

온톨로지 기반의 지식맵 서비스 시스템의 설계 및 구현

김정민^o 박철만 정준원 이한준 정호영 민경섭 김형주
서울대학교

{jmkim^o, cmpark, jwjung, hjlee, hyjung, ksmin, hjk}@snu.ac.kr

Seoul National University

Jungmin Kim^o Chulman Park Junwon Jung Hanjun Lee Hoyoung Jung
Kyungsub Min Hyungjoo Kim
School of Computer Engineering, Seoul National University

요 약

지식관리시스템의 지식 분류 체계가 지식맵이며 이 지식맵을 이용하여 지식 서비스가 제공된다. 그러나 현재의 지식맵은 비표준적인 지식 네트워크 구조와 정적인 분류체계라는 문제를 가진다. 문제 해결 방법은 지식맵에 온톨로지를 적용해서 표준성, 의미 기반의 지식 네트워크 구조, 동적인 지식분류, 자동화된 지식 서비스를 제공하는 새로운 개념의 지식맵을 구현하는 것이다. 본 논문에서는 토픽맵 모델을 기반으로 온톨로지를 생성, 저장, 검색하는 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.

1. 서 론

지식맵(knowledge map)은 수많은 지식이 저장된 지식베이스(knowledge base)에서 원하는 지식을 쉽게 검색할 수 있는 지도 역할을 한다. 그러나 현재의 지식맵은 구조적으로 몇 가지 문제점을 가지고 있다. 첫째, 시스템 종속적이며 계층구조의 느슨한 지식 분류 체계를 가진다. 둘째는 지식맵의 지식내용이 데이터 중심 보다는 업무 프로세스 중심적이다. 셋째는 지식맵의 표준화된 데이터모델이 없으므로 분산 구조의 지식맵을 통합하거나 지식 교환이 어렵다. 넷째는 표준화된 질의 언어가 없기 때문에 복잡한 질의를 처리할 수 없다는 것이다. 이러한 문제들은 해결하기 위해서는 지식을 의미 기반으로 조직화하고 표현할 수 있는 표준 데이터 모델을 두는 것이다.

지식을 사람이나 기계들 사이에 공유하기 위해서는 상호간에 이해 가능한 용어들로 개념을 정의한 다음 표준화된 모델과 형식화된 언어로 표현해야 한다. 시맨틱을 정의하는 방법으로 인공지능 분야에서 연구되어 온 온톨로지(ontologies)가 있다. 온톨로지는 특정 도메인의 개념 및 지식을 명세화하기 위해 그 지식을 설명하는 표준 용어(term)들을 정의하고 용어들 사이의 관계(taxonomy and relation)를 정의하는 것이다. 같은 지식 도메인의 서로 다른 웹사이트들에서 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 지식서비스를 제공할 수 있다.[1]

시맨틱웹(Semantic Web)에서 온톨로지를 표현하고 인코딩(encoding)하기 위해 사용되는 데이터 모델로 RDF(Resource Description Framework)[2], RDF Schema[2], DAML+OIL[3] 그리고 Topic Map[4]이 있다. 본 논문에서는 토픽맵을 기반으로 온톨로지를 생성, 변경, 저장 및 검색하는 온톨로지 관리 시스템(Ontology Management System) K-Box 시스템을 제안한다.

2. 관련연구

토픽맵 관리 시스템으로 상용화된 시스템은 Ontopia에서 개발한 Ontopia Knowledge Suite(OKS)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 리파지토리 등 다양한 스토리지를 지원하는 스토리지 래퍼와 토픽맵 캐싱, 그리고 토픽맵 제약조건을 정의하는 토픽맵 스키마 지원 등의 향상된 기능을 가지고 있다.

3. 지식맵 데이터 모델

K-Box 시스템에서는 지식맵 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 데이

터 모델로 사용되고 있다. 토픽맵 모델의 핵심 요소는 토픽(Topic), 어커런스(Occurrence), 어소시에이션(Association)으로 볼 수 있다. 토픽은 객체지향 모델에서의 클래스나 객체에 해당하는 것으로 표현하고자 하는 대상을 가리키고 어커런스는 토픽에 종속되는 개념으로 해당 토픽에 대한 실제 내용이 담겨있는 자원의 주소(URL로 표현된다)나 지식, 데이터 자체를 가리킨다. 그리고 토픽들 사이에는 토픽타입과 토픽간의 상하관계와 토픽들간의 연관성을 표현하는 어소시에이션이 있다.[4]

온톨로지를 구축하기 위해서는 지식을 설명할 수 있는 용어(term)를 정의하고 용어들 사이의 관계를 정의한다. 용어들 사이의 관계는 분류 체계를 표현하는 taxonomy와 개념간의 연관관계를 표현하는 relation으로 나뉘어진다.

