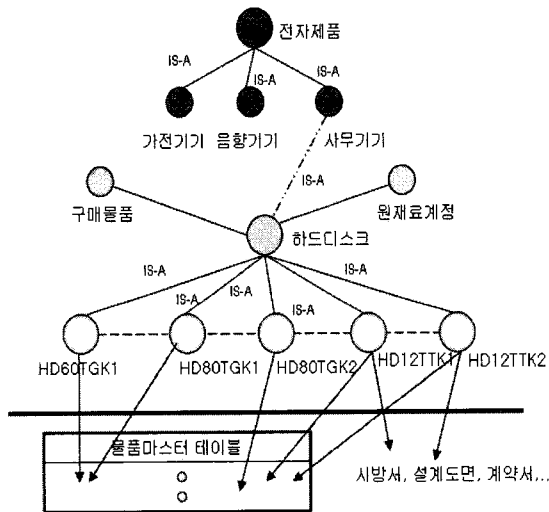
물품 온톨로지는 크게 지식층(knowledge layer)과 자원층(resource layer)으로 나누어진다. 지식층에는 물품 정의 및 분류와 관련된 지식을 구성하는 계층이고 자원층은 지식층에 정의된 지식과 연결된 데이터베이스, 문서, 웹페이지, 이미지 등의 실제 자원들로 구성된 계층이다. <그림 1>은 물품 온톨로지의 개념적인 구조를 보이고 있다.



<그림 1> 물품 온톨로지의 개략적 구조

지식층의 상위에는 HS, UNSPSC 등의 표준 품목분류 체계를 둘 수 있으며 이 경우 전자카탈로그 시스템에서 기업 내부에 자체적으로 정의된 품목 분류와 표준 품목 분류 사이에 매핑을 설정할 수 있다. 이러한 매핑 정보에 따라 기업은 B2B 전자상거래에 참여할 수 있으며 파트너 관계의 타 기업과의 자동화된 데이터 처리도 가능하다.

물품 온톨로지의 지식층에는 물품에 관한 지식 외에 기업의 조직을 정의하는 조직층, 비즈니스와 활동을 정의하는 비즈니스층, 그리고 비즈니스의 세부 활동인 TASK층을 독립적으로 구축할 수 있으며 이들 계층들과 물품들을 연결함으로써 온톨로지를 확장할 수 있다.



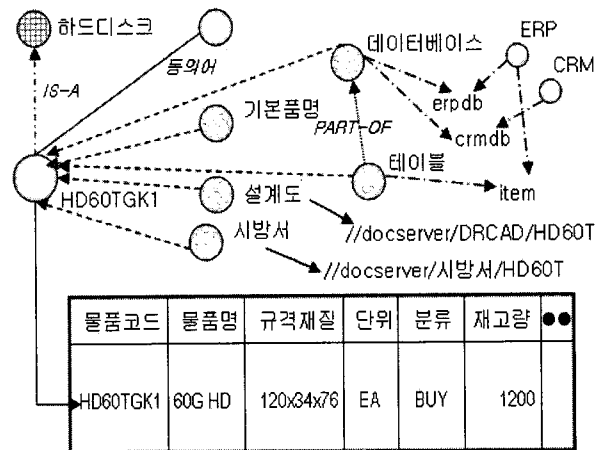
<그림 2> 지식층과 자원층의 연결 구조

<그림 2>는 지식층과 자원층의 연결을 예를 들어 보이고 있다. 지식층의 모든 개체(entity)는 자원층의 특정 자원으로의 연결을 가질 수 있다. 여기서 개체는 온톨로지를 구성하는 요소들로서 품목명, 물품명, 속성명, 관계명 등을 가리킨다. <그림 2>에서는 하드디스크 품목에 속하면서 물품코드를 가지는 개체가 ERP 시스템의 데이터베이스, 시방서, 설계도면, 공급 계약서 등의 자원에 대한 하이퍼링크를 가지고 있음을 보인다.

지식층의 하나의 개체는 다른 개체와의 구분을 위한 식별자로서 유일한 ID를 가지며 여러 명칭들 중에서 가장 기본적인 명칭을 기본명(base name)으로 가진다. 이때 하나의 개체가 가지는 명칭은 약어, 특수어, 전문어, 별명 등이 될 수 있다. 이와

함께 개체의 특성을 설명하는 속성들이 정의된다. 속성에는 크게 내부 속성과 외부 속성으로 나누어진다. 내부 속성은 속성 값이 온톨로지 내부에 정의되는 것이고 외부 속성은 속성 값이 온톨로지 외부 즉, 자원층에 존재하는 것으로 URI(Unique Resource Identifier)로 기술되는 참조 주소가 된다.

