

K-Box: 토픽맵 기반의 온톨로지 관리 시스템

(K-Box: Ontology Management System based on Topic Maps)

김정민[†] 박철만[†] 정준원^{**}
(Jungmin Kim) (Chulman Pak) (Junwon Jung)

이한준[†] 민경섭[†] 김형주^{***}
(Hanjun Lee) (Kyoungsub Min) (Hyoung-Joo Kim)

요약 시맨틱 웹은 지능적인 서비스를 제공하기 위한 새로운 웹 환경이다. 시맨틱 웹을 구현하기 위해 많은 새로운 노력들이 진행되어 왔다. 이 중에서 가장 기본적이고 중요한 것은 의미 정보를 표현, 저장, 질의, 관리해 주는 온톨로지 관리 시스템이다. 우리는 이와 같은 기능을 제공하기 위한 새로운 시스템으로서, 토픽맵 모델을 기반으로 한 효율적인 온톨로지 관리 시스템인 K-Box를 구현하였다. K-Box는 온톨로지 관리를 위한 기본적인 기능들을 제공하며, 이질적인 저장소들을 일관된 인터페이스로 접근할 수 있도록 함으로써 저장 장치 독립성을 제공하였다. 또한, 저장 관리되는 모든 온톨로지들의 무결성을 보장하기 위한 새로운 기법과 사용자 관심을 중심으로 한 온톨로지 검색 지원을 위한 방법을 제안하였다.

마지막으로, 우리는 여러 온톨로지들을 적용해 봄으로써 K-Box 시스템이 효율적으로 사용 가능함을 확인하였다.

키워드 : 시맨틱 웹, 온톨로지, 토픽맵, 온톨로지 관리 시스템

Abstract The Semantic Web introduces the next generation of the Web by establishing a semantic layer of machine-understandable data to enable machines(i.e intelligent agents) retrieve more relevant information and execute automated web services using semantic information. Ontology-related technologies are very important to evolve the World Wide Web of today into the Semantic Web in representation and share of semantic data.

In this paper, we proposed and implemented the efficient ontology management system, K-Box, which constructs and manages ontologies using topic maps. We can use K-Box system to construct, store and retrieve ontologies.

K-Box system has several components: Topicmap Factory, Topicmap Provider, Topicmap Query Processor, Topicmap Object Wrapper, Topicmap Cache Manager, Topicmap Storage Wrapper.

Key words : Semantic Web, Ontology, Topic Map, Ontology Management System

1. 서론

웹에는 전 세계의 기업, 단체 및 개인들에 의해 하루에도 수많은 정보들이 생성되고 있으며 이러한 정보들

을하고자 하는 사용자는 단순히 정보가 담겨있는 웹사이트의 URL 주소만 알면 언제든지 필요한 정보에 접근할 수 있다.

그러나 웹에 담겨있는 정보의 양이 대용량일 뿐만 아니라 그 양이 급속도로 증가하고 있기 때문에 사용자가 원하는 정보를 정확히 찾아내기란 쉬운 일이 아니다. 예를 들어 Tim Berners Lee가 저자인 문서를 찾기 위해 검색엔진을 이용하는 경우 Tim Berners Lee의 각 단어가 들어있는 수많은 웹 페이지들의 리스트를 넘겨받게 된다. 검색엔진은 단순히 검색 키워드가 들어있는 웹 페이지들을 찾아서 수집해주는 역할만 하기 때문에 이들 웹 페이지들의 내용을 살펴보면 원하는 정보를 추출해 내는 것은 사람이 해야 한다. 이는 현재의 웹 구조

· 본 연구는 BK21과 G7 프로젝트에서 지원받았음

† 비회원 : 서울대학교 전기컴퓨터공학부
jmkim@oopsla.snu.ac.kr
cmpark@oopsla.snu.ac.kr
leehj@oopsla.snu.ac.kr
ksmin@oopsla.snu.ac.kr

** 학생회원 : 서울대학교 전기컴퓨터공학부
jwjung@oopsla.snu.ac.kr

*** 종신회원 : 서울대학교 전기컴퓨터공학부 교수
hjk@oopsla.snu.ac.kr

논문접수 : 2003년 2월 19일

심사완료 : 2003년 9월 19일

가 사람이 읽고 이해할 수 있도록 되어 있기 때문이다.

이러한 문제를 해결하기 위해서는 웹 콘텐츠에 담겨 있는 정보의 내용 - 시맨틱(semantic) 또는 지식(knowledge) - 을 사람이 이해하듯이 기계가 웹 콘텐츠의 시맨틱을 이해하고 사용자의 필요에 맞는 정보를 제공하거나 웹서비스를 제공하도록 웹의 구조를 확장하는 것이다. 이러한 확장된 개념의 웹 구조를 시맨틱 웹(Semantic Web)이라고 한다[1,2].

시맨틱 웹을 구축하기 위해서는 먼저 웹 콘텐츠에 들어있는 시맨틱을 사람이나 기계들 사이에 공유할 수 있도록 상호간에 이해 가능한 용어들로 시맨틱을 정의하고 이를 표준화된 모델과 형식화된 언어로 표현해야 한다. 시맨틱을 정의하는 방법에는 인공지능 분야에서 지식을 정의하는 방법으로 제안한 온톨로지(ontologies)가 있다. 온톨로지는 특정 도메인의 개념 및 지식을 명세화하기 위해 그 지식을 설명하는 표준 용어들을 정의하고 용어들 사이의 계층(taxonomy) 및 연관(relation) 관계를 정의하는 것이다. 같은 지식 도메인의 서로 다른 웹 사이트들에서 같은 온톨로지를 공유할 경우 에이전트는 이들 사이트들에 들어있는 정보를 관련있는 토픽(topic)끼리 연결하여 동시에 검색할 수 있으며 각 사이트에 존재하는 에이전트들 사이에 온톨로지에 기반하여 상호 메시지를 주고받음으로써 자동화된 웹서비스를 제공할 수 있다[3].

시맨틱 웹에서 온톨로지를 표현하고 엔코딩하기 위해 사용되는 데이터 모델로 RDF(Resource Description Framework)[4], RDF Schema[4], DAML+OIL[5] 그리고 Topic Map[6]이 있다. 본 논문에서는 토픽맵을 기반으로 온톨로지를 생성, 변경, 저장 및 검색하는 온톨로지 관리 시스템(K-Box)을 제안한다. 토픽맵은 주제중심으로 개념을 명세화하고 개념들간의 연관관계를 정의하는 모델로서 ISO(International Organization for Standardization)에서 표준으로 제안하고 있다. 현재의 웹이 시맨틱 웹으로 발전되어가기 위해서는 지식을 담고 있는 각 사이트들이 온톨로지에 기반한 웹포털로 진화해야 한다. 이를 위해 K-Box 시스템은 시맨틱 웹 포털의 기반 시스템 역할을 한다.

K-Box 시스템은 온톨로지의 데이터모델로 토픽맵을 사용하기 때문에 토픽맵을 생성하고 변경, 저장하며 키워드 검색 및 주제 검색 등을 지원하기 위한 여러 컴포넌트들로 구성된다. K-Box의 컴포넌트들은 토픽맵 객체들의 단일 인터페이스 제공을 위한 토픽맵 오브젝트 래퍼(Topic Map Object Wrappers), 토픽맵 객체들을 저장하기 위한 스토리지 래퍼(Storage Wrappers), 사용자가 토픽맵에 접근할 수 있게 하는 토픽맵 제공자(Topic Map Provider), 대용량의 토픽맵의 효율적인

검색을 위한 토픽맵 캐쉬 관리자(Topic Map Cache Manager), 토픽맵 객체들을 생성하는 토픽맵 생성자(Topic Map Factory), 토픽맵 관리를 위한 토픽맵 관리자(Topic Map Manager), 토픽맵의 가져오기 및 내보내기등 여러 유용한 기능들을 제공하는 토픽맵 도구(Topic Map Utilities) 등이 있다.

