

토픽맵을 이용한 이 기종 상품분류체계 온톨로지 통합에 관한 연구

A Study of Integrating Ontologies of Heterogeneous Product Classification Schemes Using XML Topic Maps(XTM)

고세영(Se Young Ko)*, 김성혁(Sung Hyuk Kim)**

초 록

상품분류체계 중 가장 많이 사용하는 UNSPSC와 HS를 선정하여 각 상품분류체계의 구성 및 구조와 그 특징을 파악하고 분류명에 대한 온톨로지를 분석하였다. 상품분류체계의 번호체계 및 각 자릿수에 대한 이해를 충분히 하고 UML을 이용하여 모델링을 하였다. 상품분류체계 중 대상 범위에 대한 토픽맵을 개발한 뒤 그래프로 표현하고 최종적으로 개별 상품분류체계에 대한 토픽맵을 통합하여 상품분류체계의 온톨로지를 통합함으로써 계층 및 구조가 상이한 상품분류체계 간의 정보 교환을 용이하게 할 수 있게 하였다.

ABSTRACT

The Topic Maps paradigm allows people and organizations to integrate and merge heterogeneous products classification systems such as UNSPSC and HS. Merging their product ontologies could combine information about classification scheme for products. We analyzed two product classification schemes for UML modeling and developed an integrated TM for watches. Examples in XTM syntax show how UNSPSC and HS can be integrated by merging their ontology.

키워드 : 토픽맵, 온톨로지, 상품분류체계, 이 기종 상품분류체계 통합

XML Topicmaps, Ontology, Product Classification, XTM

* 주)인포와이즈

** 숙명여자대학교 정보과학부

1. 서 론

1.1 연구배경 및 필요성

e-Business가 활성화 되면서 다양한 서비스를 제공하는 많은 거래 사이트들이 생겨났고 방대한 양의 상품들이 인터넷을 통해 거래되고 있다. 수출입, 기업간의 상거래시 관세, 유통 등의 관리를 위해 여러 번의 상품코드변환 즉, 상품분류체계의 정보 교환이 일어나고 있으며 현 체계에서는 그 작업이 번거로울 뿐만 아니라 이에 따라 많은 시간과 노력이 들고 많은 비용이 발생된다. 따라서 원활한 전자상거래를 위해서는 물류 이동 시마다 빈번히 발생하는 상품코드를 자동으로 부여하거나 변환할 수 있어야 한다. 그러나 기존의 테이블매핑의 상품분류체계 정보 교환 방식으로는 이를 해결하기가 어렵다.

본 연구에서 제시하는 이 기종 상품분류체계 온톨로지 통합은 상품분류체계의 온톨로지를 활용하여 각기 다른 상품분류체계를 상호 유동적으로 쉽게 변환하고 검색할 수 있도록 하는 기능을 지원한다. 상품분류체계가 주기적으로 수정, 변경, 삭제되더라도 그 때마다 각 상품분류체계의 분류단계와 분류명, 분류코드 등을 검색하여 수정하는 등의 번거로운 작업이 필요 없게 되며 그에 따른 많은 시간과 노력이 줄어들게 된다. 또한 상품정보를 표현하는 데 있어 단순하게 코드를 중심으로 하지 않고 분류명의 온톨로지를 분석하여 토픽맵을 구성하고 표현하였으므로 상품분류체계에 관계없이 구조가 상이한 상품분류체계를 통합할 수 있으며 이를 이용하여 상품정보

를 교환하고자 하는 경우에도 상품분류체계의 구조를 손상시키지 않아 상품분류체계가 갖는 고유한 구조를 유지하면서 서로 정보를 자유롭게 교환할 수 있다. 또한 상품 검색 시 이용자가 원하는 상품을 더 빠르고 더 정확하게 검색할 수 있도록 도와주는 기능을 한다.

1.2 연구범위 및 방법

본 연구에서 토픽맵을 이용한 상품분류체계 온톨로지 통합 대상이 되는 상품분류체계에 온라인과 오프라인을 통틀어 가장 많이 사용하는 UNSPSC와 HS를 선정하였고 선정된 상품분류체계에서 토픽맵을 구현할 범위를 선택하였다. 일반적으로 상품분류체계는 전체 상품을 그 대상으로 하므로 상당히 광범위하다. 본 연구에서는 전체를 대상으로 하는 것은 무리라는 판단 하에 상품분류체계 중 일부를 프로토타입으로 구현하도록 하였다.

다음 단계로 선정된 각 상품분류체계의 구성 및 구조와 그 특징을 파악하고 분류명에 대한 온톨로지를 분석하였다. 또한 상품분류체계의 번호체계 및 각 자릿수에 대한 이해를 충분히 하고 상품분류체계 계층을 조사하였으며 모델링을 위해 UML을 이용하였다.

토픽맵 도구 (Topic Map Designer V1.1과 XMLSPY)를 이용하여 대상 범위의 상품분류체계에 대한 토픽맵을 개발한 뒤 그래프로 표현하고 최종적으로 개별 상품분류체계에 대한 토픽맵을 통합하여 상품분류체계의 온톨로지를 통합함으로써 계층 및 구조가 상이한 상품분류체계 간의 정보 교환을 용이하게 할 수 있게 하였다.

2. 이론적 배경

2.1 토픽맵(Topic Maps)

토픽맵은 표준화된 지식관리 데이터 모델로 등장한 개념으로, 대용량의 정보를 분류하고 구조화하며 의미론적인 연관관계를 설정하여 원하는 정보를 쉽고 정확하게 찾을 수 있도록 한다. 일종의 색인의 역할을 하는 것으로 지도(map)와 같은 첨부된 레이어(layer)를 통해 원하는 정보를 전달하는 데 그 목적이 있다[1].

