

토픽맵 기반 디지털 콘텐츠 표준화

서울시립대학교 | 이재호*

1. 개요

디지털 콘텐츠의 효과적인 활용을 위해서는 디지털 콘텐츠의 발굴, 생성, 저작, 표현, 저장, 검색, 서비스, 유통, 보안, 추적 등의 전 과정에 걸친 표준화를 통한 상호 호환성 및 운용성의 확보가 필요하다. 이를 위해서는 디지털 콘텐츠 표현을 위한 표준도 필요하지만 위의 전 과정에 필요한 디지털 콘텐츠 메타데이터(metadata)의 표준화도 요구된다. 더불어 자동화된 컴퓨터 시스템에 의한 효율적 검색 및 처리를 위해서는 메타데이터 의미(semantics) 간의 호환성도 필연적으로 요구된다.

새로운 콘텐츠의 지속적인 등장과 발전에 따라 이러한 메타데이터의 호환적 의미는 확장성을 전제로 표현될 수 있어야 한다. 끊임없이 증가하는 방대한 디지털 콘텐츠 간의 확장적 상호운용성을 확보하고 유지하기 위해서는 인터넷 상의 정보 자원이 의미 기반 기계가독형으로 표현됨으로써 웹이 단순한 자원의 군집이 아닌 의미의 결합체로의 발전이 이루어 져야 한다. 이러한 변화에서 중추적 역할을 하는 것이 바로 온톨로지이다.

명시적 개념화를 추구하는 온톨로지는 상호 공유될 의미를 표준적인 방법으로 표현하여 소프트웨어 간의 의미 수준의 상호운용성을 확보하는 방안을 제공한다. 디지털 콘텐츠를 비롯한 자원의 내용에 관한 메타정보를 기계가독형으로 구축하여 의미 기반 검색을 비롯한 디지털 콘텐츠의 활용을 강화할 수 있는 메타데이터 시스템 개발의 핵심은 국제적으로 통용되는 표준화된 온톨로지의 구축에 있다. 온톨로지 관련 표준은 디지털 콘텐츠의 체계적이고 명확한 정의와 상호 연계를 통한 정보 자원의 효율적 검색, 통합, 재사용을 가능케 하는 규격으로서 미래 정보 및 지식관리에 지대한 영향을 미칠 것으로 예측된다. 체계적으로 개발된 표준적 온톨로지는 그 자체로도 귀중한 디지털 콘텐츠

자산으로서 지식과 정보를 다양하게 접근하여 관련 지식 또는 정보를 서로 긴밀히 연결된 상태로 볼 수 있게 한다. 동일한 원천 지식과 정보에 대해서도 각각도의 온톨로지를 적용하면 사용자의 요구에 따른 차별적 지식체계를 제공할 수 있다.

온톨로지 기술표준은 국제표준기구인 ISO/IEC에서 추진하는 토픽맵(Topic Maps)과 W3C가 추진하고 있는 RDF/OWL 중심의 시맨틱웹(Semantic Web)접근 방식이 있다. 이들은 자원의 포괄적 표현과 자원 관계의 의미 있는 연계를 위한 대표적 온톨로지 관련 표준으로서 경쟁적 측면과 동시에 상호 보완적인 요소를 가지고 있다.

토픽맵(Topic Maps)은 주제 중심 지식의 표현과 상호호환을 위한 국제표준화기구(ISO, 아이소) 표준으로서 정보의 연계를 통한 정보의 발견(findability)에 중점을 두고 있다. 토픽맵은 일반적 온톨로지 분야에서 추구하는 명시적 개념화를 위한 지식의 표현과 연계 및 추론과는 다소 주안점이 다르나 그 효과 측면에서는 같은 목적을 위해 사용될 수 있다.

국제민간표준포럼인 W3C가 추진해온 시맨틱웹은 URI, XML 네임스페이스와 같은 기초 표준과 이를 기반으로 한 활용한 RDF, RDF스키마, OWL 마크업 언어 등으로 구성된 구조를 갖고 있다.