K-Box 시스템은 클라이언트-서버 구조로 사용자의 토픽맵 접근을 지원하는 프론트-엔드(front-end)와 토픽맵 엔진 역할의 백-엔드(back-end)로 나누어지며 K-Box API를 제공함으로써 지식 관리 시스템, 콘텐츠 관리 시스템 등의 내부 온톨로지 관리 모듈로 활용될 수 있다.

온톨로지의 무결성을 지원하기 위해 토픽맵 생성자 내부에 토픽맵 파서를 둬으로써 부적합한 토픽들이나 토픽들간의 연관관계가 생성되지 않도록 보장한다.

생성된 토픽맵은 관계형 데이터베이스, 객체지향 데이터베이스, 자바 객체 저장소, XML 저장소 등에 저장할 수 있다. K-Box 시스템은 저장장치 독립성을 제공하기 위해 스토리지 래퍼를 두고 있다.

K-Box 시스템에서의 토픽맵 검색은 토픽 중심의 키워드 검색과 주제 중심의 디렉토리 검색 그리고 그래프 구조 중심의 비주얼 네비게이션 검색으로 나눌 수 있다. 그리고 사용자 프로파일을 둬으로써 사용자가 K-Box에 로그인했을 때 그 사용자의 관심 토픽들과 그 주위의 연관된 토픽들을 자동으로 검색해서 보여주도록 사용자 중심의 검색을 지원한다.

2장의 관련연구에서는 기존의 토픽맵 관리 시스템과 온톨로지 관리 시스템들에 대해 살펴본다. 그리고 3장에서는 토픽맵 데이터 모델을 이용하여 온톨로지를 정의하고 표현하는 과정에 대해 설명하고 있으며 4장에서는 K-Box 시스템의 아키텍처와 각 컴포넌트들에 대해 자세한 구현 내용을 설명한다. 5장에서는 시스템의 구현 결과를 보여주고 있으며 6장에서 결론 및 향후연구에 대해 소개한다.

2. 관련연구

W3C와 ISO에서는 온톨로지를 표현하는 기본 모델로 각기 RDF와 토픽맵을 표준으로 제안하고 있다. 이 두 모델은 서로 경쟁적인 입장이기보다는 상당히 많은 유사성으로 인해 통합 및 상호 보완 관계를 가진다. 즉, RDF로 생성된 온톨로지를 토픽맵 모델로 변환이 가능하고 반대로 토픽맵으로 생성된 온톨로지 모델을 RDF 모델로 변환이 가능하다. 주요 차이점은 RDF가 자원(resource)중심인데 반해 토픽맵은 주제(subject)중심이다. RDF는 URL로 접근 가능한 자원들에 대한 메타데이터를 생성하기 위한 모델을 제시하고 토픽맵은 개념

(concept)이나 사물에 대하여 정형화된 명세를 생성하기 위한 모델을 제시한다. 온톨로지는 개념의 명세화이므로 실제 자원이 없더라도 지식 구조를 표현하기 위해 사용할 수 있다. 이런 관점에서 볼 때 RDF보다는 토픽맵이 온톨로지 모델로서 더 적합하다고 볼 수 있다[7].

토픽맵 관리 시스템으로 상용화된 시스템은 Ontopia에서 개발한 Ontopia Knowledge Suite(OKS)[8]와 Mondeca에서 개발한 Mondeca Intelligent Topic Manager(ITM)가 있다. 이 시스템들은 토픽맵을 생성하고 관계형 데이터베이스에 저장하며 편집, 검색, 네비게이션 기능을 제공하는 것으로 지식 관리 시스템 또는 콘텐츠 관리 시스템의 엔진으로 사용된다. 본 논문의 온톨로지 관리 시스템은 이들 시스템과 유사한 기능을 가지고 있을 뿐만 아니라 관계형 데이터베이스, 객체지향 데이터베이스, XML 저장소 등 다양한 스토리지를 지원하는 스토리지 래퍼와 토픽맵 캐싱, 그리고 토픽맵 제약 조건을 정의하는 토픽맵 스키마 지원 등의 향상된 기능을 가지고 있다.

OntoEdit[9]는 온톨로지 공학 측면에서 접근한 시스템으로 온톨로지 구축 방법론과 그 방법론에 따라 온톨로지를 구축해 가는 과정을 시스템으로 구현해 놓았다. 그리고 여러 온톨로지 설계자들이 협력하여 온톨로지를 구축할 수 있도록 하였다. 이 시스템은 온톨로지를 서비스하기 이전에 각 사이트에서 온톨로지를 설계하고 생성하고자 할 경우에 적용 가능한 시스템이다. Karlsruhe Ontology(KAON) tool suite[10]는 온톨로지를 저장하고 검색하기 위해 하위 수준의 KAON-API와 상위 수준의 KAON-API를 제공한다. KAON의 프로그래밍 인터페이스를 이용하여 온톨로지를 관리하거나 검색하는 상위 어플리케이션을 개발할 수 있다. KAON-API는 RDF-API로 직접적으로 매핑이 될 수 있으며 RDF Schema, DAML+OIL로 표현된 온톨로지를 지원하며 하위 수준의 API에서는 여러 종류의 스토리지에 구현되도록 인터페이스를 제공한다. 그러나 KAON의 검색은 온톨로지 전체를 브라우징하는 정도의 간단한 질의를 처리하는 수준이다.

3. 온톨로지 데이터 모델

K-Box 시스템에서는 온톨로지 데이터 모델로 토픽맵을 사용한다. 토픽맵은 ISO에서 표준으로 제정한 것으로 초기에는 전자 색인을 위한 데이터 모델로 고안되었으나 현재에는 지식 관리 시스템의 지식맵, 콘텐츠 관리 시스템의 콘텐츠 맵 그리고 시맨틱 웹의 온톨로지 등의 데이터 모델로 사용되고 있다. 토픽맵 모델의 핵심 요소는 토픽(Topic), 어커런스(Occurrence), 어소시에이션(Association)으로 볼 수 있다. 토픽은 객체지향 모델에

서의 클래스나 객체에 해당하는 것으로 표현하고자 하는 대상을 가리키고 어커런스는 토픽에 종속되는 개념으로 해당 토픽에 대한 실제 내용이 담겨있는 자원의 주소(URI로 표현된다)나 지식 데이터 자체를 가리킨다. 그리고 토픽들 사이에는 토픽타입과 토픽간의 상하관계와 토픽들간의 연관성을 표현하는 어소시에이션이 있다 [1,11].

온톨로지를 구축하기 위해서는 그 분야의 전문지식을 가지고 있는 전문가에 의해 개념을 설명할 수 있는 용어를 정의하고 용어들 사이의 관계를 정의한다. 용어들 사이의 관계는 상위 개념과 하위 개념 및 개념과 인스턴스간의 관계인 분류(taxonomy)와 개념간의 연관성을 표현하는 연관관계(relation)으로 나뉘어진다.

온톨로지가 정의되면 용어를 토픽으로, 분류를 토픽타입과 토픽간의 관계로, 그리고 연관관계를 토픽들 사이의 어소시에이션으로 매핑하여 초기 토픽맵을 생성한다. 그림 1에서는 온톨로지 정의 및 토픽맵 데이터 모델을 생성하는 과정에 대해 보여주고 있다[3].

토픽맵을 표현하는 기술 언어로 XTM(XML Topic Maps)[12]이 있다. XTM은 토픽맵 모델의 각 요소를 나타내는 태그 집합과 엘리먼트들간의 구조를 정의하고 있으며 토픽맵을 기술하는 표준 포맷으로 사용되고 있다. 그림 2에서는 실험자료로 사용하는 서울대학교 객체지향 시스템 연구실의 웹사이트 온톨로지를 XTM으로 기술한 문서의 일부분을 보여주고 있다. 여기서 oopsla, member, paper는 클래스에 해당하는 토픽 타입이고 HJKIM은 객체에 해당하는 member의 인스턴스이다.