토픽맵은 토픽의 위치정보를 나타내는 다양한 차원의 토픽 스페이스(topic space)를 정의한다. 특정 문헌이나 문서 내 표현하고자 하는 내용을 토픽으로 정의하고 이들 간의 관계를 나타내는 것이다. 정보관리에 있어 GPS와 같은 역할을 하는 토픽맵은 정보관리 및 지식표현의 각 도메인 간의 다리역할을 하며 대용량의 데이터와 매시간 증가하는 정보 pool의 조직 및 네비게이션을 위한 최상의 솔루션이 된다[5].

최초로 제안된 토픽맵 표준 스펙은 2000년 1월 HyTM으로 표현된 "ISO/IEC 13250 Topic Maps"이다. ISO/IEC 13250 제정으로 정보 자원들을 상호 연관성에 따라 연결하고 조직하여 지식 구조를 기술하는 것이 가능해졌다. 1996년 ISO는 topic map을 수용하여 이에 대한 많은 논의와 연구를 거친 후 1998년 10월 투표를 실시하였다. 그 결과 2000년 1월 ISO 13250으로 표준화 되었다. ISO 13250은 SGML architecture에 따라 기술되고 HyTime으로 정의되어 있다[16].

W3C에서는 토픽맵이 등장하기 이전부터 웹에 의미정보를 표현할 수 있는 방법을 연구해왔다. 즉, 웹 페이지들에 그 페이지 자체를 설명하는 메타데이터(metadata)를 기술함으로써 웹 페이지의 하이퍼링크 뿐만 아니라 시맨틱(semantic)링크에 의해서 보다 정확한 검색을 하고자 하였다. 그 결과로 RDF(The Resource Description Framework)라는 모델이 제시되었으며 웹 자원의 시맨틱 네트워크(semantic network)를 구축하기 위한 도구로 사용되고 있다. 또한 토픽맵과의 통합 및 상호 교환에 관한 연구가 수행되고 있다.

토픽맵은 향후 여러 분야에 활용될 수 있다. 콘텐츠관리시스템의 경우 토픽맵을 이용하여 콘텐츠들의 연관관계를 설정하고 구조화하여 관리할 수 있다. 또, 전자카탈로그의 경우 UNSPSC와 같은 상품속성분류 체계와 토픽맵의 접목을 통해 유사 상품 내 네비게이션(navigation) 및 검색이 용이해질 수 있다. 그 외에도 Knowledge Representation, Ontological Engineering, Semantic Network, 그리고 RDF 등과의 통합 및 응용을 통해 사용자들에게 보다 용이하게 자동화된 의미기반의 표현을 제공할 수 있을 것이다[1].

토픽맵의 기본 요소는 토픽(Topic), 어소시에이션(Association), 어커런스(Occurrence)로 나누어 볼 수 있다.

2.2 선행연구

Helka Folch(2000)는 전기회사인 Electricite de France (EDF)의 연구개발부서의 지식관리시스템을 개발하는 Scriptorium 프로젝트

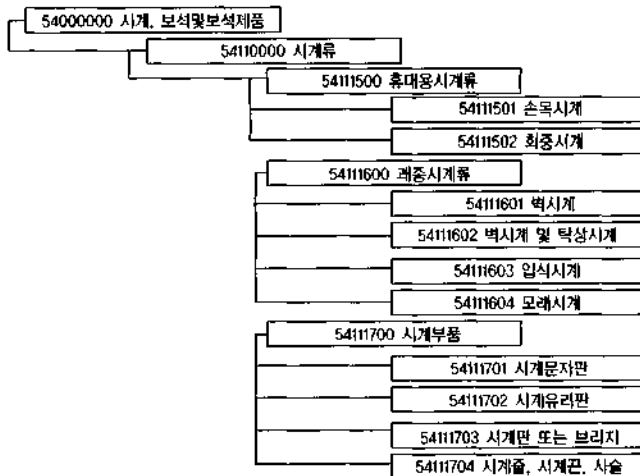
의 프레임워크로 토픽맵을 이용하였다. 대용량의 지식 수집 및 이종의 언어자료(corpora)와 자료들의 시맨틱을 브라우징하고 네비게이션 하는데 토픽맵을 이용한 것이다. 이종의 문서 컬렉션을 전자문서화 하는데 있어 시소러스나 색인과 같은 기존의 네비게이션 방법은 이미 정의된 시맨틱 카테고리나 분류에 기반하기 때문에 동적이고 비조직화된 텍스트를 기술하거나 브라우징하는데 어려움이 많았다. 이들은 텍스트 데이터의 통계적인 분석을 통해 언어자료로부터 시맨틱 클래스를 추출하고 관리 효율성이 가장 높은 사이즈의 언어자료를 추출함으로써 문서 컬렉션에 대한 이전과는 다른 관점을 만들어냈다. 다음 Topic Map Standard (ISO/IEC 13250)를 이용하여 문서 컬렉션에 대한 토픽맵을 작성하였는데 이는 성공적으로 문서 컬렉션에 대한 시맨틱 인터페이스를 제공하였다. 네비게이션은 언어자료의 시맨틱 스페이스를 3D로 표현함으로써 해서 그 효과를 배가시켰다[3].