이 글에서는 디지털 콘텐츠의 정보를 표현하고 이를 효과적으로 활용하기 위한 메타정보 표현 및 연계 수단으로서의 토픽맵의 의의를 제시하는 목적에서 토픽맵의 개요 및 구성과 더불어 관련된 표준화 활동을 소개한다.

2. 토픽맵의 개요

2.1 토픽맵의 구성

토픽맵(Topic Maps)은 지식을 코드화하고 코드화된 지식을 관련 정보 자원과 연계하기 위한 추상적 구조를 일컫는다. 토픽맵의 개념적 모델은 그림 1에 보인 바와 같이 정보와 콘텐츠 자원 영역인 정보층(infor-

* 중신회원

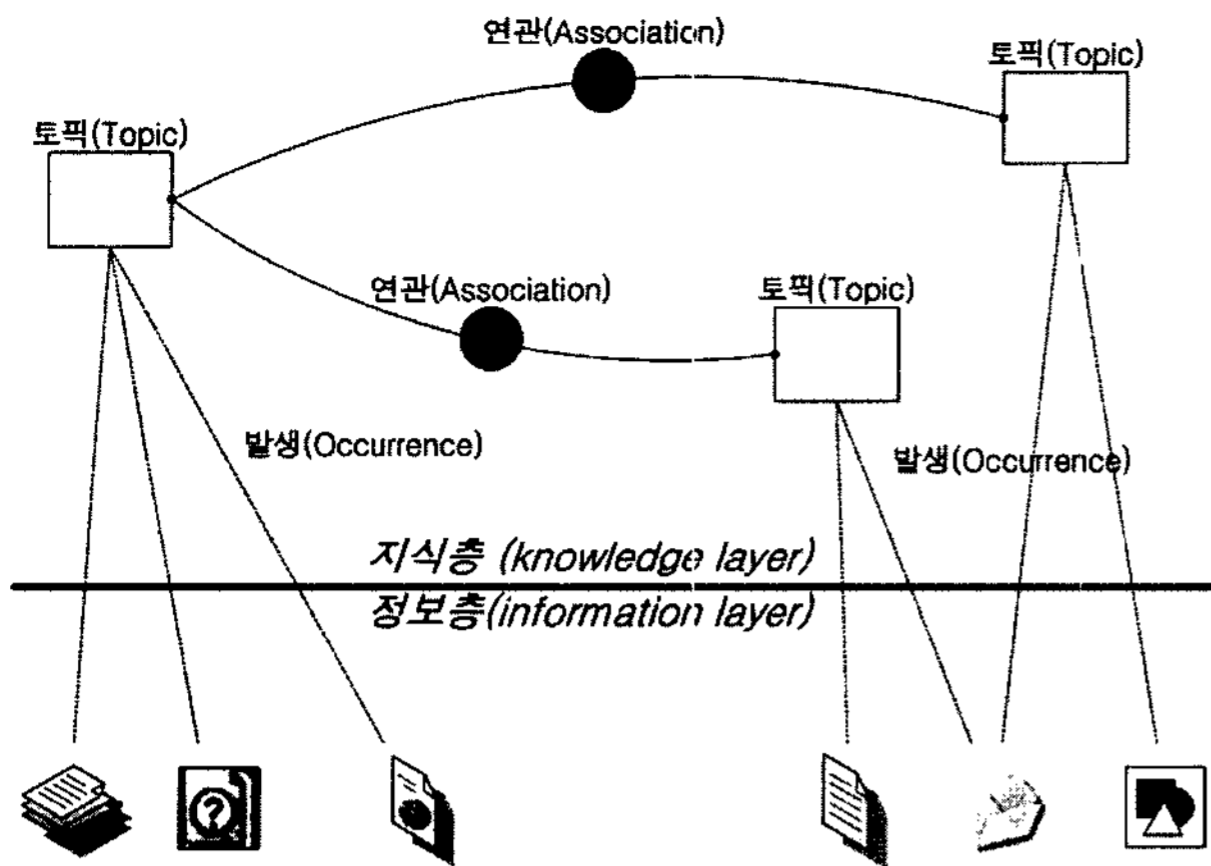


그림 1 토픽맵의 개념적 구성 요소

mation layer)과 이와 연관된 메타정보와 지식구조 표현 영역인 지식층(knowledge layer)으로 구성된 이원화된 구조를 기반으로 하고 있다.