4. K-Box의 설계와 구현

4.1 시스템 구조

K-Box 시스템은 토픽맵 데이터를 저장하고 관리하는 역할을 수행하기 위해 그림 3에서 볼 수 있듯이 토픽맵 객체들의 단일 인터페이스 제공을 위한 토픽맵 오브젝트 래퍼(Topic Map Object Wrappers), 토픽맵 객체들을 저장하기 위한 스토리지 래퍼(Storage Wrappers), 사용자가 요청하는 토픽맵 객체를 전달하는 토픽맵 제공자(Topic Map Provider), 대용량의 토픽맵의 효율적인 검색을 위한 토픽맵 캐쉬 관리자(Topic Map Cache Manager), 토픽맵 객체들을 생성하는 토픽맵 생성자(Topic Map Factory), 토픽맵 관리를 위해 일관된 인터페이스를 제공하는 토픽맵 관리자(Topic Map Manager), 토픽맵 가져오기(import), 내보내기(export) 및 여러 기능들을 제공하는 토픽맵 도구(Topic Map Utilities)들로 구성된다.

토픽맵 오브젝트 래퍼는 토픽맵을 구성하고 있는 오브젝트들의 구현에 비종속적으로 일관된 인터페이스를

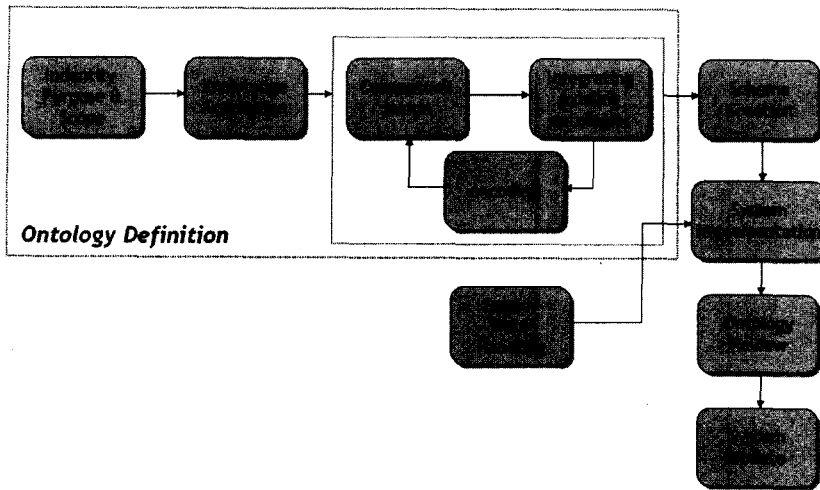


그림 1 온톨로지 정의 및 데이터 모델 생성 절차

```

<?xml version="1.0" encoding="UTF-8"?>
<topicMap id="oopsla-tm" xmlns="http://www.topicmaps.org/xtm/1.0/"
xmlns:xlink="http://www.w3.org/1999/xlink">
<!-- Topic map for OOPSLA homepage. (c) 2001 OOPSLA, SNU -->
<topic id="oopsla">
<subjectIdentity><subjectIndicatorRef xlink:href="#oopsla-tm"/></subjectIdentity>
<baseName>
<baseNameString>OOPSLA Lab topic map</baseNameString>
</baseName>
<occurrence><resourceRef xlink:href="http://oopsla.snu.ac.kr"/></occurrence>
</topic>
<topic id="member">
<baseName> <baseNameString>OOPSLA Member</baseNameString> </baseName>
</topic>
<topic id="paper">
<baseName> <baseNameString>OOPSLA Paper</baseNameString> </baseName>
</topic>
<topic id="HJKIM">
<instanceOf><topicRef xlink:href="#member"/></instanceOf>
<baseName> <baseNameString>김형주</baseNameString> </baseName>
<occurrence><resourceRef http://oopsla.snu.ac.kr/~hjk/></occurrence>
</topic>

```

그림 2 서울대학교 객체지향시스템 연구실 온톨로지를 표현하는 XTM 문서의 일부분

제공해 주는 역할을 한다. 토픽맵 오브젝트 래퍼는 자바의 인터페이스로 정의가 되며, 각각의 오브젝트들의 메소드들을 정의하고 있다.

토픽맵 스토리지 래퍼는 토픽맵 오브젝트 래퍼를 상속받아 각 스토리지에 맞게 구현된 클래스들을 의미한다. 토픽맵 스토리지 래퍼는 토픽맵의 각 객체들을 실제의 스토리지에 영구적으로 저장하고 로드하는 등의 작업을 수행한다. 토픽맵 스토리지 래퍼는 토픽맵 오브젝트 래퍼를 통해서 하위의 스토리지에 관계없이 일관적

인 인터페이스를 제공해 준다.

토픽맵 생성자는 토픽맵의 각 구성요소들을 생성하는 역할을 수행한다. 토픽맵 생성자는 토픽맵 스토리지 래퍼별로 독립적으로 제공되며, 토픽맵 오브젝트 래퍼의 생성자 인터페이스를 통해 일관적인 인터페이스를 제공한다.

토픽맵 제공자는 스토리지에 저장된 토픽맵을 메모리로 로드하여 사용자에게 제공하는 역할을 한다. 토픽맵 제공자도 역시 토픽맵 스토리지 래퍼별로 독립적으로

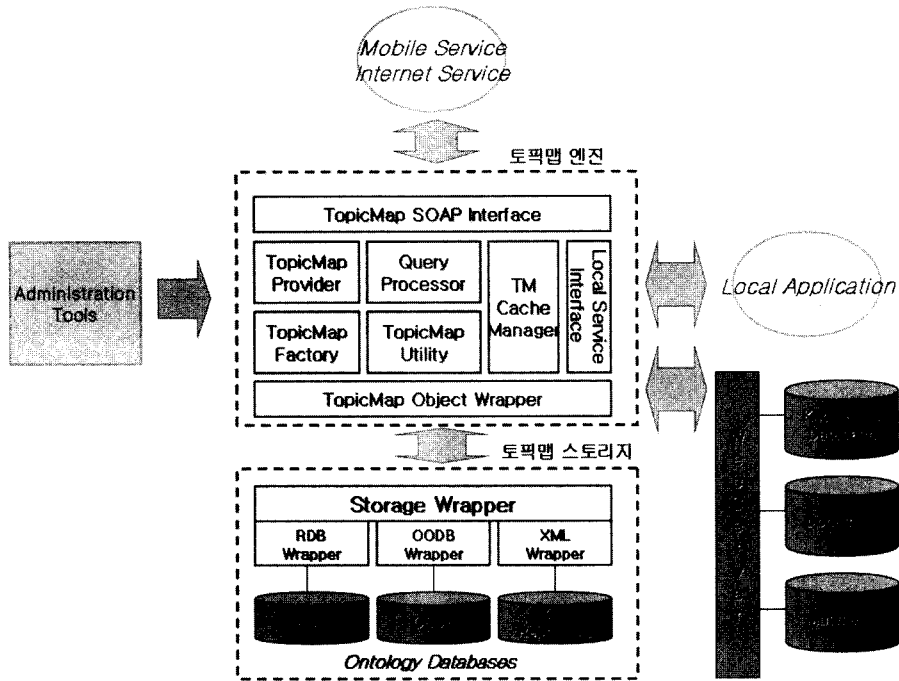


그림 3 K-Box 시스템의 토픽맵 기반 온톨로지 관리를 위한 구조

제공되며, 토픽맵 오브젝트 래퍼의 제공자 인터페이스를 통해 일관적인 인터페이스를 제공할 수 있다.