온톨로지가 정의 되면 용어를 토픽으로, taxonomy를 토픽 타입과 토픽간의 관계로, 그리고 relation을 토픽들 사이의 어소시에이션으로 매핑하여 초기 토픽맵을 생성한다. 그림 1에서는 온톨로지 정의 및 토픽맵 데이터 모델을 생성하는 과정에 대해 보여주고 있다.

의하고 있다.

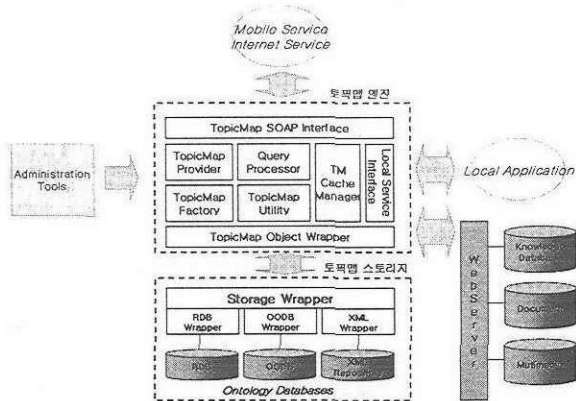


그림 2 K-Box 시스템의 토픽맵 기반 온톨로지 관리구조
토픽맵 스토리지 래퍼는 토픽맵 오브젝트 래퍼를 상속받아 각 스토리지에 맞게 구현된 클래스들을 의미한다. 토픽맵 스토리지 래퍼는 토픽맵의 각 객체들을 실제의 스토리지에 영구적으로 저장하고 로드하는 등의 작업을 수행한다. 토픽맵 스토리지 래퍼는 토픽맵 오브젝트 래퍼를 통해서 하위의 스토리지에 관계없이 일관적인 인터페이스를 제공해 준다.

토픽맵 생성자는 토픽맵의 각 구성요소들을 생성하는 역할을 수행한다. XTM 문서가 입력으로 들어오면 SAX 파서가 이 문서를 분석하여 구문상의 오류를 체크하고 토픽맵 DTD에 적합한 경우 엘리먼트 정보와 이벤트를 토픽맵 파서 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. 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 제공되며, 토픽맵 오브젝트 래퍼의 제공자 인터페이스를 통해 일관적인 인터페이스를 제공할 수 있다.

토픽맵 캐시 관리자는 대용량의 토픽맵의 검색 효율을 높이기 위한 것으로 특정 토픽을 검색하면서 그 토픽과 연관 관계에 있는 다른 토픽들의 정보도 미리 검색하여 캐시에 저장한 다음 사용자의 검색 조건에 부합되는 토픽이 있으면 디스크를 접근하지 않고 캐시에서 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 토

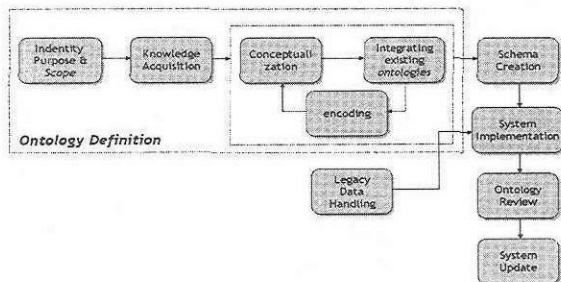


그림 1 온톨로지 정의 및 데이터 모델 생성 절차

토픽맵을 표현하는 기술 언어로 XTM(XML Topic Maps)[5]이 있다. XTM은 토픽맵 모델의 각 요소를 나타내는 태그 집합과 엘리먼트들 사이의 구조를 정의하고 있으며 토픽맵을 기술하는 표준포맷으로 사용되고 있다.

4. K-Box 시스템 설계 및 구현

4.1 K-Box 시스템 구조

K-Box 시스템은 토픽맵 데이터를 저장하고 관리하는 역할을 수행하기 위해 그림 2에서 볼 수 있듯이 토픽맵 객체들의 단일 인터페이스 제공을 위한 토픽맵 오브젝트 래퍼, 토픽맵 객체들을 저장하기 위한 스토리지 래퍼, 사용자가 요청하는 토픽맵 객체를 전달하는 토픽맵 제공자, 대용량의 토픽맵의 효율적인 검색을 위한 토픽맵 캐시 관리자, 토픽맵 객체들을 생성하는 토픽맵 생성자, 토픽맵 관리를 위해 일관된 인터페이스를 제공하는 토픽맵 관리자, 토픽맵의 импорт(import), 익스포트(export) 및 여러 기능들을 제공하는 토픽맵 도구들로 구성된다.