개체는 개념적으로 외연(extension)에 해당하는 상위 개체를 부모로 가질 수 있으며 내포(intension)에 해당하는 하위 개체를 자식으로 가질 수 있다. 이때 상위 개체와 하위 개체 사이에는 IS-A 관계가 성립된다. 또한 다른 개체와의 사이에 의미적 연관성이 존재하는 경우 관계를 정의하는 개체를 생성하고 이 관계 개체의 멤버로써 다른 개체와 의미적으로 연결될 수 있다. 예를 들어, 인텔 CPU와 AMD 전용 마더보드 사이에는 disjoint-of의 관계가 정의될 수 있다.



<그림 3> HD60TGK1 개체의 연결 구조

<그림 3>은 HD60TGK1 개체를 중심으로 연결되는 망의 구조를 개략적으로 보이고 있다. HD60TGK1은 기본품명, 설계도, 시방서, 데이터베이스, 테이블을 속성으로 가지며 각 속성에는 속성 값이 기술되어 있다. 여기서 HD60TGK1 개체는 ERP 시스템이 사용하는 erpdb 데이터베이스의 item 테이블에 저장되어 있음을 가리키므로 본 논문의 토평맵 응용

프로그램에서는 이 속성 값을 이용하여 개체와 테이블의 특정 레코드와 연결한다. 물품 마스터 테이블의 경우 물품코드 컬럼이 주요키이므로 별도의 키 컬럼 속성은 가지지 않는다.

HD60TGK1 개체와 다른 개체와의 관계를 살펴보면 하드디스크와 IS-A의 관계를 가지고 다른 하나의 개체와 동의어(synonym) 관계를 가진다. IS-A와 PART-OF는 개체들 사이의 계층적 구조를 형성하는 관계이고 동의어 관계, 상호 연결 불가 관계, 대체 가능 관계 등은 의미적으로 개체들을 연결하는 수평적 연관관계이다. 온톨로지는 시소러시와 달리 의미관계를 임의로 정의할 수 있으며 제약조건이나 추론을 위한 로직, 규칙을 정의할 수 있다.

#### 4. 물품 온톨로지 구축 프로세스

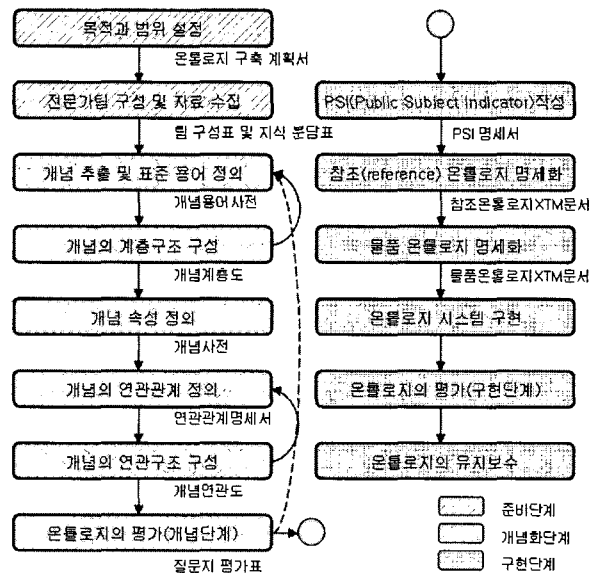
온톨로지는 개념의 명세화, 구조화, 색인화이므로 이를 구축하는 것은 매우 어려운 일이며 많은 시간과 노력이 소요되는 일이다. 이에 따라 온톨로지를 자동으로 구축하기 위해 텍스트 마이닝(text mining)에서는 자연어 처리, 기계 학습 등으로 단어를 추출하고 단어들 사이의 통계 정보에 따라 연관성을 부여하여 관계를 설정한다. 그러나 한 단어가 지니는 의미는 상황에 따라 달라질 수 있으며 이에 따라 단어들 사이에 넓은 의미와 좁은 의미를 기계적으로 해석하고 찾지란 불가능한 일이다.

그러므로 온톨로지 구축을 위해서는 그 도메인의 전문가들이 서로 협의하여 개념을 추출하고 개념들 사이의 연관관계를 해석하며 개념 정의를 표준화하는 과정을 필요로 한다. 이 장에서는 물품 온톨로지 구축을 위한 일련의 과정들에 대해 기술하고 있다.

컴퓨터 프로그램을 개발하기 위해 소프트웨어 공학에서 소프트웨어 개발 방법론이 있듯이 온톨로지를 구축하기 위해서도 온톨로지 구축 방법론이 있다[6,7,8]. 그러나 이 방법론들은 온톨로지 구축을 위한 큰 흐름만을 명시하고 있으므로 도메인에 따라 개념을 명세화하기 위한 세부 과정은 별도로 연구되어야 한다.

본 논문에서는 온톨로지 구축 방법론의 큰 흐름을 따르면서 세부적으로는 물품 정보를 온톨로지화 하기 위한 별도의 세부 과정을 정의하였다.

<그림 4>는 물품 온톨로지 구축 프로세스를 보이고 있다.