토픽맵 캐쉬 관리자는 대용량의 토픽맵의 검색 효율을 높이기 위한 것으로 특정 토픽을 검색하면서 그 토픽과 연관관계에 있는 다른 토픽들의 정보도 미리 검색하여 캐쉬에 저장한 다음 사용자의 검색 조건에 부합되는 토픽이 있으면 디스크를 접근하지 않고 캐쉬에서 그 토픽을 제공한다.

토픽맵에 대한 검색은 검색 엔진에서와 마찬가지로 키워드 검색과 디렉토리 검색을 지원하고 비주얼 네비게이션 도구를 이용한 그래프 탐색도 지원한다.

토픽맵 관리자는 토픽맵 컨테이너의 구현에 상관없이 일관된 인터페이스를 사용자에게 제공한다. 토픽맵 관리자는 K-Box에 등록되어 있는 토픽맵과 이 토픽맵이 어떤 토픽맵 스토리지 래퍼에 의해 제공되고 있는지에 대한 정보를 관리한다. 사용자로부터 토픽맵을 접근하고자 하는 요청이 오면, 토픽맵 관리자는 적절한 토픽맵 스토리지 래퍼의 토픽맵 제공자를 생성한다. 토픽맵 제공자는 사용되는 스토리지 또는 구현에 종속적인 부가 정보를 필요로 하는 경우가 있는데, 이를 위해 별도의 설정 파일을 두어 이러한 정보를 제공한다. 토픽맵 관리자에 의해 생성된 토픽맵 제공자는 사용자에게 전달되며, 사용자는 토픽맵 오브젝트 래퍼의 제공자 인터페이스를

통해 토픽맵에 접근할 수 있다. 즉, 토픽맵 관리자를 통해서 토픽맵을 접근하는 사용자는 토픽맵 오브젝트 래퍼의 상위 인터페이스를 통해서 접근할 수 있기 때문에, 토픽맵 스토리지 래퍼의 구현에 독립적으로 일관적인 인터페이스를 사용할 수 있게 된다.

4.2 토픽맵 생성과 온톨로지 무결성

토픽맵 생성자는 XTM 문서를 읽어서 토픽맵을 생성하거나 이미 생성된 토픽맵의 구성 요소 객체를 추가, 삭제, 변경할 수 있는 기능을 제공한다. XTM 문서가 입력으로 들어오면 SAX 파서가 이 문서를 분석하여 구문상의 오류를 체크하고 토픽맵 DTD에 적합한 경우 엘리먼트 정보와 이벤트를 토픽맵 파서(Topic Map Parser) 객체에게 전달한다. 토픽맵 파서는 토픽맵 생성자 객체를 생성하고 이 토픽맵 생성자가 입력으로 들어온 엘리먼트의 종류에 따라 각 구성 요소에 해당하는 구현 객체들을 생성하여 토픽맵 구성 요소들을 생성해 나간다. 그림 4에서는 토픽맵 생성과 관련된 클래스들 사이의 연결 관계를 보여준다.

토픽맵 파서는 온톨로지의 무결성을 보장하기 위해 온톨로지 제약조건을 사전에 검사한다. SAX 파서의 경우 XTM 문서의 구문 오류를 검사하는 역할이지만 토픽맵 파서는 연관관계의 다중성(cardinality), 상위 토픽이 가질 수 있는 하위 토픽의 종류, 어커런스 형

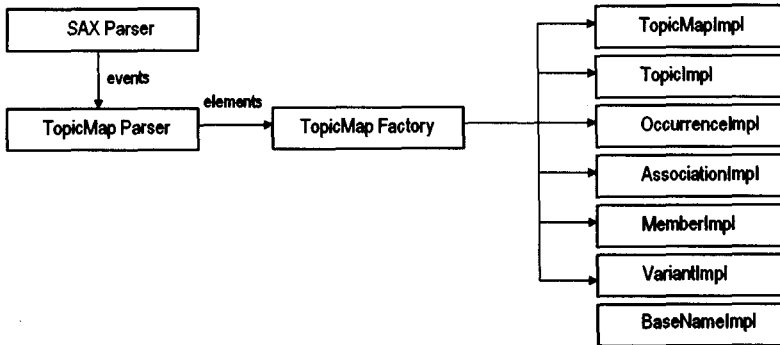


그림 4 토픽맵 생성을 위한 클래스들

표 1 토픽맵 파서에 의해 검사 가능한 제약조건의 종류

제약조건 대상	제약조건 유형	설명
TOPIC	hasSubType	하위 토픽으로 가질 수 있는 토픽
	isOccurrence	어커런스를 가지는지 여부
	isTerminal	하위 토픽을 가질 수 있는지 여부
OCCURRENCE	instanceOf	특정 타입의 어커런스 허용
	min, max	어커런스의 카디널리티
	data, ref, any	어커런스의 종류
ASSOCIATION	min, max	어소시에이션의 카디널리티
	roleTopicRef	어소시에이션의 player type

(Occurrence Type), 연관관계에 있어서 멤버의 역할(role) 등의 제약조건을 검증하기 위해 토픽맵 스키마를 참조한다.

토픽맵 스키마는 제약조건 규칙(rule) 테이블로서 온톨로지를 설계할 때 설계자에 의해 정의된다. 토픽맵의 토픽 객체 집합과 토픽맵 스키마의 규칙 집합을 대응시킴으로써 토픽맵 파서는 특정 토픽에 변경이 발생할 때 그 토픽과 연결된 규칙들을 찾아서 제약조건 검사를 수행한다. 표 1에서는 토픽맵 파서가 검사하는 제약조건의 종류를 보여주고 있다.

4.3 토픽맵 스토리지 독립성

토픽맵 생성자에 의해 생성된 메모리 내의 토픽맵을 저장장치에 저장하기 위해 K-Box 시스템에서는 스토리지 래퍼를 이용하여 토픽맵 데이터의 영속성을 유연하게 지원하고 있다.

스토리지 래퍼는 객체 지향적으로 설계되어, 스토리지 래퍼를 위한 공통의 인터페이스를 가지며, 이 인터페이스에 따라 각 스토리지의 내부 구현은 숨겨지도록 설계되어 있다. 각각의 래퍼들은 토픽맵 데이터의 효율적인 저장을 위해 각각의 스토리지 시스템의 특성에 따라 별도의 최적화 기법을 사용할 수 있다. 본 논문에서 구현

한 K-Box는 관계형 데이터베이스 시스템과 객체지향 데이터베이스 시스템, XML 저장소 방식의 스토리지 래퍼를 두고 있으며 각각의 스토리지를 토픽맵 스토리지로 사용하기 위해 고려해야 할 특성으로는 아래와 같은 것들이 있다.

객체지향 데이터베이스 시스템은 토픽맵 데이터모델을 직접적으로 표현할 수 있는 장점이 있으나 안정적인 객체지향 데이터베이스를 사용하기가 쉽지 않다. K-Box 시스템은 자바로 구현되어 있으므로 자바 객체를 영속적으로 저장해주는 자바 객체 저장소를 객체지향 데이터베이스 대신 사용할 수 있으나 대용량의 처리와 트랜잭션 처리에 부적합한 문제점이 있다.

XML 저장소의 경우, XML로 표현된 토픽맵 데이터(XTM 문서)를 별도의 가공없이 스토리지에서 지원이 가능하다는 장점이 있다. 그러나 XML 저장소가 지원하는 DOM이나 SAX는 실제 토픽맵 데이터 모델과는 그 형태가 다르다. 따라서 DOM 혹은 SAX로 표현된 XML 데이터를 토픽맵 데이터 모델로 변환하는 작업이 필요하다.