Mary Nishikawa(2002)는 정유 회사인 Schlumberger의 인트라넷 내 정보를 효율적으로 분류, 조직하고 좀 더 개념을 확장하여 기술정보를 고객 및 파트너와 공유하기 위해 토픽맵을 활용하였다. 특히 전자상거래에 많이 이용되는 상품분류체계인 ECCMA의 UNSPSC 중 segment 71 Mining and Oil and Gas Service를 대상으로 이를 XTM으로 구현하여 토픽맵의 PSI(Published Subject Indicator)를 특정 주제에 대한 정보의 연결점으로 이용하여 정보관리를 하는 방안을 제시하였다[4].

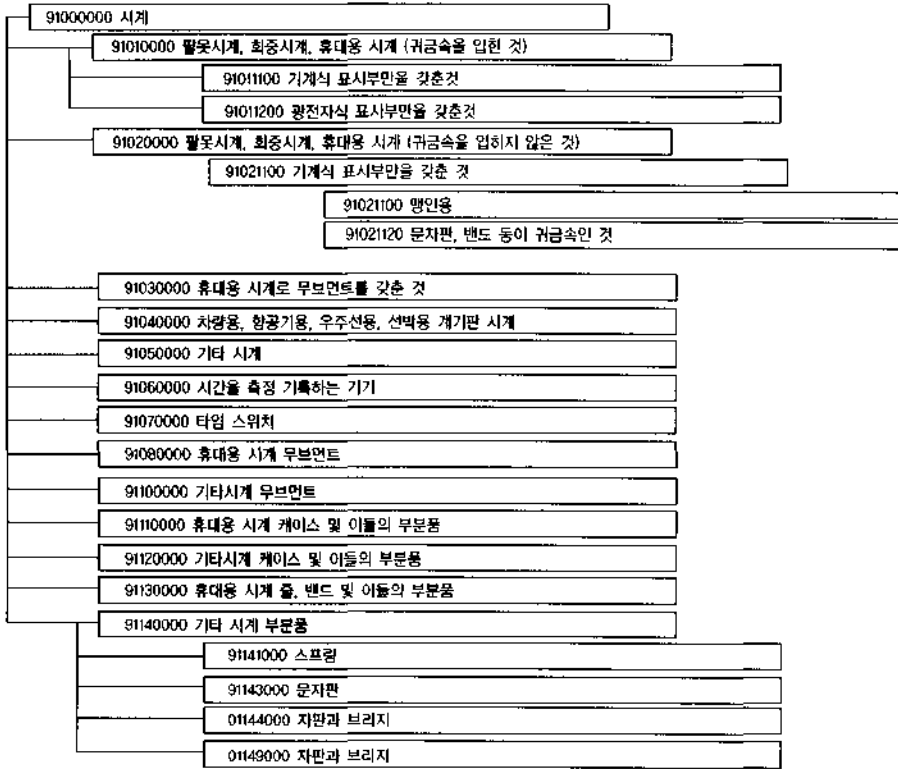
3. 토픽맵 설계를 위한 모델링

3.1 계층분석

UNSPSC에서 시계 및 부품에 대한 분류체계를 살펴보면, '시계, 보석 및 보석제품'이 segment로 최상위 분류단계 이고 다음 단계



〈그림 1〉 UNSPSC 시계 분류체계



〈그림 2〉 HS 시계 분류체계 중 일부

는 '시계류'로 분류명이 할당된 family이다. '시계류' family는 '휴대용시계류', '폐종시계류', '시계부품' 이렇게 3개의 class를 가지며 각 class는 최종단계인 commodity를 갖는다는 것을 알 수 있다. 또한 분류단계와 그 명칭을 분석하면 시계의 부분품의 종류가 분류의 기준임을 알 수 있다.

분류체계의 계층 및 구성을 도식화하면 〈그림 1〉과 같다.

HS에서 시계 및 부품에 대한 분류체계를 살펴보면, UNSPSC와는 다르게 상당히 세분되어 있음을 알 수 있다. 우선 시계를 그 종류에 따라 분류하지 않고 휴대용 시계와 일반시

계로 나누고 그것과 함께 귀금속을 입힌 것과 그렇지 않은 것으로 세분하여 chapter를 구성하였다. 또한 chapter에 시계의 무브먼트와 케이스 및 부분품을 분류하고 이들을 세분한 항목들을 heading과 subheading에 포함시켰다. 분류체계가 상세하다는 것 외에 UNSPSC와 차이를 보이는 부분으로는 분류기준에 있다. 단순히 종류만으로 분류하지 않고 용도와 재질을 분류기준으로 이용하여 반영시켰다는 점이다. 각 항목별로 중요하다고 판단되는 기준을 분류단계 별로 다르게 적용했다고 할 수 있다.

분류체계 계층 및 구성의 일부를 도식화하면 〈그림 2〉와 같다.

3.2 온톨로지 분석 및 구성

각 분류체계의 분류명을 분석하여 별도의 온톨로지를 구성하였다. UNSPSC에서는 주로 시계의 종류 및 부품으로 분류되어 있는 반면 HS에서는 시계의 종류와 용도, 재질을 기준으로 분류되어 있다. 두 분류체계의 분류명을 하나씩 나열하여 이에 대해 종류를 가장 큰 분류로 하고 용도와 재질에 따라 세분류로 나누어 분류명에 대한 온톨로지를 XML화 하였다.

3.3 UML 설계

UML은 분석 및 설계에 있어 효과적인 표기법을 제공하는 언어이다. 다시 말해, UML은 시스템 분석과 설계 과정상의 산출물을 명세화, 시각화, 문서화 하여 표준화 시키는데 사용되는 것으로 다양한 시스템 분석 및 설계 방법론이 통합된 언어라고 할 수 있다[2].