<그림 4> 물품 온톨로지 구축 프로세스

구축 프로세스는 준비 단계, 개념화 단계, 구현 단계의 세 가지 단계로 나누어지고 다시 14개의 세분화된 과정으로 이루어져 있다.

##### 4.1 준비 단계

하나의 도메인에는 많은 다양한 종류의 지식들이 존재하므로 필요에 의해 하나 이상의 서로 다른 온톨로지들이 구축될 수 있다. 그러므로 온톨로지를 구축하기 전에 온톨로지에 포함할 개념의 형태, 개념의 표현 깊이, 지식 서비스 내용 등에 적합한 목적과 범위를 정하는 것이 우선이다. 물품 온톨로지를 예를 들면, 범위는 기업 내부의 제품, 상품, 구매품, 부품 등 유형 물품과 비용, 서비스 등 유형 물품을 대상으로 한다. 목적은 ERP, CRM 등 기업의 기간 시스템에 정확한 물품 지식 제공과 확장성, 유연성을 부여하기 위해서이다.

목적과 범위가 결정되면 이에 적합한 지식을 보유한 전문가들을 모집하여 팀을 구성한다. 전문가의 역할은 지식을 담고 있는 자원들을 분석하여 주요 개념들을 찾고 각 개념의 의미를 정확히 표현할 수 있는 표준 용어를 정의하는 것이다. 또한 개념들 사이에 계층적 포함 관계 및 의미적 연관

관계를 설정하는 것이다.

물품 온톨로지의 경우 설계 기술, 제조, 구매, 영업, 재무 등의 물품 정보에 관한 전문가들로 팀을 구성하고 지식 자원은 시스템의 데이터베이스, 설계도, 시방서, 명세서, 계약서 등 각종 문서 자료들을 대상으로 한다.

## 4.2 개념화 단계

개념화 단계는 온톨로지를 설계한 다음 비형식적인 언어나 표현 도구로 온톨로지를 생성하는 단계로서 지식자원 분석, 개념 추출, 개념 용어 정의, 개념 계층 구조 정의, 개념 속성 정의, 개념 연관관계 정의 등의 과정을 반복적으로 수행한다.

개념추출 및 표준용어 정의 과정에서는 지식자원을 분석하여 온톨로지에 포함할 주요 개념들을 파악하고 개념의 의미를 정확히 표현할 수 있는 용어를 선정하여 표준 용어로 정의하는 과정이다. 개념을 추출하는 방법은 접근 방향에 따라 하향식(top-down), 상향식(bottom-up), 상하향식(middle-out) 방법이 있다. 하향식은 가장 넓은 의미를 가지는 최상위 개념을 루트로 두고 그 개념의 하위 개념들을 파악해 나가는 방법이고 상향식은 반대로 가장 아래 수준의 구체적인 개념들을 의미적 유사성에 의해 분류한 다음 상위 개념을 정의하면서 위로 올라가는 방법이다. 상하향식은 계층 구조의 중간 위치의 개념들을 먼저 추출한 다음 그 개념들에서 위로, 아래로 개념들을 찾아가는 방법이다.

이미 ERP 시스템이 구축된 상태에서 물품 온톨로지를 생성하는 경우 상향식 방법으로 접근한다. 즉, 가장 아래 수준의 구체적인 물품들이 코드로 정의되어 있으므로 이 물품들의 품명, 성격, 처리방법 등의 기준에 의해 유사한 물품들을 분류하면서 상위 개념들을 찾아나간다. 이때 ERP에 등록된 물품명은 필드 길이의 제한으로 인해 축약된 형식의 기호를 사용하지만 개념 추출 및 용어 정의 시에는 축약되지 않은 전체 물품명을 분석 대상으로 한다. 그리고 물품명 외에 규격, 특성 등을 분석하여 상위 그룹을 결정한다.

예를 들어, 하드디스크의 경우 규격에 따라 데스크탑용과 노트북용으로 분류되고 전송율에 따라 5400, 7200 rpm 등으로 분류되고 제조사에 따라

시게이트, 삼성 등으로 분류할 수 있다.

개념 추출 후 각 개념에는 정확한 의미를 표현하는 용어를 선정 또는 정의한다. 용어의 형식은 명사형으로써 단일명사, 복합명사, 명사구 등이 될 수 있다. 예를 들어, '보조기억장치', '노트북 하드디스크', '하드디스크 삼성 7200RPM' 등이 개념의 용어가 될 수 있다.

개념이 추출되는 과정에서 유사성에 따라 개념들을 분류하여 상위로 파생되어 올라가기 때문에 자연스럽게 계층 구조를 형성하게 된다. 그러므로 개념 추출과 계층 구조 정의는 최상위 개념을 결정할 때까지 반복적으로 수행되는 과정이다.

개념들이 도출되면 개념 속성 정의 과정에서는 상위의 개념부터 하위로 내려가면서 개념의 속성을 파악하고 정의한다. 상위에서 하위로 내려오는 이 유는 상위의 개념이 하위 개념을 포함하는 포괄적 의미를 가지므로 상위 개념의 속성 또한 하위 개념으로 상속된다.