관계형 데이터베이스는 실제 가장 많이 쓰이는 저장 도구로써, 성능이나 안정성, 보편성 등이 이미 검증되었다는 장점을 지닌다. 그러나 그래프 데이터 모델에 대한 타 저장 시스템들의 표현력에 비해 관계형 데이터베이스의 표현력은 많이 떨어진다. 그렇기 때문에, 토픽맵 데이터 모델을 위한 관계형 스키마 생성, 토픽맵 데이터의 관계형 데이터로의 변환/역변환 등의 문제가 발생한다. 이 과정에서 관계형 데이터베이스 내에서 조인이 과도하게 발생하게 되고, 이는 성능상의 오버헤드로 나타나게 된다. 이는 토픽맵의 연관 정보 탐색과 같은 토픽간의 연관성 경로를 따라 탐색이 일어나는 경우 많이 발생하게 되며, 이후에 토픽맵 상에서의 추론 등 연관 정보에 대한 탐색이 빈번하게 일어나는 응용에 있어 큰 문제점으로 작용할 수 있다.

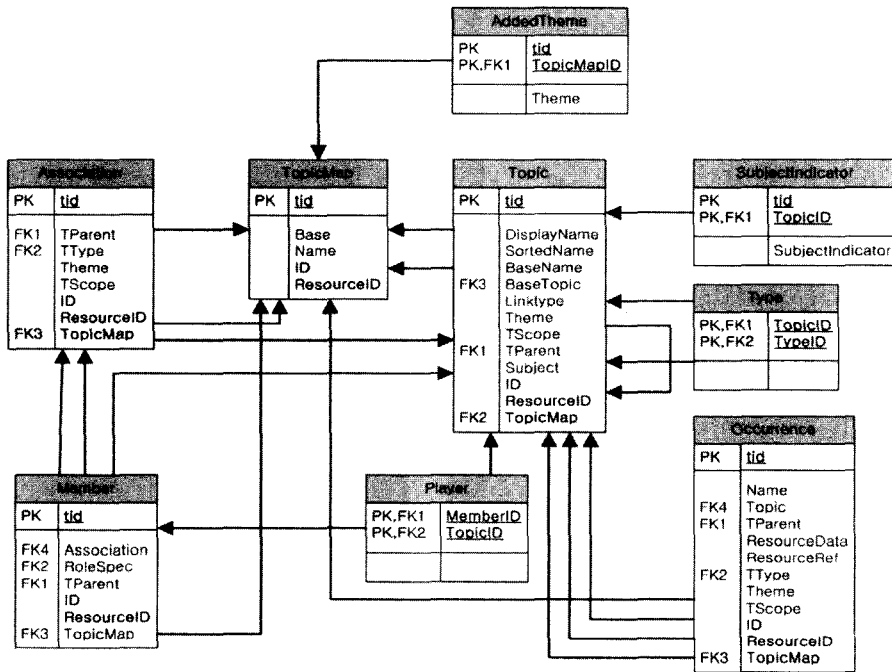


그림 5 토픽맵 저장을 위한 관계형 데이터베이스 스키마

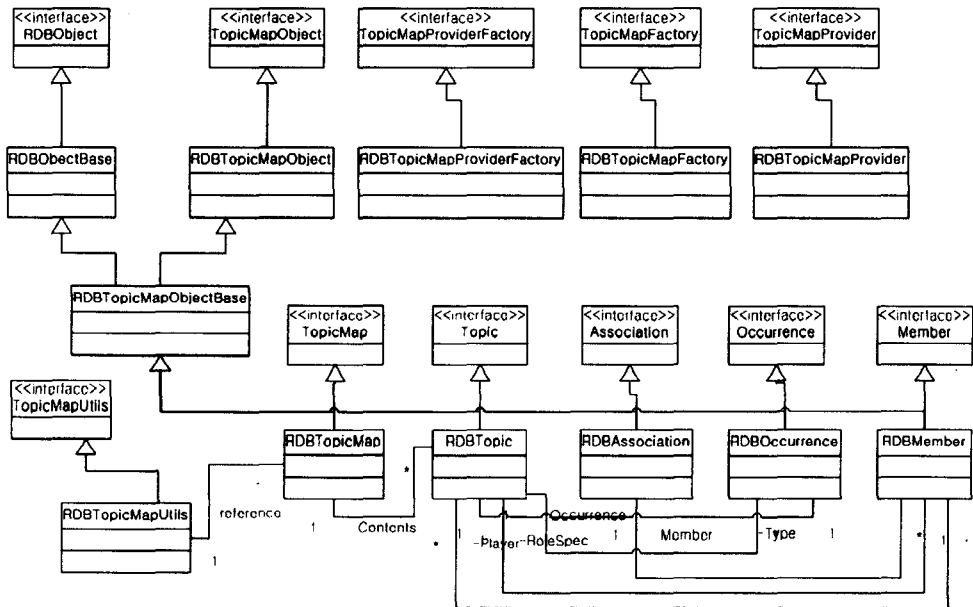


그림 6 관계형 데이터베이스에 저장하기 위한 스토리지 래퍼의 클래스 계층도

이러한 문제는 관계형 데이터베이스를 이용한 XML 저장소에서의 질의 처리 문제와 어느 정도 유사하다. XML 저장소의 질의 처리에 있어 가장 큰 쟁점이 되고

있는 부분은 질의어에 표현된 정규 경로식(regular path expression)을 어떻게 효율적으로 처리할 것인가 하는 점이다. 이를 위해 1-Index, 2-Index, T-Index[13] 및

반구조적 데이터를 위한 인덱스[14,15] 등의 인덱스를 이용하여 검색을 효율적으로 수행하도록 하는 방법과 시그니처 정보를 이용하여 탐색 경로를 줄이는 방법 [16,17]과 같은 질의 처리를 개선시키는 방법 등이 제안되었다. 이러한 질의 처리 기법들은 모두 XML을 대상으로 제안되었으나 토픽맵 시스템의 연관 정보 탐색 문제에도 도입해 볼 수 있을 것이다.

K-Box 시스템에서는 관계형 데이터베이스를 주 스토리지로 선택하였으며 구현 및 실험도 오라클 8i에서 수행하였다. 그림 5는 토픽맵을 관계형 데이터베이스에 저장하기 위한 스키마 구조를 보여주고 있으며 그림 6은 관계형 데이터베이스에 저장하기 위한 스토리지 래퍼의 클래스 구조도이다. 토픽맵 인터페이스를 상속받아 관계형 데이터베이스에 저장하기 위해 RDB인터페이스를 생성하였다.

4.4 사용자 중심의 토픽맵 검색

K-Box 시스템에서는 토픽 중심의 키워드 검색과 주제 중심의 디렉토리 검색 그리고 포커스를 이동해 가면서 검색하는 그래프 네비게이션을 지원한다. 그림 7에서 보면 왼쪽에 트리구조로 나오는 부분이 주제 중심의 검색부분이고 상단의 검색창은 키워드 검색부분이다. 그리고 토픽맵 브라우저를 실행하면 그래프 형태의 토픽맵 구조를 보여주는 비주얼 네비게이션 검색이 가능하다.

사용자가 검색할 키워드를 입력하면 질의 처리기가 현재 사용하는 스토리지에 따라 적합한 질의어로 변환

한다. 관계형 데이터베이스에 저장된 경우 SQL문으로 TOPIC 테이블의 BaseName 필드에 대한 조건식을 표현하여 키워드가 들어있는 토픽을 검색하게 된다. 스토리지에서 검색된 토픽들은 토픽맵 제공자에 의해 메모리 내에 객체로 생성되어 사용자에게 제공된다.

그리고 키워드에 해당하는 토픽들의 리스트를 확인한 사용자가 그 중에서 특정 토픽을 선택하면 그 토픽과 관련된 상세 정보를 OCCURRENCE 테이블과 ASSOCIATION 테이블에서 검색하여 보여준다.