지식층은 주제(subject)를 표현하는 토픽(topic)과 이러한 주제 간의 연관 관계를 나타내는 연관(association)으로 구성된다. 지식층에 토픽으로 표현된 주제와 정보층의 정보 자원(information resource)과의 연관 관계는 발생(occurrence)으로 표현된다.

토픽은 주제(subject)를 표현하기 위한 단위이다. 존재 여부나 특정 성격에 관계없이 이에 대해 무엇이든 무슨 수단에 의해서든 서술될 수 있는 모든 대상이 주제가 될 수 있다.

토픽은 고유의 식별자를 통하여 구분되며 단일한 토픽타입(topic type)을 갖는다. 더불어 토픽은 다양한 명칭(name)들을 갖는 것이 가능하다.

발생(occurrence)은 토픽으로 표현된 주제와 연관된 정보를 갖고 있는, 디지털 콘텐츠와 같은 정보 자원과의 관계를 일컫는다. 발생은 발생타입(occurrence type)을 갖는다. 이러한 토픽맵의 핵심요소인 토픽(Topic), 연관(Association), 발생(Occurrence)을 토픽맵의 TAO라 일컫는다.

2.2 토픽맵의 병합(merging)

토픽맵은 병합 규칙에 의하여 단일 토픽맵으로 통합될 수 있다. 이는 별개로 개발된 지식정보 시스템을 단일 시스템으로 통합하여 일관된 관점(view)을 제공하고, 전사적인 지식 공유를 가능하게 하며 나아가 분산 지식관리와 조직 간 지식공유를 가능하게 하는 토픽맵의 핵심적 특징이다. 병합 규칙은 토픽마다 부여된 고유 식별자를 기반으로 이루어진다.

2.3 토픽맵의 식별자

토픽맵의 병합은 토픽별로 부여된 고유 식별자에

의해서 가능해진다. 고유 식별자는 다음의 요소로 정의된다.

2.3.1 주제 소재지정자(subject locator)

토픽으로 표현하고자 하는 대상인 주제가 정보 자원 자체인 경우에는 정보 자원의 주소를 토픽의 고유 식별자로 활용할 수 있다. 이때 토픽 표현 주제에 해당하는 정보 자원의 주소를 주제 소재지정자(subject locator)라 한다.

2.3.2 주제 지시자(subject indicator)

토픽으로 표현하고자 하는 주제가 정보 자원 자체가 아닌 경우가 대부분이다. 이 경우에는 해당 정보 자원과 정보 자원의 주소가 없기 때문에 주제 소재지정자를 사용할 수가 없다. 이때 필요한 것이 주제 지시자(subject indicator)이다. 주제 지시자는 토픽이 나타내는 주제를 모호하지 않게 사용자가 식별할 수 있도록 만들어진 정보 자원이다. 주제 지시자는 정보 자원이기 때문에 주소를 갖게 되고 토픽의 주제를 식별하는 데 이 주소를 주제 식별자(subject identifier)로 사용할 수 있게 된다.

2.3.3 주제 식별자(subject identifier)

주제 지시자(subject indicator)를 가리키는 소재지정자(locator)를 말한다.