예를 들어, 하드디스크 개념의 속성으로 기본명, 전송율, 제조사, 규격, 시방서, 계약서, 이미지, 데이터베이스, 테이블 등을 정의하고 이 속성들은 하드디스크 하위의 개념들에 상속된다.

개념의 속성은 개념 자체를 설명하기 위한 속성과 개념과 자원을 연결하기 위한 속성으로 나눈다. 그리고 속성 값은 온톨로지 내부에 저장되는 것으로 텍스트 설명(text description)과 다른 개념으로의 참조(#conceptID 형식)가 있고 온톨로지 외부의 지식 자원으로의 링크를 지정하는 URI 주소 값이 있다.

개념의 연관관계 정의와 연관관계 구성 과정에서는 개념들 사이의 의미적 연관성을 파악하고 연관성을 설명하는 개념 및 용어를 정의한 다음 이 관계를 이용하여 두 개념을 연결하는 과정이다. 기업에서는 하나의 물품을 부서에 따라 서로 다른 명칭으로 부르는 경우가 많다. 이 경우 표준 용어만 정의할 경우 사용자가 익숙한 용어로 검색하는 것이 불가능해진다. 그러므로 이들 용어를 동의어 관계로 연결한다. 이외에도 대체관계, 부산물관계, 조합 불가 관계 등 여러 의미적 연관성을 부여할 수 있다.

온톨로지 평가 과정에서는 개념적 차원에서 온톨로지의 크기, 깊이, 구성 등을 평가하는 과정으로서 전문가에 의해 질문지와 정답지를 미리 작성

한 다음 질문지의 각 문제에 대해 온톨로지를 탐색할 경우 해답을 찾을 수 있는가를 평가한다. 이러한 평가과정을 통해 개념 사이의 연결 오류, 개념 정의 오류등을 파악하고 온톨로지의 내용이 충분하지도 파악한다.

### 4.3 구현 단계

구현단계에서는 컴퓨터상에서 온톨로지를 구축하고 이를 서비스하는 단계로서 비형식언어로 작성된 온톨로지를 형식언어로 변환하는 과정과 이 언어를 해석하고 시스템 내에 온톨로지를 생성한 다음 사용자의 질의에 검색 결과를 제공할 수 있는 시스템을 구현하는 과정으로 이루어진다.

본 논문에서는 물품 온톨로지의 데이터 모델로 토픽맵을 사용한다. 즉 물품 온톨로지 자체가 토픽맵인 것이다. 이 토픽맵을 기술하는 형식 언어로서 XTM(XML Topic Maps)[9]이 있다.

컴퓨터상에 토픽맵을 구축하기 위해서는 토픽맵의 내용을 XTM 문서로 작성하고 이 XTM 문서를 토픽맵 응용프로그램을 이용하여 메모리내에 토픽맵으로 구축한 다음 데이터베이스에 저장하거나 XTM 문서로 다시 변환한다. <표 1>은 하드디스크 개념을 XTM 구문으로 기술한 것을 보이고 있다.

구현단계의 PSI(Public Subject Indicator) 작성은 토픽맵 모델에서 토픽(topic) 즉, 물품 온톨로지의 하나의 개체에 유일성을 부여하는 것으로 온톨로지 상위에 존재하는 약속된 설명이라고 볼 수 있다. 토픽맵에서 서로 다른 이름의 개체가 동일한 PSI를 가질 경우 이름과 상관없이 동일한 하나의 개체인 것으로 인식하게 된다.

참조 온톨로지는 물품 온톨로지의 스키마 온톨로지이다. 참조 온톨로지를 먼저 정의함으로써 물품 온톨로지를 구현하거나 또는 유지보수 과정에서 온톨로지의 정당성을 검사할 수 있으며 일관된 형식을 유지하도록 유도할 수 있다. 또한 물품 온톨로지를 생성할 도메인이 다르더라도 동일한 참조 온톨로지를 사용할 경우 속성명, 관계명, 관계구조 등에서 일치함으로써 쉽게 통합할 수 있다. 여러 공장 또는 계열사를 가지는 기업에서 서로 다른 유형의 품목들에 대해 각각 물품 온톨로지를 구축한 다음 이들을 하나로 통합 또는 연결시키고자 할때 참조 온톨로지에 의거하여 통합하게 된다.