UML을 구성하는 여러 다이어그램 중 Class, Interface, Relation을 이용하여 각 클래스와 그 관계를 가시화하고 상세한 내용을 명

```

<?xml version="1.0" encoding="UTF-8"?>
<ontogrp id="clock">시계
  <onto id="watch">시계류</onto>
  <onto id="cl_po">휴대용시계류</onto>
  <onto id="wat">손목시계
    <subonto id="w_ca">손목시계 케이스귀금속제
      <kind id="w_ma">기계식</kind>
      <kind id="w_el">광전자식</kind>
      <kind id="w_er">전기구동식</kind>
      <kind id="w_au">자동권식</kind>
      <kind id="w_et">기타</kind>
    </subonto>
    <subonto id="w_no">손목시계 일반
      <kind id="wno_ma">기계식
        <use id="w_fb">맹인용</use>
        <leafkind id="w_b">문자판 밴드동이 귀금속</leafkind>
      </kind>
      <kind id="wno_el">광전자식
        <leafkind id="el_b">배터리 또는 축전지 구동식</leafkind>
      </kind>
      <kind id="wno_er">전기구동식</kind>
      <kind id="wno_au">자동권식</kind>
      <kind id="wno_et">기타</kind>
    </subonto>
  </onto>
  <onto id="part">시계부품
    <subonto id="pa_sp">스프링</subonto>
    <subonto id="pa_jw">시계용보석</subonto>
    <subonto id="pa_pl">문자판</subonto>
    <subonto id="pa_br">지관 또는 브리지</subonto>
    <subonto id="pa_gl">시계유리판</subonto>
    <subonto id="pa_ot">기타</subonto>
  </onto>
</ontogrp>

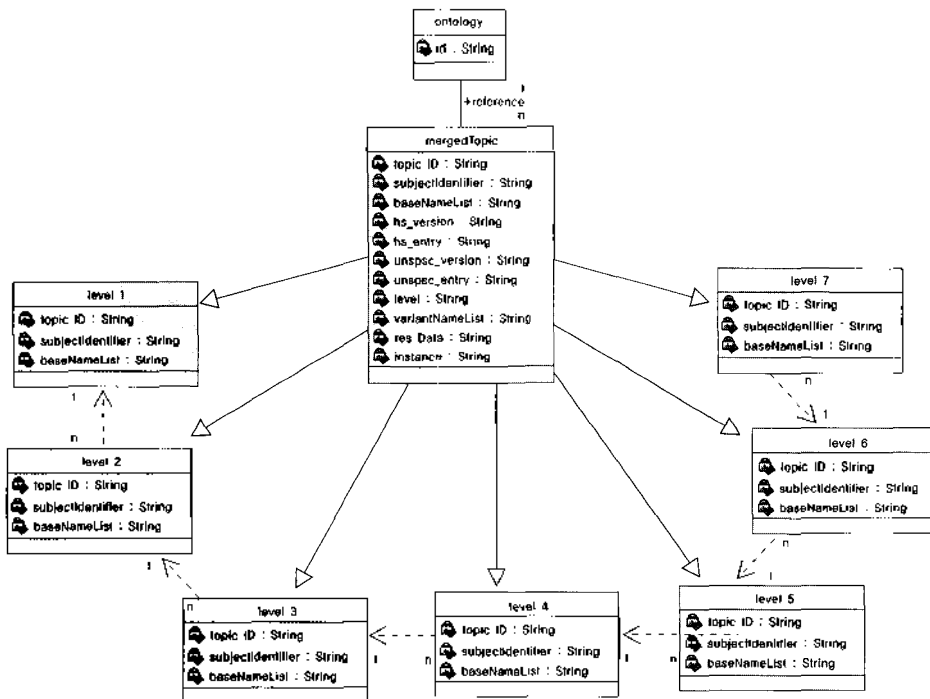
```

<예제 1>

세화 하는 클래스 다이어그램(Class Diagram)으로 설계가 이루어졌다. UNSPSC 토픽맵의 토픽은 문자열로 구성된 ID, SI, basename, version, entry를 갖고 그 구성은 가장 상위단계인 segment, segment에 종속되는 family, family에 종속되는 class, class에 종속되는 commodity로 분류단계가 구성되어 있다. 상위단계와 하위단계는 1:N의 관계이다. HS 토픽맵의 토픽은 문자열로 구성된 ID, SI, basename, version, entry를 갖고 그 구성은 가장 상위단계인 section, section에 종속되는 chapter, chapter에 종속되는 heading, heading에 종속되는 subheading으로 분류단계가 구성되어 있다. 상위단계와 하위단계는 1:N의 관계이다.

각 분류체계의 구성에 대한 분석과 온톨로지를 바탕으로 UNSPSC 토픽맵과 HS 토픽맵을 통합한 통합 토픽맵을 모델링하면 <그림 3>과 같다.

통합 토픽맵의 토픽은 문자열로 구성된 ID, SI, basename, 각 상품분류체계별 version, 상품분류체계별 entry, 단계를 나타내는 level, 다른 상품분류를 참조할 때 필요한 resData 항목을 포함한다. 그 구성은 가장 상위단계인 level1, level1에 종속되는 level2, level2에 종속되는 level3, level3에 종속되는 level4, level4에 종속되는 level5, level5에 종속되는 level6, level6에 종속되는 level7로 단계가 구성되어 있다. 상위단계와 하위단계는 1:N의 관계이다.