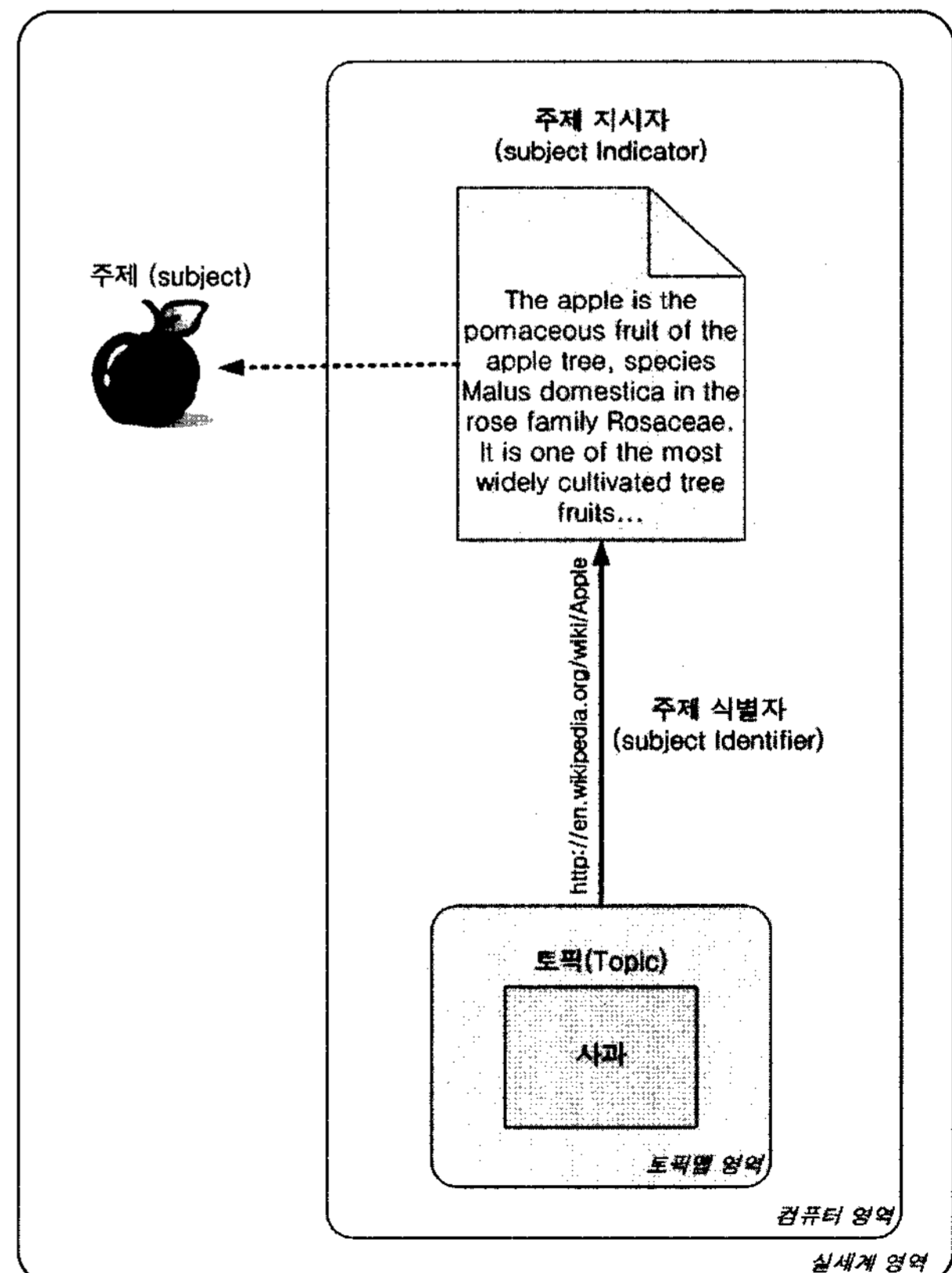


그림 2 주제 지시자 (Subject Indicator)

주제(subject)와 주제 지시자(subject indicator), 주제 식별자(subject identifier)의 관계를 도식적으로 표현하면 그림 2와 같다.