<표 1> 하드디스크를 기술하는 XTM 구문

```

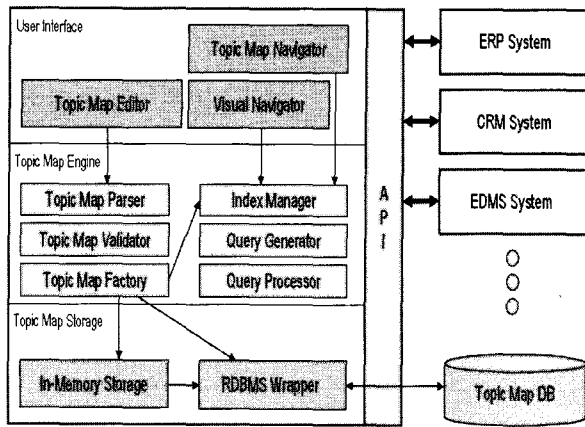
<topic id="hard_disk">
  <instanceOf>
    <topicRef xlink:href="#disk_storage"/>
  </instanceOf>
  <basename><basenameString>하드디스크
    </basenameString></basename>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#description"/>
    </instanceOf>
    <resourceData>하드디스크는 보조기억...
    </resourceData>
  </occurrence>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#transfer-rate"/>
    </instanceOf>
    <resourceData>전송율은 보통 5400rpm..
    </resourceData>
  </occurrence>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#conn-type"/>
    </instanceOf>
    <resourceData>연결형식에 따라 E-IDE..
    </resourceData>
  </occurrence>
</topic>
<topic id="HD60TGK1">
  <instanceOf>
    <topicRef xlink:href="#hard_disk"/>
  </instanceOf>
  <basename><basenameString>시게이트..
    </basenameString></basename>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#description"/>
    </instanceOf>
    <resourceData>이 디스크는 시게이트...
    </resourceData>
  </occurrence>
</topic>

```

물품 온톨로지를 XTM 구문에 따라 명세화를 하고 나면 이를 컴퓨터상에 구현하여 사용자에게 물품 정보를 제공하거나 사용자의 요구에 따라 온톨로지의 내용을 변경할 수 있어야 한다. 이를 위해 본 논문에서는 토픽맵을 처리할 수 있는 토픽맵 응용 시스템을 구현하였다.

## 5. 토픽맵 응용 시스템 구조

토픽맵 응용 시스템은 기본적으로 XTM 파일들을 해석하여 토픽맵을 생성하고 검색 및 저장하는 기능을 가진다. <그림 5>는 토픽맵 응용 시스템의 구조를 개략적으로 보이고 있다.



<그림 5> 토픽맵 응용 시스템의 구조

시스템의 주요 구성요소는 사용자 인터페이스, 토픽맵 생성 및 검색 엔진, 토픽맵 저장소, 토픽맵 API 등의 4가지로 나누어볼 수 있다. 사용자 인터페이스는 JSP와 스트럿츠(struts)로 구현된 웹 인터페이스를 제공하는 것으로 토픽맵의 요소를 추가, 변경, 삭제할 수 있는 편집기(Topic Map Editor)와 토픽맵의 내용을 텍스트 기반으로 검색하는 토픽맵 네비게이터(Topic Map Navigator) 및 그래픽 기반으로 검색하는 비주얼 네비게이터(Visual Navigator)로 구성되어 있다.

토픽맵 엔진은 토픽맵 생성을 위한 부분과 생성된 토픽맵의 내용 검색을 위한 부분으로 나누어지는데, 생성은 토픽맵 파서(Topic Map parser), 정당성 검사기(Topic Map Validator), 구축기(Topic Map Factory)로 구성되고 검색은 색인의 생성 및 관리를 담당하는 색인 관리자(Index Manager), 키워드 검색을 지원하는 질의 생성기(Query Generator), 그리고 질의를 처리하는 질의 처리기(Query Processor)로 구성되어 있다.

토픽맵 저장은 메모리 내에 생성하는 인-메모리 저장기(In-Memory Storage)와 관계형 데이터베이스로 사상시켜 저장하는 RDBMS 래퍼(RDBMS

Wrapper)로 구성된다.

토픽맵을 생성하는 과정을 살펴보면, 먼저 입력으로 들어오는 XTM 파일의 구문을 XML 프로세서를 통하여 파싱을 한 다음 다시 토픽맵 파서에서 엘리먼트 종류에 따른 추가 파싱 작업을 해주어야 한다. XTM 구문으로 작성한 문서를 XML 프로세서의 SAX 파서의 입력으로 하면 SAX 파서는 입력된 문서를 처음부터 스캔하면서 엘리먼트 시작과 엘리먼트 끝 이벤트를 일으킨다. SAX 파서는 각 엘리먼트에 따라 처리 루틴을 가지고 있는 것이 아니고 단지 XML 문서의 엘리먼트를 추출하고 문서의 정당성을 검사하는 것이므로 각 엘리먼트의 타입에 따라 세부 처리하는 루틴은 따로 구현하여야 한다. 이 역할을 담당하는 클래스가 TM parser 클래스이다.