<그림 3> 통합 토픽맵 UML 설계

4. 모델링에 기반한 토픽맵 개발

계층 및 온톨로지 분석과 UML을 이용한 설계를 토대로 각 분류체계에 대한 토픽맵을 개발하였고 도구¹는 "Topic Map Designer V 1.1"과 "XMLSPY"를 사용하였다. Topic Map Designer V 1.1은 토픽맵 에디터 및 그래프 뷰어로 독일의 Dresden University of Technology의 Ronald Heckel이 개발한 도구이다. 기존의 토픽맵 에디터와 달리 네비게이션과 그래픽 뷰가 가능하다는 장점이 있어 본 연구에서 개발한 토픽맵을 네비게이션 하는 브라우저로 선정하였다.

4.1 UNSPSC 토픽맵

우선 분류체계의 항목을 의미하는 엔트리(entry), 버전관리를 위한 버전번호, 분류체계 계층 상의 segment, family, class, commodity 간의 관계를 토픽으로 선언하였다. 각 항목에 대해서는 분류번호를 토픽 ID로, 분류명을 basename으로 지정하였으며 해당 토픽에 대한 온톨로지를 subject indicator로 선언하여 XML로 작성한 온톨로지 파일에서 참조하도록 하였다. 또한 어소시에이션을 사용하여 해당 코드가 분류체계 내 어느 계층인지를 나타내고 토픽으로 선언한 상품분류체계의 버전을 표시하도록 하였다.

<예제 1>은 위에 언급한 모델링방법과 도구를 이용하여 구현한 UNSPSC 토픽맵의 일부이다.

<그림 4>는 Topic Map Designer를 이용하여 정의된 UNSPSC의 토픽 및 어소시에이션 및 관계를 나타내는 화면이다. 정사각형으로 표시된 것이 토픽이고 삼각형으로 표시된 것이 어소시에이션이다.

4.2 HS 토픽맵

HS분류체계를 토픽맵으로 개발하는 작업은 UNSPSC와 동일하게 이루어졌다. 분류체계의 항목을 의미하는 엔트리, 버전관리를 위한 버전번호, 분류체계 계층 상의 section, chapter, heading, subheading간의 관계를 토픽으로 선언하였다. 각 항목에 대해서는 분류번호를 토픽 ID로, 분류명을 basename으로 지정하였으며 해당 토픽에 대한 온톨로지를 subject indicator로 선언하여 XML로 작성한 온톨로지 파일에서 참조하도록 하였다. 또한 어소시에이션을 사용하여 해당 상품분류체계가 분류체계 내 어느 계층인지를 나타내고 토픽으로 선언한 상품분류체계의 버전을 표시하도록 하였다.

<예제 2>는 위에 언급한 모델링방법과 도구를 이용하여 구현한 HS 토픽맵의 일부이다.

<그림 5>는 Topic Map Designer를 이용하여 정의된 HS의 토픽 및 어소시에이션과 그 관계를 나타낸 화면이다.

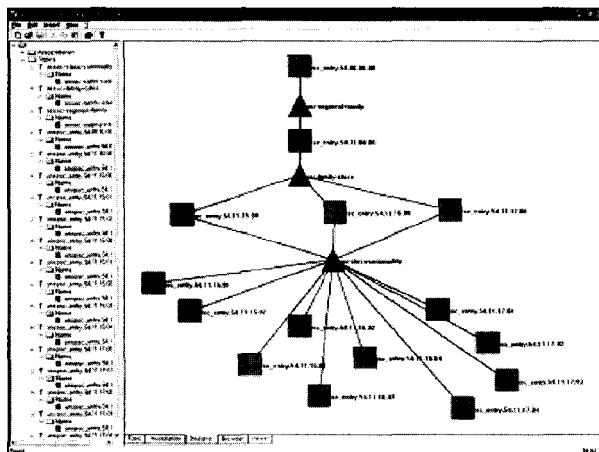
¹ 토픽맵을 개발하고 편집하기 위한 도구에는 tmproc, TM4J, Seman Text, GooseWorks Toolkit, empolisk42, Perl XTM 등의 엔진과 Topic Map Designer, The 'V' Topic Map Browser, TMTab, LTM, AsTMa, mapelizer 등의 에디터(네비게이터 겸용)가 있다.


```

<?xml version = "1.0" encoding = "UTF-8"?>
<topicMap xmlns = "http://www.topicmaps.org/xtm/1.0/"
xmlns:xlink = "http://www.w3.org/1999/xlink">
  <topic id = "entry">
    <baseName>
      <baseNameString>HS Entry</baseNameString>
    </baseName>
  </topic>
  <topic id = "version">
    <baseName>
      <baseNameString>HS Version</baseNameString>
    </baseName>
  </topic>
  ...
  <topic id = "assoc-section-chapter">
    <baseName>
      <baseNameString>HS Section Consists Of Chapters</baseNameString>
    </baseName>
    <baseName>
      <scope>
        <topicRef xlink:href = "#section"/>
      </scope>
      <baseNameString>Chapters</baseNameString>
    </baseName>
    <baseName>
      <scope>
        <topicRef xlink:href = "#chapter"/>
      </scope>
      <baseNameString>Belongs To Section</baseNameString>
    </baseName>
  </topic>
  ...