토픽맵 오브젝트 래퍼는 토픽맵을 구성하고 있는 오브젝트들의 구현에 비종속적으로 일관된 인터페이스를 제공해주는 역할을 한다. 토픽맵 오브젝트 래퍼는 자바의 인터페이스로 정의가 되며, 각각의 오브젝트들의 메소드들을 정

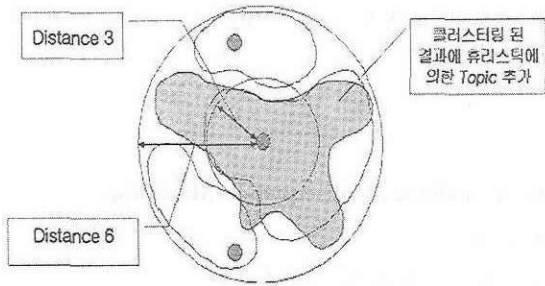


그림 3 클러스터링을 통해 생성된 캐쉬

픽맵 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 온톨로지 무결성을 위한 제약조건 검사

토픽맵 파서는 온톨로지의 무결성을 보장하기 위해 온톨로지 제약조건을 사전에 검사한다. SAX 파서의 경우 XTM 문서의 구문 오류를 검사하는 역할이지만 토픽맵 파서는 어소시에이션의 카디널리티(cardinality), 상위 토픽이 가질 수 있는 하위 토픽의 종류, 어커런스 타입, 어소시에이션의 롤(role) 등의 제약조건을 검증하기 위해 토픽맵 스키마를 참조한다.

토픽맵 스키마는 제약조건 룰(rule) 테이블로서 온톨로지 설계할 때 설계자에 의해 정의된다. 토픽맵의 토픽 객체 집합과 토픽맵 스키마의 룰 집합 사이에 매핑을 시킴으로써 토픽맵 파서는 특정 토픽에 변경이 발생할 때 그 토픽과 연결된 룰들을 찾아서 제약조건 검사를 수행한다. 표 1에서는 토픽맵 파서가 검사하는 제약조건 종류의 종류를 보여주고 있다.

표 1 토픽맵 파서에 의해 검사 가능한 제약조건 종류

제약조건 대상	제약조건 유형	설명
TOPIC	hasSubType	하위토픽으로 가지는 토픽타입
	isOccurrence	어커런스를 가지는지 여부
	isTerminal	하위 토픽을 가지는지 여부
OCCURRENCE	instanceOf	특정 타입의 어커런스 허용
	min, max	어커런스의 카디널리티
ASSOCIATION	data, ref, any	어커런스의 종류
	min, max	어소시에이션의 카디널리티
	roleTopicRef	어소시에이션의 player type

5. 구현결과

실험 자료는 서울대학교 객체지향시스템 연구실 웹사이트와 대우전자 세탁기 연구소 지식자원들을 대상으로 하였다. 먼저 각각의 온톨로지를 생성하고 이 온톨로지를 XTM으로 작성하여 K-Box 시스템에 입력함으로써 토픽맵을 구축하였다.

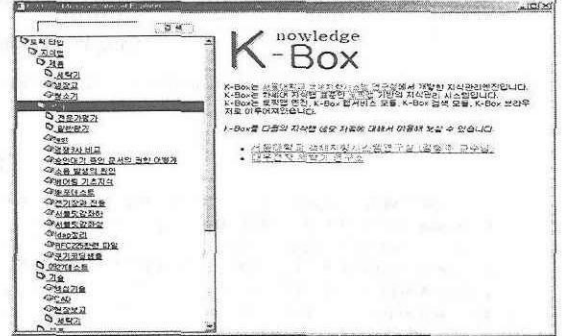


그림 4 K-Box 시스템 메인 윈도우

6. 결론

현 지식맵 구조의 문제점을 해결하기 위해 표준적, 개방적, 그리고 시맨틱 기반의 지식맵을 작성하는 방법으로 본 논문에서는 토픽맵 기반의 지식맵 서비스 시스템을 제안하였다. 토픽맵은 RDF와 함께 온톨로지를 컴퓨터상에 구축할 수 있는 표준화된 데이터 모델이다. 본 논문의 지식맵 서비스 시스템은 기본적으로 지식맵의 생성, 저장, 검색 및 네비게이션을 지원하는 지식맵 서버 역할을 수행한다. 또한 대용량의 지식맵 환경에서 검색 성능을 높이기 위해 토픽맵 사이의 연결 가중치를 계산하여 중심 토픽과 연관관계가 있는 다른 토픽들을 미리 읽어오도록 지식맵 캐싱을 수행하며 토픽맵 스키마에 의한 제약조건을 적용함으로써 비일관된 구조의 지식맵이 생성되지 않도록 일관성을 유지한다.

참고 문헌

[1] James Hendler. "Agents and the Semantic Web", *IEEE Intelligent Systems Journal*, March/April 2001
 [2] 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>
 [3] Horrocks, I., et. al. "The DAML+OIL Language Specification", <http://www.daml.org/2001/03/daml+oil.daml>
 [4] Michel Biezunski, Martin Bryan and Steve Newcomb. ISO/IEC 13250 TopicMaps.
 [5] Steve Pepper and Graham Moore. "XML Topic Maps(XTM) 1.0", TopicMaps.Org.
 [6] 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.