TM parser 클래스는 SAX 파서로부터 엘리먼트 시작 이벤트를 받으면 어떤 엘리먼트인지를 보고 각 엘리먼트에 따라 필요한 초기화 작업을 한 다음 그 엘리먼트를 실제로 처리하는 TM builder 클래스로 넘겨준다. TM builder 클래스는 하위의 구현 클래스들을 이용하여 실제 토픽맵을 구축하는 클래스이다. TM builder가 한 엘리먼트에 해당하는 토픽 객체를 생성하고 완료하면 TM parser는 다시 그 엘리먼트 끝 이벤트를 처리하여 한 엘리먼트 처리를 종료한다. 그런 다음 SAX 파서가 넘겨주는 다음 엘리먼트를 동일한 방법으로 처리한다.

생성된 토픽맵에 대한 검색은 최상위 토픽에서부터 원하는 토픽까지 연결 구조를 따라 찾아가는 방법과 키워드 검색에 의해 같은 키워드를 토픽명이나 어커런스에 가지는 토픽들을 검색한 다음 원하는 토픽을 찾는 방법이 제공된다. 특정 토픽을 선택하게 되면 해당 토픽의 상세 정보를 보이는데, 토픽의 텍스트 설명, 자원 연결 정보 등의 어커런스와 다른 토픽과의 상하 계층 관계 및 의미적 연관 관계를 보여준다.

토픽맵의 표준 질의 언어는 TMQL(Topic Map Query Language)이지만 이 언어는 아직 드래프트 수준으로 정식으로 릴리즈되지 못한 상태이다. 따라서 본 시스템에서는 온토피아(<http://www.ontolpia.net>)에서 제안한 tolog[10]를 토픽맵 질의어로 사용한다. 다음은 하드 디스크 토픽의 하위에 연결된 모든 토픽들을 검색하는 tolog 질의이다.



instance-of(\$DISK, hard\_disk)?

본 시스템의 질의 생성기는 웹 인터페이스의 질의 폼(form)에 입력한 내용을 근거로 tolog 질의를 생성한다. 질의 처리기는 생성된 또는 사용자가 입력한 tolog 질의를 처리하여 토픽맵으로부터 검색 결과를 생성한다.

<표 2> 토픽, 어커런스, 연관관계 테이블 구조

Topic 테이블			
컬럼	데이터형	제약조건	설명
topicID	int	not null	시스템ID
basename	string	not null	토픽명
parentID	int	null	상위토픽ID
scope	string	null	범위
theme	string	null	주제
linktype	string	not null	토픽유형
subject	string	null	subject값
resourceID	string	null	자원 ID
topicmap	int	not null	토픽맵
ID	string	null	유저토픽ID

Occurrence 테이블			
컬럼	데이터형	제약조건	설명
topicID	int	not null	토픽ID
occno	int	not null	어커런스NO
typeID	int	null	어커런스형
scope	string	null	범위
theme	string	null	주제
resData	string	null	자원설명
resRef	string	null	자원링크
resourceID	string	null	자원 ID
topicmap	int	not null	토픽맵

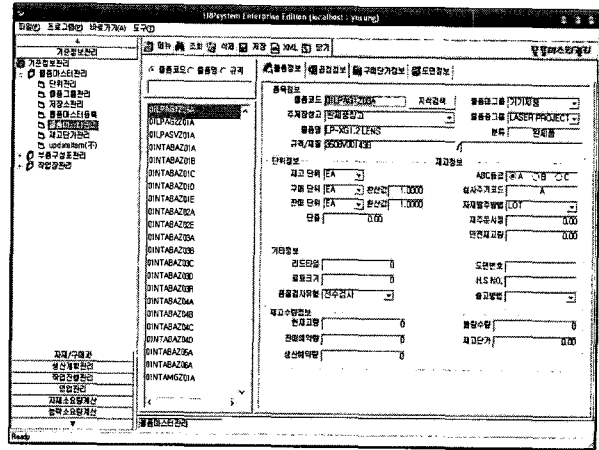
Association 테이블			
컬럼	데이터형	제약조건	설명
topicID	int	not null	assoc. ID
parentID	int	not null	상위assoc.ID
type	string	null	assoc.유형
scope	string	null	범위
theme	string	null	주제
cardinality	int	not null	멤버수
reified	int	not null	0/1
topicmap	int	not null	토픽맵

생성된 토픽맵은 관계형 데이터베이스에 저장할 수 있다. 이를 위해 토픽맵 모델에 적합한 관계형 데이터베이스 스키마를 설계하였으며 RDBMS 래퍼는 토픽맵 모델을 관계형 스키마로 사상시킴

으로써 토픽맵의 요소들을 각각의 테이블에 레코드로 추가하여 저장한다. <표 2>는 토픽, 어커런스, 연관관계를 저장하기 위한 테이블 구조를 보이고 있다.

## 6. 구현 결과

본 논문의 토픽맵 응용 시스템은 아래 <그림 6>의 상용 ERP 시스템과 연계하였으며 핸드폰 및 전자제품의 사출품 제조회사의 실 데이터를 기반으로 실험하였다.