3. 토픽맵 표준화 현황

주요 디지털콘텐츠 표현 및 온톨로지 관련 표준화 활동을 위하여 국내에서는 지식경제부 기술표준원 산하에 ISO/IEC JTC1/SC34 전자문서 처리언어 국내 전문위원회가 구성되어 활발한 활동이 이루어지고 있다. 국제적으로는 ISO/IEC JTC1/SC34 WG3에서 토픽맵 관련 표준화 활동이 이루어지고 있다. 토픽맵과 관련된 주요 표준화 프로젝트 내역은 다음과 같다.

- 13250-1, Part 1: Overview and basic concepts, Motomu Naito
- 13250-2:2006, Part 2: Data model(TMDM), 완료
- 13250-3:2007, Part 3: XML syntax(XTM), 완료
- 13250-4, Part 4: Canonicalization, Lars Marius Garshol, Jaeho Lee
- 13250-5, Part 5: Reference model, Robert Barta, Patrick Durusau, Steven R. Newcomb
- 13250-6, Part 6: Compact Syntax(CTM), Lars Heuer, Gabriel Hopmans, Dmitry Bogachev
- 13250-7, Part 7: Graphical Notation(GTM), Lars Marius Garshol, Jaeho Lee
- 18048, Query Language(TMQL), Robert Barta, Lars Marius Garshol
- 19756, Constraint Language(TMCL), Dmitry Bogachev, Graham Moore
- Expressing Dublin Core Metadata using Topic Maps, Patrick Durusau

3.1 토픽맵 데이터 모델(Topic Maps Data Model)

토픽맵 데이터 모델 (TMDM) 표준은 토픽맵 요소에 대한 개념을 정의하는 토픽맵 메타모델을 규정하는 핵심 표준으로서 토픽맵의 질의어인 TMQL(TM Query Language)과 제약언어인 TMCL과 같은 토픽맵 관련 표준의 기초가 된다. 이러한 개념을 표기하기 위해서는 Topic Maps XML Syntax(XTM)와 사용자의 가독성을 증대시킨 Topic Maps Compact Syntax(CTM) 등이 있다.

3.2 토픽맵 XML 구문(Topic Maps XML Syntax)

토픽맵 데이터 모델(TMDM) 규격으로 정의된 토픽맵을 XML을 사용하여 표기하는 표준이다.

3.3 토픽맵 정규화(Topic Maps Canonicalization)

토픽맵 정규화 규격은 토픽맵의 데이터 구조를 연속

된 바이트의 순차적 배열로 변환하는 표준적인 직렬화 방법을 규정한다. 같은 주제와 개념을 표현한 토픽맵은 정규화를 통하여 같은 직렬화 결과를 갖게 됨으로 직렬화된 데이터를 직접 비교하여 두 토픽맵의 일치 여부를 파악하는 것이 가능해 진다.

3.4 토픽맵 간결 구문(Topic Maps Compact Syntax)

Topic 토픽맵 간결 구문(CTM) 규격은 토픽맵 데이터 모델(TMDM)로 정의된 토픽맵을 간편하고 단순한 구문으로 이루어진 언어를 써서 표기하기 위한 표준이다. XTM이 기계에 의한 토픽맵의 해석과 처리를 용이하기 위한 표준인 반면 CTM의 사람이 토픽맵 데이터 모델 표기하기 상호 의사 소통을 원활하게 하기 위한 표준이다. 또한 CTM으로 표현된 토픽맵은 TMQL 표준으로 질의문을 구성할 때도 사용될 수 있다.

3.5 토픽맵 도식적 표기(Topic Maps Graphical Notation)

토픽맵 도식적 표기(GTM) 규격은 TMDM으로 정의된 토픽맵을 인간 사용자가 토픽맵을 사용하여 도메인 지식을 모델링하기에 편리한 도식적 표기법을 규정한다.