```

<예제 2>



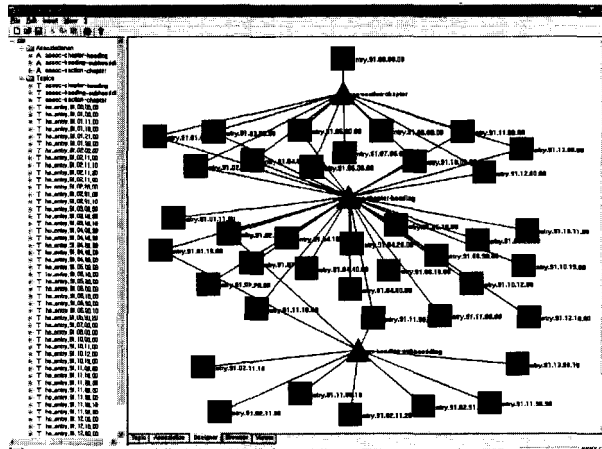
<그림 4> UNSPSC 토픽맵 구조 화면

```

<?xml version = "1.0" encoding = "UTF-8"?>
<topicMap xmlns = "http://www.topicmaps.org/xtm/1.0/"
xmlns:xlink = "http://www.w3.org/1999/xlink">
  <topic id = "entry">
    <baseName>
      <baseNameString>HS Entry</baseNameString>
    </baseName>
  </topic>
  <topic id = "version">
    <baseName>
      <baseNameString>HS Version</baseNameString>
    </baseName>
  </topic>
  ...
  <topic id = "assoc-section-chapter">
    <baseName>
      <baseNameString>HS Section Consists Of Chapters</baseNameString>
    </baseName>
    <baseName>
      <scope>
        <topicRef xlink:href = "#section"/>
      </scope>
      <baseNameString>Chapters</baseNameString>
    </baseName>
    <baseName>
      <scope>
        <topicRef xlink:href = "#chapter"/>
      </scope>
      <baseNameString>Belongs To Section</baseNameString>
    </baseName>
  </topic>
  ...

```

<예제 3>



<그림 5> HS 토픽맵 구조 화면

5. 프로토타입 구현

모델링에 기반하여 개발한 UNSPSC와 HS의 토픽맵을 통합하여 서로 상이한 상품분류체계를 하나로 통합하였다. 전체적인 구조는 분류가 비교적 단순하고 상품 각각이 아닌 상품군을 분류하는 UNSPSC를 기본 골격으로 하고 HS는 UNSPSC가 다루지 못하는 부분을 보완하거나 세분류를 추가하는 형태가 되었다.

일반적으로 토픽맵을 통합하는 방법은 이름을 기반으로 하는 통합(name-based merge)과 주제를 기반으로 하는 통합(subject-based merge), 두 가지이다. 이름을 기반으로 하는 통합은 서로 동일한 <baseName>을 가지는 경우 둘 이상의 토픽을 하나로 통합하는 방식이고 주제를 기반으로 하는 통합은 서로 동일한 <subjectIndicator>를 가지는 경우 둘 이상의 토픽을 하나로 통합하는 방식이다. 그러나 대부분의 주제를 표현할 때 다양한 이름이 존재하므로 이름을 기반으로 통합하는 방식은 자칫 오류를 범하기 쉽다. 따라서 정확한 통합을 위해서 주제를 기반으로 하는 통합 방식을 선택하였다.

통합의 기반이 되는 각 주제는 앞서 설계한 온톨로지 파일이다. 상품분류체계의 각 단계별 그 명칭과 온톨로지 파일을 비교한 후 단계명이 나타내는 주제에 해당하는 온톨로지 파일의 태그를 참조하도록 하여 서로 다른 상품분류체계를 하나의 토픽맵으로 설계하였다. 그 예는 <예제 4>와 같다.

상이한 상품분류체계에 대한 토픽맵을 하나로 표현할 때 각 상품분류체계의 분류단계를 나타내는 section 또는 segment 와 같은 분류단계명은 큰 의미를 가지지 않으므로 각 단계에 대한 명칭을 통일하여 "level"로 표현하였다. 두 상품분류체계 모두 4단계 구조로 이를 통합하는 과정에서 통합분류체계를 4단계로 나타내고자 하였으나 이는 각 상품분류체계를 무너뜨리는 결과를 가져오게 되었다. 결과적으로 7단계로 계층구조를 구성하였고 융통성 있는 구조를 갖도록 하기위해 단계에는 제한을 두지 않는 것을 원칙으로 하였다. 최상위 단계를 level1, 다음 단계를 level2 등 각 단계를 순차적으로 표현하고 총 7단계로 나타내었으며 최하위를 level7로 표시하였다. 프로토타입 구현을 위해 운영체제는 Windows 2000 server, 프로그래밍언어는 XML과 JSP.

```

<topic id="hs_entry.9L13.20.00">
  <instanceOf>
    <topicRef xlink:href="#level5"/>
  </instanceOf>
  <subjectIdentity>
    <subjectIndicatorRef xlink:href="http://dllab.sookmyung.ac.kr/ontology.xml#ha_no"/>
  </subjectIdentity>
  <baseName>
    <baseNameString>비금속제의 것</baseNameString>
  </baseName>
</topic>

```

<예제 4>

Servlet/JSP engine으로는 Tomcat 4.1이 사용되었으며 브라우저로는 Topicmap Designer v 1.1이 사용되었다.

위에 언급한 과정을 통해 UNSPSC토픽맵과 HS토픽맵을 통합하여 프로토타입을 구현하였다. 각 단계를 토픽으로 정의하고 그들의

```

<?xml version = "1.0" encoding = "UTF-8"?>
<topicMap xmlns = "http://www.topicmaps.org/xtm/1.0/" xmlns:xlink = "http://www.w3.org/1999/xlink">
...
  <topic id = "level1">
    <baseName>
      <baseNameString>Level 1</baseNameString>
    </baseName>
  </topic>
...
  <topic id = "assoc-level1-level2">
    <baseName>
      <baseNameString>Level 1 Consists Of Level 2</baseNameString>
    </baseName>
    <baseName>
      <scope>
        <topicRef xlink:href = "# level1"/>
      </scope>
      <baseNameString>Level 2 </baseNameString>
    </baseName>
    <baseName>
      <scope>
        <topicRef xlink:href = "# level2"/>
      </scope>
      <baseNameString>Belongs To Level 1</baseNameString>
    </baseName>
  </topic>
  <topic id = "unspsc_entry.54.00.00.00">
    <instanceOf>
      <topicRef xlink:href = "# level1"/>
    </instanceOf>
    <subjectIdentity>
      <subjectIndicatorRef xlink:href = "http://dlla.sookmyung.ac.kr/ontology.xml# clock"/>
    </subjectIdentity>
    <baseName>
      <baseNameString>시계 및 보석</baseNameString>
      <variant>
        <variantName>
          <resourceData>hs_entry.91.00.00.00 시계 및 그 부속품</resourceData>
        </variantName>
        <parameters>
          <topicRef xlink:href = "# hs_entry.91.00.00.00"/>
        </parameters>
      </variant>
    </baseName>
  </topic>
...