<그림 6> ERP 시스템의 물품 마스터 관리화면

물품 온톨로지를 구축하기 위해 전문가탐을 구성하고 ERP 시스템에 등록된 물품마스터 데이터의 품명 기준으로 전체 리스트를 작성한 다음 오류, 중복 등의 문제있는 데이터를 제거하는 데이터 정제 작업을 선행하였다.

데이터 정제 작업 이후 물품과 관련된 설계도, 명세서, 계약서 등 문서와 이미지 자료들을 수집하였고 코드 부여 체계를 조사하였다.

자료 수집 이후 전문가에 의해 개념 추출과 분류 작업을 수행하였고 주요 품목의 특성을 파악하여 속성을 정의하였다. 그리고 품목간 관계에 따른 의미적 연관성을 정의하였다. 이러한 과정으로 생성된 물품 온톨로지의 구조를 <표 3>에 몇 가지 기준으로 수치화 하였다.

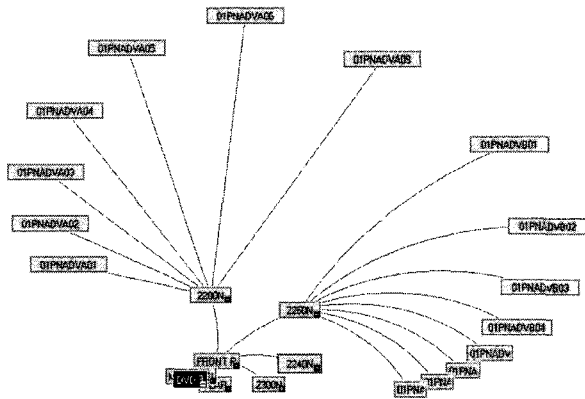
참조 온톨로지에는 PSI를 가지는 토픽들과 속성 및 연관관계를 정의하는 토픽들로 생성되었다. PSI를 가지는 토픽은 표준 용어 토픽으로서 '통신제품',

<표 3> 생성된 물품 온톨로지의 특성

온톨로지	토픽수	연관관계수	깊이	크기
참조	1281	57	0	1.6MB
물품	7382	1584	5	14.8MB

‘핸드폰’, ‘사출품’과 같이 다른 도메인에서도 참조 가능한 용어들을 정의하고 있다. 속성 정의 토픽은 물품 온톨로지의 각 개념의 속성인 ‘재질’, ‘규격’, ‘성분’, ‘색상’ 등의 용어를 정의하는 토픽이다. 연관관계 정의 토픽은 물품 온톨로지의 개념들 사이에 ‘동의어’, ‘의존관계’, ‘조합 불가관계’ 등의 의미적 연관성을 표현하는 용어를 정의하는 토픽이다.

물품 온톨로지는 ERP 시스템의 물품마스터에 저장된 5300여개 물품코드 토픽과 이들을 분류하는 상위 품목 토픽들을 정의하고 있으며 1500여개의 연관관계 토픽을 가지고 있다. <그림 7>에서는 물품 온톨로지의 일부분을 그래프 방식의 비주얼 화면으로 보이고 있다. 그래프 방식의 비주얼 검색은 노드와 노드 사이의 연결선을 따라서 특정 토픽을 중심으로 연결된 다른 토픽들을 쉽게 파악할 수 있도록 한다.

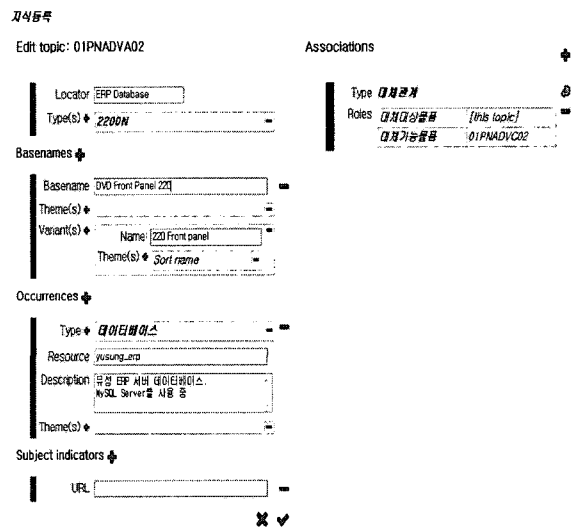


<그림 7> DVD케이스사출품 토픽과 그 하위 토픽들의 연결구조를 보이는 비주얼 검색 화면

<그림 7>은 중심토픽인 ‘DVD 케이스 사출품’ 토픽과 이에 계층적으로 연결된 하위 토픽들을 보여준다. 그래프를 생성하기 위해 본 논문에서는 스타트리(www.inxight.com) SDK를 사용하였으나 이외 여러 하이퍼볼릭 그래프 도구들을 사용하여

토픽맵과 연결할 수 있다.