ISO에서는 토픽맵의 검색을 위한 질의어로 SQL과 유사한 구문을 가지는 TMQL(Topic Map Query Language)를 제안하고 있으나 현재 작업문서 수준으로 현재까지 완료된 버전이 배포되지 않았다. K-Box 시스템에서는 향후 TMQL이 표준으로 배포되면 이 질의어를 지원할 계획이다.

K-Box 시스템에서는 사용자 프로파일을 이용하여 사용자 중심의 검색이 가능하다. 사용자 중심의 검색이란 대용량의 토픽맵에서 루트 토픽부터 특정 토픽을 그래프를 따라 가면서 찾는 데 소용되는 시간을 줄이고 사용자의 관심 토픽 정보를 프로파일로부터 읽어서 해당 토픽을 직접 접근하는 것이다. 사용자가 K-Box 시스템에 로그인하면 토픽맵 제공자가 사용자의 프로파일을 읽어서 관심 토픽을 검색하여 초기 화면에 보여주게 된다. 이 프로파일은 사용자가 토픽을 추가할 때 자동적으로 그룹화될 수 있는 상위 토픽들을 검색하는 데에도

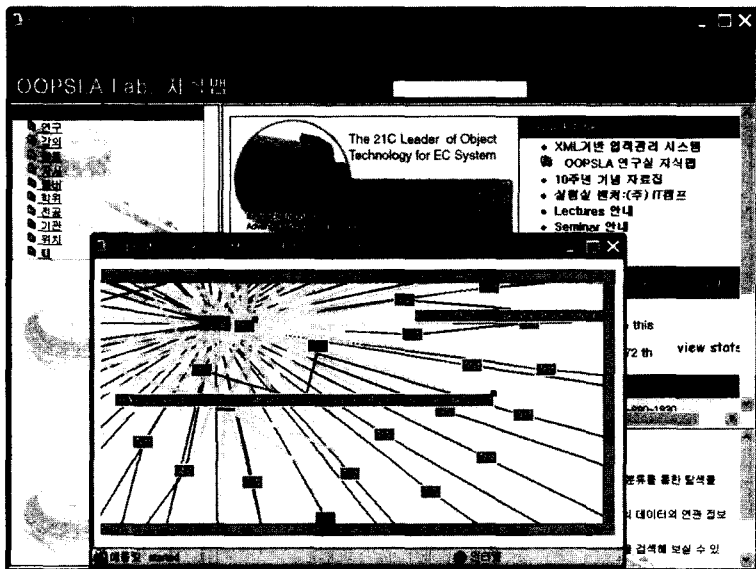


그림 7 K-Box 시스템의 검색유형 - 키워드 검색, 디렉토리 검색, 그래프 네비게이션

사용될 수 있다.

4.5 토픽맵 서비스

K-Box 시스템은 클라이언트-서버 기반으로 동작하며 자바가상기계가 설치된 클라이언트에서는 별도의 소프트웨어 없이 웹 브라우저만으로 토픽맵 서버에 접속할 수 있다. 그리고 웹기반의 분산객체지원을 위해 SOAP (Simple Object Access Protocol)을 사용하고 있다.

K-Box의 SOAP 클래스 구조는 SOAP 지원을 위해 토픽맵 인터페이스를 상속받아 만든 SOAPTopicMap, SOAPTopic, SOAPOccurrence, SOAPAssociation 등의 인터페이스들과 이를 구현한 SOAPTopicMapImpl, SOAPTopicImpl, SOAPOccurrenceImpl, SOAPAssociationImpl 클래스들로 구성된다.

4.6 토픽맵 캐쉬

수백만 개의 토픽들로 구성된 대용량의 토픽맵에서 무엇보다 중요한 것은 검색 및 네비게이션의 성능 문제이다. 특히 네트워크를 통하여 온톨로지 서버에 접속하는 에이전트들에게 있어서 검색하고자 하는 토픽 및 그 토픽과 연관 관계에 있는 토픽들을 빠른 시간 내에 접근할 수 있도록 지원할 수 있어야 한다. 이를 위해 K-Box 시스템은 토픽맵 관리자 내부에 캐쉬 생성기와 캐쉬 관리자를 둬으로써 토픽맵 캐싱에 의하여 검색 및 네비게이션 속도를 향상시켰다.

캐쉬 생성기는 토픽맵 정보에 기반한 클러스터링과 휴리스틱 기법으로 캐쉬를 생성하고 캐쉬 관리자는 사용자별 캐쉬 관리 및 캐쉬 교체를 담당한다. 그림 8은 중심 토픽에서 그래프 상의 거리 3과 6 범위내에 존재하는 연관 토픽들을 클러스터링하는 것을 보여준다. K-Box에서 토픽에 대한 요청을 캐쉬 관리자에게 전송하면 캐쉬 관리자는 캐쉬를 탐색하여 해당 토픽이 있으면 캐쉬에서 읽어서 돌려주고 없으면 캐쉬 생성기에게 새로운 캐쉬 생성을 요청한다. 캐쉬 생성기는 스토리지에 있는 토픽들을 가져온 다음 토픽맵 캐쉬 알고리즘을 통해 캐쉬를 생성하고 캐쉬 관리자에게 돌려준다. 캐쉬 관리자는 기존 캐쉬와 새로 생성된 캐쉬 중 새로 추가

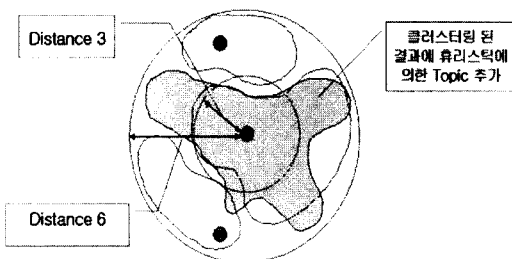


그림 8 클러스터링을 통해 생성된 캐쉬

된 부분만을 추가하고 캐쉬의 용량에 맞게 조절한 후 전송할 캐쉬를 만들어 클라이언트에게 전송한다.

5. 구현결과

실험 자료는 서울대학교 객체지향시스템 연구실 웹사이트와 대우전자 세탁기 연구소 지식직원들을 대상으로 하였다. 먼저 각각의 온톨로지를 생성하고 이 온톨로지를 XTM으로 작성하여 K-Box 시스템에 입력함으로써 토픽맵을 구축하였다. 아래에 보여지는 그림에서는 토픽맵에 기반한 검색이 일반적인 검색에 비해 효율적임을 보이고 있다.

5.1 온톨로지 기반의 연관검색

그림 9는 HTML로 작성된 웹페이지를 보여주고 있으며 그림 10은 동일한 자료를 K-Box의 온톨로지에 기반하여 검색한 결과를 보이고 있다. 그림 9의 HTML 작성된 웹페이지에서는 논문 제목의 하이퍼링크를 따라 가면서 검색하도록 되어 있다. 만일 저자의 다른 논문을 검색하거나 공동저자의 논문을 검색하고자 할 경우 또는 논문이 실린 저널에 대한 소개나 동일 저널에 실린 다른 논문들을 검색하고자 할 경우 그림 9 형태의 웹페이지는 사람이 수작업으로 페이지 내용을 읽으면서 찾아다니야 한다. 하이퍼링크가 없는 경우는 다른 검색엔진을 이용해서 자료를 찾아야 한다.