3.6 토픽맵 질의 언어(Topic Maps Query Language: TMQL)

토픽맵 질의 언어(TMQL) 규격은 토픽맵으로 표현된 정보를 질의, 검색, 수정하기 위한 언어와 방법을 정의한다. 질의어는 관계형 데이터베이스 질의어인 SQL에 FLWR(For, Let, Where, Return) 구문을 갖춘 형태를 취한다. 응용 프로그램은 TMQL 질의어를 사용하여 토픽맵에서 필요한 요소를 얻게 된다. 토픽맵의 저장 방법에 관한 사항은 TMQL에서 정의하지 않으며 TMQL은 응용 프로그램과의 인터페이스 역할만 담당한다.

3.7 토픽맵 제한 언어(Topic Maps Constraint Language)

토픽맵 제한 언어(TMCL) 규격은 토픽맵 데이터 모델(TMDM)에 정의된 토픽맵 모델의 스키마와 제약 사항을 정의하는데 목적이 있으며, 보다 풍부한 개념 표현을 위한 토픽맵 기반의 온톨로지 모델링 언어로서 토픽맵의 표현력을 향상시키고 토픽맵 표현에 대한 구문적, 의미적 틀을 제공하기 위해 활용될 수 있다.

3.8 토픽맵 기반 더블링크어 메타데이터

더블링크어는 국제적으로 가장 널리 쓰이고 있는 통합메타데이터 스키마이다. 현재까지 XML, RDF 인코딩이 이루어진 상태이며 이를 토픽맵 기반으로 확장하여 토픽맵의 활용성을 증대하기 위한 목적으로 진행되고

표 1 토픽맵 관련 솔루션

솔루션	개발사	라이선스	플랫폼
Ontopia Knowledge Suite	Ontopia	상용	Java
Knowledge Management Suite	Empolis	상용	Java
Intelligent Topic Manager	Mondeca	상용	Java
TMCORE05	Networked Planet	상용	.Net
L4 Tool Suite	Moresophy	상용	Java
TM4J	TM4J Project	오픈소스	Java
TMAPI	SourceForge	오픈소스	독립적
TM4JScript	TM4JScript Project	오픈소스	JavaScript
Perl XTM	Bond Univ.	오픈소스	Perl
ZTM (Zope Topic Maps)	SourceForge	오픈소스	Zope

있다.

4. 토픽맵 활용 현황

토픽맵과 같은 국제표준규격을 활용하여 디지털 콘텐츠 정보를 구성하면 상호운용성과 재사용성의 증가 및 정보 간의 연계를 통하여 해당 기관뿐만 아니라 국가적으로도 통합 검색 및 서비스 수준의 비약적 향상을 가져올 수 있다. 노르웨이는 토픽맵을 활용한 국가적 디지털 콘텐츠의 연계와 구성을 위한 국가 온톨로지 구축에 선도적 역할을 하고 있다. 과학기술분야 연구 정보를 토픽맵을 토대로 구성하여 국가 지식 포털(www.forsking.no)을 구축하여 논문뿐만 아니라 논문에 포함된 이미지까지도 토픽맵의 토픽으로 표현하여 모든 정보를 의미적 연계를 구성하였다. 이외에도 다음과 같은 토픽맵 기반 연계 정보 서비스를 개발하여 제공하고 있다.

정보기술 교육	ww.itu.no
소비자 정보	www.fobrukerportalen.no
음식 정보	www.matportalen.no
부동산 정보	www.avhending.no
정당정보	www.hoyre.no
국가교육포털	www.utdanning.no
국가문화 포털	www.kulturnett.no
소비자 포털	forbrukerportalen.no

미국의 Department of Energy는 여러 핵에너지 생산 연구소의 극비급 자료의 일관적 분류 및 연계 체계 유지와 철저한 보안 관리를 위하여 토픽맵 기반 온톨로지를 활용하여 지식포털을 구축하였으며 Internal Revenue Service는 사용자 편의성의 제고를 목적으로 온톨로지 기반 지식 포털을 지향하고 있다. 이 밖에도 독일에서는 토픽맵 기반 지능형 온톨로지를 구축하여 정부 기관들과 회사들의 체계적 디지털 콘텐츠를 비

롯한 지식 자원의 효율적 관리를 지원하고 있다.