<그림 8>은 물품 온톨로지에 새로운 물품 토픽을 등록하는 것을 보이고 있다. 토픽 ID는 01PNADVA02 이고 기본명은 DVD Front Panel 220을 기술되어 있다. 그리고 데이터베이스 속성을 가지고 있으며 01PNADVC02와 대체관계를 가지는 것으로 연결되어 있다. 온톨로지상에 등록되는 물품이 ERP 시스템의 물품마스터에 없을 경우 자동으로 물품마스터에 추가된다.



<그림 8> 물품토픽 등록을 처리하는 JSP 화면

## 7. 결 론

본 논문에서는 기업의 핵심 데이터 중의 하나인 물품마스터의 관리 어려움에서 오는 데이터 부정확 문제를 해결하기 위한 접근법으로 물품 온톨로지 및 온톨로지 관리 시스템을 제안하였다.

기업의 기간 시스템인 ERP 시스템의 운영이 성공적이지 못한 여러 가지 이유 중의 하나가 ERP 데이터베이스에 저장된 데이터가 부정확하기 때문이다. 이러한 데이터의 부정확성을 야기시키는 문제는 데이터베이스 구조 자체의 비확장성 및 비유연성을 들 수 있다. 또한 물품에 대한 정보가 ERP 시스템 내에 통합되지 않고 분산되어 있기 때문이며 물품코드 부여 체계가 표준화 및 시스템화 되어 있지 않기 때문이다.

물품 온톨로지는 기업의 유형, 무형의 물품에 대

한 지식을 표준화, 구조화함으로써 개별 물품코드의 나열이 아닌 계층적 품목 분류 체계와 물품 사이의 의미적 연관관계를 제공한다. 또한 ERP, CRM 등 기간 시스템의 데이터베이스와 설계서, 공정도, 시방서 등 물품과 관련된 지식 자원들을 통합하여 검색할 수 있도록 지원한다.

본 논문에서는 물품 온톨로지를 생성하기 위한 구축 프로세스를 3개의 단계와 14개의 과정으로 세분화하여 제시하였으며 물품 온톨로지의 논리적 구조를 설계하였다. 그리고 토픽맵 모델 기반의 물품 온톨로지를 구축할 수 있도록 토픽맵 응용 시스템을 구현하였다.

또한 상용 ERP 시스템과 토픽맵 응용 시스템을 연계하여 실제 제조기업의 물품 온톨로지를 구축해 보았으며 물품에 관한 지식 검색과 신규 물품 코드의 생성을 물품 온톨로지를 이용하여 처리할 수 있었다.

ERP 데이터의 정확성 유지는 물품 온톨로지의 구축과 함께 사용자의 의사소통과 협동이 중요한 요소이므로 이를 평가하기 위해서는 장기간의 운영이 필요하다. 그러나 물품 온톨로지의 구축과 이를 관리하는 시스템은 그 자체만으로도 기업의 물품 지식을 효율적으로 관리할 수 있는 가능성을 제공한다고 볼 수 있다.

## 참 고 문 헌

- [1] 최성, 김계철(2003). ERP와 데이터 리엔지니어링. 한국정보처리학회, 10(1), pp.121-133.
- [2] 윤정모, 김계철(1999). ERP 시스템의 개념과 발전과정. 한국정보처리학회, 6(5), pp.9-18.
- [3] Steve Pepper(2000). TAO of the Topic Maps : finding the way in the age of infoglut. XML Europe 2000.
- [4] O. Lassila and R. Swick(1999). Resource description framework(RDF) model and syntax specification. W3C.
- [5] P. Ciancarini, M. Pirruccio, F. Vitali, R. Gentilucci, V. Presutti(2003). Metadata on the Web. On the integration of RDF and Topic Maps. In Proceedings of Extreme Markup Languages.
- [6] M. Uschold and M. Gruninger, ONTOLOGIES: Principles, Methods and Applications, Knowledge Engineering Review, 11(2), 1996.
- [7] M. Uschold, Building Ontologies: Towards A Unified Methodology, In Proceedings of Expert Systems, 1996.
- [8] M. Fernandez-Lopez, A. Gomez-Perez, and J. Pazos Sierra, Building a Chemical Ontology Using Methontology and the Ontology Design Environment, IEEE Intelligent Systems, 14(1), pp.37-46, 1999.
- [9] TopicMaps.org Authoring Group. XML Topic Maps(XTM) 1.0. [Online] available: <http://www.topicmaps.org/xtm/1.0/>.
- [10] L. M. Garshol(2003). tolog 1.0. available: <http://www.ontopia.net/topicmaps/materials/tolog-spec.html>.



정 현 숙 (Hyun-Sook Chung)

- 정회원
- 1993년 2월 : 대구가톨릭대학교 물리학과 (이학사)
- 2월 : 대구가톨릭대학교 컴퓨터 공학과 (공학석사)
- 2003년 8월 : 연세대학교 컴퓨터과학과 (공학박사)
- 2004년 9월 ~ 현재 : 가톨릭대학교 컴퓨터정보 공학부 교육전담 전임강사
- 관심분야 : 모션그래픽스, 이러닝, 시맨틱웹, 온톨로지