이에 반해 그림 10의 온톨로지 기반의 연관검색은 논문이라는 클래스에 속하는 인스턴스들(여기서는 논문제목)을 나열하고 선택된 논문에 대한 내용을 검색창에 보임과 동시에 그 논문과 연관된 다른 정보(예, 저자, 저널)를 동시에 보여준다. 그림 10을 보면 선택된 논문의 저자가 '이상원'이며 공동저자로 지도교수가 '김형주'

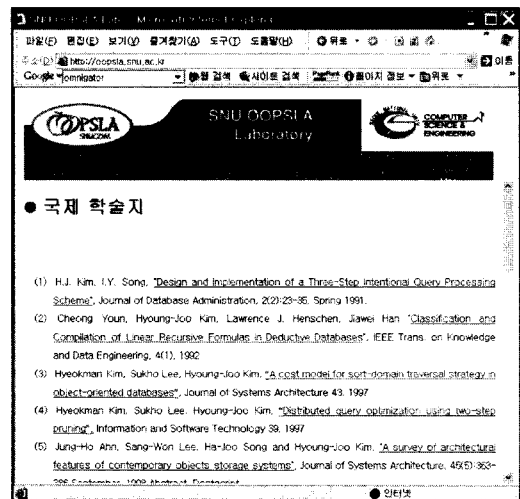


그림 9 HTML로 작성된 웹페이지

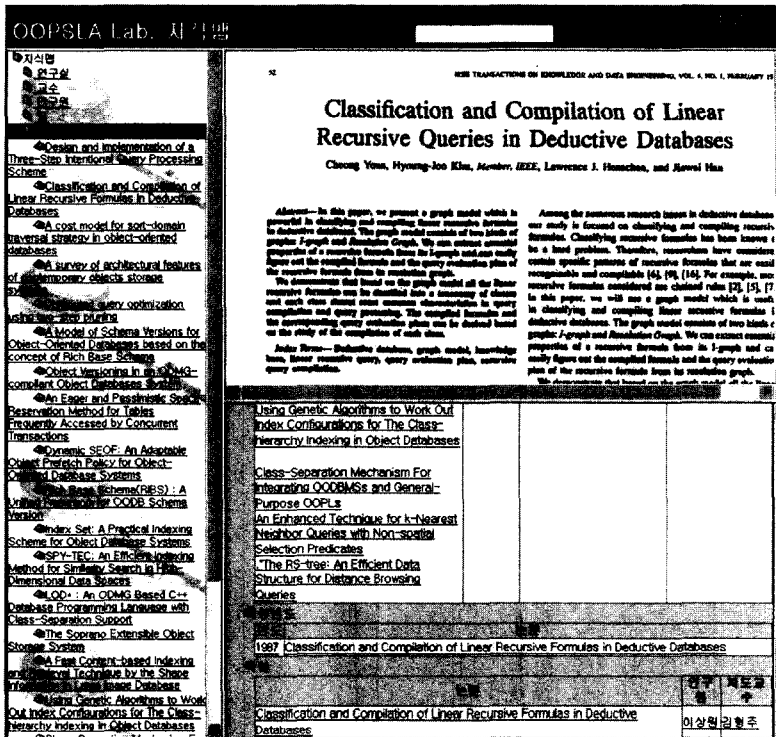


그림 10 온톨로지 기반의 연관정보 검색

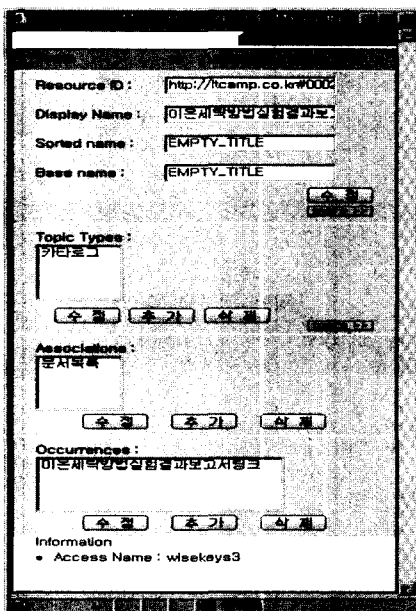


그림 11 토픽 편집 다이얼로그

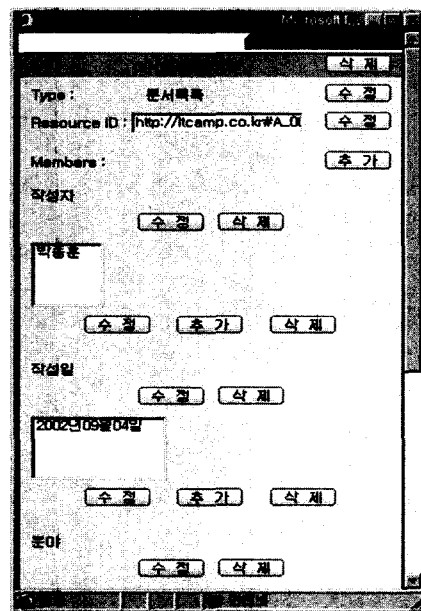


그림 12 토픽간의 연관관계 편집

라는 정보를 보이고 있으며 이상원을 선택할 경우 이상원이 중심 토픽이 되어 이상원의 홈페이지, 논문 리스트

등을 보여준다. 만일 그림 10에서 작성년도를 선택할 경우 동일한 년도에 발간된 논문 리스트를 보여준다.

5.2 온톨로지 편집

토픽 편집기에서는 토픽의 기본 속성과 토픽의 타입, 토픽과 연관된 어소시에이션 정보들, 토픽의 어커런스 정보들을 편집할 수 있다. 어소시에이션 정보 편집에서는 어소시에이션의 타입, 어소시에이션에 참여하고 있는 멤버 토픽, 멤버의 Role Spec을 편집할 수 있다.

토픽의 편집은 시스템 관리자나 사용자 중에 편집 권한을 가진 사용자만 가능하다. 새로운 지식을 추가할 때 지식의 토픽을 새로 만들거나 아니면 기존 토픽의 어커런스를 변경한다. 또한 그 토픽과 관련있는 토픽들이 있다면 서로 간에 연관관계를 어소시에이션 편집 폼을 이용하여 추가한다.

5.3 K-Box와 타 시스템과의 기능비교

표 2는 K-Box와 관련연구에서 소개한 타 시스템, 온토피아 Knowledge Suite, OntoEdit, KAON의 주요 기능을 비교하고 있다[18].

다른 시스템에 비해 K-Box는 다양한 온톨로지 저장장치를 지원하며 온톨로지의 검색 성능을 향상시키기 위한 캐싱 기능을 제공한다. 또한 프로그래밍 언어 및 하드웨어 플랫폼, 프로토콜에 독립적인 구조를 가지므로 웹서비스 기술과 연계하여 온톨로지 서비스를 제공할 수 있다.

6. 결론 및 향후연구방향

시맨틱 웹은 기계 - 지능형 에이전트 - 가 웹 콘텐츠에 담겨있는 의미 정보를 이해함으로써 보다 정확한 정보 검색과 자동화된 웹서비스를 제공하는 확장된 개념의 웹을 가리킨다. 이 시맨틱 웹을 구축하기 위한 여러 기술들 중에 온톨로지 기술은 핵심적인 요소이다. 온톨로지는 개념 또는 지식을 형상화, 구체화한 것으로 지식을 설명하는 용어와 그 용어들간의 관계를 정의함으로써 생성된다. 온톨로지는 시맨틱 웹에서 에이전트들간에

지식을 공유하고 커뮤니케이션하기 위한 기반이 된다.

시맨틱 웹을 구축하기 위해서는 지식을 담고 있는 각 사이트들이 먼저 지식도메인에 적합한 온톨로지를 구축하고 변경되는 지식에 따라 온톨로지를 진화시켜 나가야 한다. 웹사이트에 온톨로지를 구축하기 위해서는 온톨로지 표현 언어를 이해하고 온톨로지 모델을 생성, 저장하며 사용자의 편집 및 검색 요구를 처리할 수 있는 온톨로지 관리 시스템이 있어야 한다. 본 논문에서는 토픽맵 기반의 온톨로지 관리 시스템을 제안하였다. 토픽맵은 RDF와 함께 온톨로지를 컴퓨터상에 구축할 수 있는 표준화된 데이터 모델이다.