국내에서도 특정 분야 정보의 시범적 지식화 단계를 넘어 도서관 정보 서비스 분야와 공공 지식 자원화 분야에서 토픽맵을 활용하기 위한 노력이 활발히 이루어지고 있으며 토픽맵 표준 개발에도 선도적인 역할을 담당하고 있다.

토픽맵을 기반으로 하는 디지털 콘텐츠 메타정보나 응용 프로그램의 효율적 구축을 지원하는 다양한 솔루션이 존재하며 표 1은 그 중의 일부를 보여주고 있다.

5. 결론

토픽맵은 기본적으로 지식 표현을 위한 주제(subject) 중심 정보 연계 모델을 제공한다. 이러한 기능은 시맨틱웹을 비롯한 여러 가지 표현법에 의해서도 가능하지만 토픽맵(Topic Maps)은 국제표준화기구(ISO) 표준으로서 영속성과 판매자 독립적인 상호호환성 제공에 특별한 의미를 갖는다. 토픽맵의 표준적 정보 연계 모델은 디지털 콘텐츠를 비롯한 정보의 의미기반 색인화, 정보 통합, 지식 관리 차원에서 가치를 발휘한다.

디지털 콘텐츠 시스템을 온톨로지 기반으로 구축하면 하나의 주제로부터 이와 연관된 콘텐츠들에 손쉽게 접근할 수 있게 되고 새로운 요구 사항을 효과적으로 반영할 수 있는 융통성 있는 스키마를 활용할 수 있기 때문에 정보 처리 능력의 비약적 증대를 기대할 수 있다.

국내외에서 끊임없이 창출되는 디지털 콘텐츠와 지식정보가 토픽맵과 같은 국제표준 규격에 의해 상호호환적으로 서로 연계되면 지식 및 정보의 공유와 재활용이 활성화되고 축적된 지식정보를 바탕으로 새로운 지식을 창출할 수 있는 선 순환적 국가적 지식체계의 확립과 이로 인한 국제 경쟁력의 향상을 기대할 수 있게 된다.

참고문헌

- [1] Lutz Maicher and Jack Park, Charting the Topic Maps Research and Applications Landscape, Springer, 2005.
- [2] Jack Park and Sam Hunting, XML Topic Maps: Creating and Using Topic Maps for the Web, Addison-Wesley, 2003.
- [3] ISO/IEC JTC1/SC34 Document Description and Processing Languages, <http://www.itscj.ipsj.or.jp/sc34/ja>
- [4] SKIM 지식 서비스, <http://www.inek.co.kr>
- [5] 이재호, 시맨틱 웹의 온톨로지 언어, 정보과학회지 제21권 제3호, pp. 18~27, 2003.
- [6] 오삼균, 온톨로지 언어의 비교 연구 - W3C OWL과 ISO 토픽맵을 중심으로, 한국비블리아 발표논문집 제11집, pp. 77~110, 2004.



이재호

1985 서울대학교 계산통계학과(학사)

1987 서울대학교 계산통계학과(석사)

1997 University of Michigan(박사)

1998~현재 서울시립대학교 전자전기컴퓨터공학부 부교수

관심분야: 에이전트 기반 시스템, 인공지능, 온톨로지 기반 시스템, 지능형 로봇 시스템

관련활동: JTC1/SC34 국내위원회 위원장, JTC1/SC34 WG3 13250-4, 13250-7 Editor

E-mail: jaeho@uos.ac.kr

소프트웨어공학 포커스센터 정립기념 해외학회 초청 세미나

- 일 자 : 2008년 6월 25일~26일
- 장 소 : 서강대학교
- 주 관 : 소프트웨어공학소사이어티
- 문 의 : 박수용 교수(서강대) 02-705-8928