```

<예제 5>

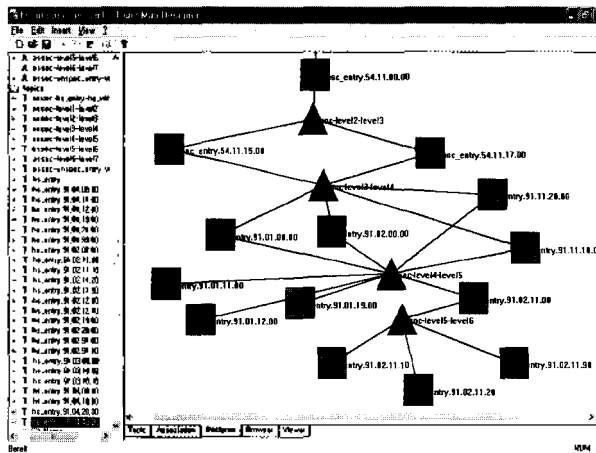
관계를 어소시에이션으로 정의하였다. 동일한 주제를 갖는 서로 다른 토픽의 경우, UNSPSC를 ID로 선언하고 HS는 variant로 선언하여 해당 HS의 코드를 parameter로 나타내었다. <예제 5>는 통합 토픽맵의 일부이다.

<그림 6>은 Topic Map Designer를 이용하여 통합 토픽맵의 구조 중 일부를 나타낸 화면으로 서로 다른 상품분류체계를 분석하여 이를 토픽맵으로 구현한 뒤 두 토픽맵을 통합하여 얻어진 결과라 할 수 있다.

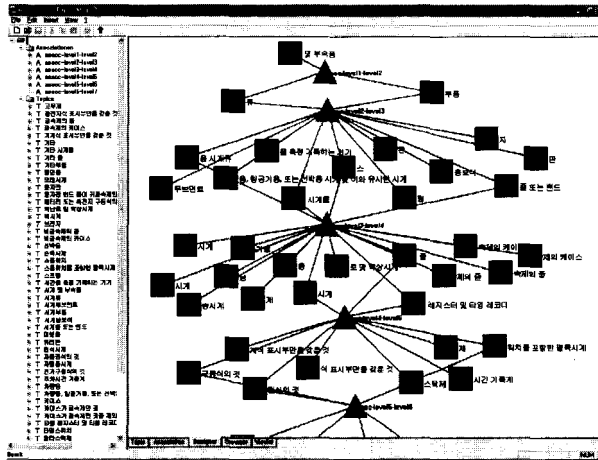
<그림 7>은 토픽맵을 통해 UNSPSC와 HS라는 이 기종 상품분류체계 간의 온톨로지를 통합한 결과이다. 시계 및 부속품이 가장 상위 단계이고 다음 시계류와 시계부품이 두번째 단계를 구성한다. 시계류는 휴대용 시계, 시간을 측정 기록하는 기기, 시계무브먼트 기타 시계류 등으로 분류되고 시계부품은 케이스, 문자판, 브리지, 시계줄 등으로 분류되어 세번째 단계를 구성한다. 네번째 단계는 세번째 단계를 세분류한 항목들로 구성되고, 네번째

단계에서 다시 세분류되는 항목이 있는 경우 이들이 다섯번째 단계를 구성한다. 이러한 방법으로 최종적으로 7단계를 구성하였다.

상품분류체계에 관계없이 변환 또는 검색을 가능하게 하는 온톨로지 통합을 이루었다. 구조가 상이한 두 개의 상품분류체계를 통합하여 각 상품분류체계의 구조를 무너뜨리거나 정보를 손상시키지 않아 체계를 그대로 유지하면서 서로 정보를 자유롭게 교환할 수 있도록 하였다. 이는 기존의 이 기종 상품분류체계 간 정보교환에 있어 테이블 매핑과 같은 코드를 기반으로 하는 방식과는 전혀 다른 새로운 방식이며 샘플링한 UNSPSC와 HS외에 기타 상품분류체계에도 적용이 가능하여 전자상거래에 있어 정보의 손실이나 중복 없이 상품정보를 교환할 수 있는 기반을 제공한다고 할 수 있다.



<그림 6> 통합 토픽맵 구조 화면



〈그림 7〉 토픽맵 통합을 통한 온톨로지 통합 화면

6. 결론 및 제언

본 연구는 토픽맵을 활용하여 이 기존 상품 분류체계간의 온톨로지를 통합한 것으로 상품분류체계를 수집하여 선정된 뒤, 각 상품분류체계의 분류단계 및 구조, 특징을 분석하여 용어와 계층구조, 토픽들 간의 연관관계를 정의하고 이들을 바탕으로 모델링하여 각각에 대한 토픽맵을 개발하고 최종적으로 이들을 통합하는 과정을 거쳤다.