본 논문의 온톨로지 관리 시스템은 토픽맵 모델을 기반으로 온톨로지를 생성, 저장, 검색 및 네비게이션이 가능한 온톨로지 서버 역할을 수행한다. 대용량의 온톨로지가 구축된 상황에서 온톨로지 검색 및 네비게이션 성능을 높이기 위해 토픽들간의 연결 가중치에 따라 검색하는 토픽과 연관관계가 있는 다른 토픽들을 미리 읽어오도록 온톨로지 캐싱을 수행하며 토픽맵 스키마에 의한 제약조건을 적용함으로써 비일관된 온톨로지 구조가 생성되는 것을 사전에 방지할 수 있다.

본 논문의 온톨로지 관리 시스템은 향후 지속적으로 연구, 개발할 온톨로지 브로커 시스템의 베이스 시스템 역할을 담당하게 된다. 웹에서는 모든 정보들이 분산되어 있으므로 온톨로지 또한 서로 다른 자연 언어와 내부 표현 모델, 지식 내용 등을 가지며 각 사이트에 분산되어 관리된다. 에이전트들이 자동화된 웹서비스를 제공하도록 하기 위해서는 여러 사이트의 상이한 온톨로지에 접근해야 하는데 이때 에이전트를 대신하여 각 사이트의 온톨로지들의 상위 수준의 통합 개념의 온톨로지를 구축하고 관리하며 에이전트에게 서비스하는 역할이 온톨로지 브로커 시스템의 중요 역할이다.

향후 연구방향은 온톨로지 중개자 시스템의 핵심 기

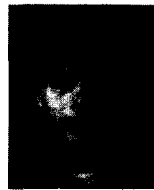
표 2 K-Box 및 타 시스템과의 기능 비교

시스템 기능	K-Box	Ontopia K.Suite	OntoEdit	KAON
온톨로지 모델	토픽맵	토픽맵	RDF	RDF
저장장치	In-memoty, RDB, ODB, XML-DB	In-memory, RDB	자체포맷의 파일	In-memory, RDB, Textfile
검색 방식	키워드검색, 디렉토리, 그래프	키워드검색, 디렉토리	키워드검색, 디렉토리, 그래프	검색기능미약 전체 브라우징
클러스터링(캐싱)	K-means	없음	없음	없음
제약조건검사	지원	지원	지원	지원
지식추론	없음	간단한 형태지원	지원	지원
온톨로지 편집	지원	지원	지원	지원
외부텍스트포맷	XTM	XTM, LTM	RDF, DAML	없음
API제공	지원	지원	없음	지원

술에 대한 것들이다. 상이한 구조의 온톨로지들의 통합을 위해서는 온톨로지들간의 유사성 검사, 기존 온톨로지들에서 새로운 사실을 발견하여 확장된 개념의 새로운 온톨로지를 생성하거나 진화시키기 위한 추론 엔진, 온톨로지들의 부분적 결합 및 분리 등에 대한 연구가 필요하며 에이전트에게 온톨로지를 서비스하기 위해서는 자동화된 온톨로지 분류 및 온톨로지 변화에 따른 제약조건 정당성 검사, 온톨로지 트랜잭션 처리 등에 대한 연구가 필요하다.

참 고 문 헌

- [1] T. Berners-Lee. "Semantic Web Roadmap," <http://www.w3.org/DesignIssues/Semantic.html>.
- [2] T. Berners-Lee, J. Hendler. O. Lassila. "The Semantic Web," *Scientific American*, May 2001.
- [3] S. Staab, H.-P. Schnurr, R. Studer, and Y. Sure. "Knowledge processes and ontologies," *IEEE Intelligent Systems, Special Issue on Knowledge Management*, 16(1), January/February 2001.
- [4] Ora Lassila and Ralph R. Swick. "Resource Description Framework(RDF) Model and Syntax Specification," W3C Recommendation 22 February 1999, URL:<http://www.w3.org/TR/REV-rdf-syntax>.
- [5] Horrocks, I., et. al. "The DAML+OIL Language Specification," <http://www.daml.org/2001/03/daml+oil.daml>
- [6] Michel Biezunski, Martin Bryan and Steve Newcomb. ISO/IEC 13250 TopicMaps.
- [7] Graham Moore. "Topic Map technology - the state of the art," *XML 2000 Conference & Exposition, Washington, USA, December 2000*.
- [8] Product White Paper. "The Ontopia Knowledge Suite," Ontopia, Inc.
- [9] Y. Sure, S. Staab, J. Angele, D. Wenke, and A. Maedche. "OntoEdit: Guiding ontology development by methodology and inferencing," *Prestigious Applications of intelligent Systems(PAIS), in conjunction with ECAI 2002, Lion, France, July 2002*.
- [10] Alexander Maedche, Boris Motik and Raphael Volz. "KAON-A Framework for semantics-based E-Services," Institute AIFB, University of Karlsruhe.
- [11] Martin Lacher and Stefan Decker. "On the integration of Topic Map data and RDF data," *Extreme Markup Languages 2001 Conference & Exposition, Montréal, Canada, August 2001*.
- [12] Steve Pepper and Graham Moore. "XML Topic Maps(XTM) 1.0," TopicMaps.Org.
- [13] Tova Milo and Dan Suciu, "Index Structures for Path Expressions," *ICDT, 1999*.
- [14] Roy Goldman and Jennifer Widom. "DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases," *VLDB, 1997*.
- [15] Jason McHugh and Jennifer Widom, "Query Optimization for XML," *VLDB, 1999*.
- [16] Sangwon Park and Hyoung-Joo Kim. "A New Query Processing Technique for XML Based on Signature," *DASFAA, April 2001*.
- [17] 박상원, 김형주, "시그니처를 이용한 XML 질의 최적화 방법," *정보과학회 논문지(데이터베이스)*, 28(1), March 2001.
- [18] Aimilia Magkanaraki, Grigoris Karvounarakis, Ta Tuan Anh, Vassilis Christophides, Dimitris Plexousakis, "Ontology Storage and Querying," Technical Report No 308., ICS-FORTH, 2002.



김 정 민

1992년 홍익대학교 전자계산학과 졸업(학사). 1994년 홍익대학교 전자계산학과 졸업(석사). 2002년 서울대학교 전기, 컴퓨터공학부 박사과정 수료. 관심분야는 Semantic Web, Ontology, IR, Database



박 철 만

1998년 경북대학교 컴퓨터공학과 졸업(학사). 2001년 서울대학교 컴퓨터 공학과(석사). 2003년 서울대학교 전기, 컴퓨터공학부 박사과정 수료. 관심분야는 Semantic Web, Ontology, XML



정 준 원

2000년 동국대학교 컴퓨터공학과 졸업(학사). 2003년 서울대학교 전기, 컴퓨터공학부 졸업(석사). 2003년~현재 서울대학교 전기, 컴퓨터공학부 박사과정 재학중. 관심분야는 Semantic Web, Ontology, XML



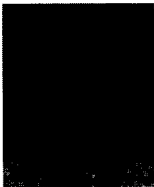
이 한 준

2001년 서울대학교 컴퓨터공학과 졸업(학사). 2001년~현재 서울대학교 전기, 컴퓨터공학부 석사과정 재학중. 관심분야는 Semantic Web, Ontology, XML



민 경 섭

1995년 한국항공대학교 컴퓨터공학과 졸업(학사). 1997년 서울대학교 컴퓨터공학과 졸업(석사). 1997년~현재 서울대학교 인지과학 박사과정 재학중. 관심분야는 XML, IR, Database



김 형 주

1982년 서울대학교 전산학과(학사). 1985년 미국 텍사스 대학교 대학원 전산학(석사). 1988년 미국 텍사스 대학교 대학원 전산학(박사). 1988년 5월~1988년 9월 미국 텍사스 대학교 POST-DOC 1988년 9월~1990년 12월 미국 조지아 공과대학 조교수. 1991년~현재 서울대학교 컴퓨터공학부 교수