토픽맵 개발에 있어 토픽을 정의하는 작업 외에 토픽간의 연관관계를 나타내는 어소시에이션을 정의하는 것이 중요하다. 구조화되어 있지 않은 정보를 구조화하는데 가장 큰 역할을 하는 것이 어소시에이션이기 때문이다. 본 연구에서는 상품분류체계에 대하여 별도의 어소시에이션을 작성하지 않고 분류단계를 그대로 어소시에이션으로 적용하였기 때문에 분류기준이 판이해 구조가 완전히 다른

상품분류체계를 통합하기에는 어려움이 따른다. 이를 극복하기 위해서는 상품분류체계의 어소시에이션을 분류단계가 아닌 분류명으로 적용하는 연구가 필요하다.

현재 개발되어 있는 토픽맵 브라우저들은 본 연구에서 개발된 토픽맵을 브라우징 할 때 상품분류체계와 분류명을 동시에 보여줄 수 있는 기능이 미흡하다. 추후 이용자가 토픽맵을 구성하는 항목 중 원하는 항목을 추출하여 보여줄 수 있는 브라우저 개발이 필요하다고 여겨진다.

기존의 상이한 상품분류체계간 통합이나 변환은 매핑테이블을 이용하여 코드를 기반으로 하는 방식을 사용하여 왔다. 1:1 테이블 매핑을 통해 두 상품분류체계간의 정보교환을 이루는 방식은 하나의 상품분류체계가 수정 또는 삭제되는 등의 변경이 이루어지는 경우, 매번 테이블을 수정해야 하는 불편이 있었다. 또한 분류기준이 상이하어 그 구성이나 구조가 다른 상품분류체계간에 하나의 상품

에 대응하는 다른 상품을 찾는 것 자체가 어려운 경우도 빈번히 발생하였다.

온톨로지를 통합하여 상품분류체계의 정보 교환을 하는 방식은 모든 전자상거래 이용자들이 각기 다르게 표현되는 상품분류체계에 대하여 상품정보를 교환할 때 또는 검색을 할 때 쉽고 빠르게, 그리고 정확하게 상품 정보를 교환할 수 있도록 지원한다. 이러한 방식은 기존의 테이블 매핑 방식과 비교하면 다음과 같은 장점이 있다.

첫째, 각기 다른 상품분류체계를 상호 유동적으로 쉽게 변환하고 검색할 수 있도록 하는 기능을 지원한다. 또한 주기적으로 변화하는 각 분류체계의 형태에도 쉽게 대응할 수 있다.

둘째, 각 기업이나 기관의 독자적인 상품분류체계를 기타 상품분류체계와 통합할 수 있도록 함으로써 상품 검색에서 거래에 이르는 과정 상 이용자가 가장 적합한 의사 결정을 할 수 있도록 지원한다.

셋째, 동일상품에 대한 표현요소 및 표현수준이 상이한 상품분류체계간의 상호 운용성을 해결하기 위해서 투자하는 많은 시간과 비용을 절감할 수 있다.

본 연구에서 제시한 모델링 및 온톨로지 통합 방식은 지식경영 및 지식관리 시스템에도 활용될 수 있다. 지식 관리 시스템은 복잡한 형태의 데이터의 표현 및 처리가 요구되고 기존의 방식으로는 이러한 요구를 수용하기가 어려우며 일반적으로 트리 형태의 구조가 대부분이나 토픽맵을 활용함으로써 지식의 여러 유형을 구체적으로 표현하고 이를 구조화할 수 있다. 앞서 보여진 것처럼 그래프 기반의 데이터 모델을 브라우저할 수 있으므로 모

든 지식 정보 표현이 가능해지고 다양한 검색을 지원할 수 있으며 디렉토리 검색 및 연관 검색을 동시 지원할 수 있다.

참 고 문 헌

- [1] 전자상거래통합포럼. 2003. 『전자상거래표준화로드맵』.
- [2] 지영수. 2002. 『Rational Rose 2000』, 서울: 홍릉과학출판사.
- [3] Helka Folch. 2000. "Navigable Topic Maps for overlaying multiple acquired semantic classifications". Markup Languages Theory and Practice, 2(3): 269-280
- [4] Mary Nishikawa. 2002. "Organizing Information in a Corporate Intranet." Extreme Markup Languages 2002: Baltimore.
- [5] Rath Hans Holger. 2000. "Topic maps: templates, topology, and type hierarchies." Markup Languages Theory and Practice, 2(1): 137-151
- [6] Steven R. Newcomb. 2002. "The Topic Maps Standard."
 <www.infoloom.com/tmstands.htm>

저 자 소 개



고세영 (E-mail : ko3030@empal.com)

1998. 숙명여자대학교 문헌정보학과(학사)

2003. 숙명여자대학교 대학원 정보학(석사)

2003. ~ 현재 (주)인포와이즈

관심 분야 : 전자상거래, XML Topicmap, 온톨로지, 표준화



김성혁 (E-mail : ksh@sookmyung.ac.kr)

1975. 연세대학교 문헌정보학과(학사)

1983. 미국 Case Western Reserve University(정보학석사)

1992. 연세대학교 대학원 문헌정보학과(박사)

1986 ~ 1991. 대신증권 전산실장

1998 ~ 1999. 첨단학술정보센터 소장 직무대행

1991 ~ 현재 숙명여자대학교 정보과학부 교수

2001 ~ 현재 EBK 전자책문서표준화위원회 위원

2003 ~ 현재 한국전자거래학회 편집이사

관심 분야 : 디지털도서관, 시맨틱웹, 온톨로지, 전자카더로그