

차세대 지식공유 기술 토픽맵 최신동향



비즈니스표준과장 박인수
(02) 509-7271 ispark@mocie.go.kr

서론

최근 IT업계에서는 시맨틱 웹(Semantic Web), 온톨로지(Ontology), 토픽맵(Topic Map), WEB 2.0, 태그(Tag) 등과 같은 용어들이 화두에 올라와 있다. 2005년 인터넷 검색 업체인 구글(Google)의 시가총액이 2005년 말 1230억 달러(123조원)를 넘어시면서 구글은 세계적으로 주목받는 회사가 되었고, IT업계에서는 구글과 같은 최신 지식공유기술을 구현하기 위하여 WEB 2.0이란 기술적 트렌드를 이용하려고 노력하고 있다.

현재 우리는 예전에는 쉽게 얻을 수 없었던 지식정보를 인터넷이란 매체를 이용하여 쉽게 접할 수 있게 되었지만, 반대로 검증되지 않은 많은 정보들이 인터넷상에 혼재되면서 보다 효과적인 정보검색을 위해서 구글과 같은 우수한 검색엔진을 이용하거나 네이버의 지식인과 같은 지식공유 환경을 이용하고 있다. 이러

한 최신 검색기술로도 인간의 지식 욕구를 채우는데 한계성을 느낀 지식 사용자나 엔지니어들은 좀 더 나은 새로운 기술을 찾게 되었고 이에 따라 시맨틱웹과 같은 새로운 기술 트렌드가 탄생하게 된 것이다.

본고에서는 지식공유에 대한 현재의 일반적인 문제점들을 해결하고자 하는 기술이자 표준인 토픽맵에 대한 이해를 돕기 위하여 관련기술들의 역사적 흐름을 살펴보고 시맨틱웹 구현의 핵심요소 기술인 토픽맵에 대하여 소개하고자 한다.

지식공유 기술의 진화 단계

1969년 9월 2일 미국 로스앤젤레스의 캘리포니아 대학과 스탠퍼드대학의 컴퓨터가 최초로 연결되어 인터넷이 시작된 이후, 80년대 말 미국립과학재단의 NSFNET을 거쳐 90년대 초 월드와이드웹(WWW)이 탄생하고 상업망이 늘어나면서 인터넷은 폭발적인



발진을 거듭했다.

초창기 미 국방성 관계자 일부와 대학 연구소 등이 주로 군사적 목적과 연구 목적으로 이용하던 인터넷은 오늘에 이르러 이미 미국의 테두리를 벗어났고, 전세계의 모든 직장과 학교와 가정, 그리고 공공기관에서 접근 가능한 통신망이 되었다.

우리나라의 경우는 특히 인터넷 대중화 정도가 빨라 외출 중에도 남는 시간에 인터넷을 활용하고 이동 중에도 인터넷에 접근이 가능하게 되었고, 인터넷은 대중들의 기호나 정보 수집 차원을 벗어나 정치, 경제, 문화, 사회의 모든 영역에 많은 변화를 일으키고 있다. 이제 부연가 새로운 일을 하고자 한다면 누구나 인터넷이라는 요소를 고려하지 않을 수 없는 세상이 되었다.

1. 월드와이드웹(WWW) 탄생

현재 인터넷 기술의 바탕으로 있는 월드와이드웹은 스위스 제네바의 유럽입자물리연구소(CERN) 컴퓨터 정보 수집 및 제어 관계 연구원이던 영국 출신의 팀 버너스 리(Tim Berners Lee)가

1969 ~	인터넷의 시작
1990 ~	월드와이드웹 탄생
1992 ~	모자이크 인터넷 브라우저 탄생
1994 ~	Yahoo 검색엔진 등장
1998 ~	Google 검색엔진 등장
2000 ~	시맨틱웹의 구조 발표
2000 ~	RDF Schema 제정
2001 ~	도목렘 표준 제정
2004 ~	Web 2.0 용어 등장

지식공유 기술 진화단계

1990년 인터넷을 편리하게 사용할 수 있는 '하이퍼텍스트(hypertext)' 프로젝트를 제안하고 '웹(WWW)'의 개념을 고안함으로써 탄생하게 되었다.

이것이 오늘날의 인터넷 주소체계인 URL과 웹페이지 언어인 HTML이다. 원래 CERN은 유럽과 세계에

많은 회원을 가지고 있었는데, 서로 멀리 떨어져 있는 연구자들끼리 연구에 대한 여러 정보를 효과적으로 공유하기 어려웠기에 이를 해결하기 위한 목적으로 그는 하이퍼텍스트를 이용한 정보 전달 방법을 제안한 것이다.

월드와이드웹(WWW)의 이름에 대해서도 「그물(the Mesh)」과 「정보의 보고(Mine of Information)」등을 검토한 끝에 인터넷이 지구촌을 거미줄(Web)처럼 둘러싸고 있다는 뜻에서 현재 널리 사용되고 있는 「웹(WWW)」으로 결정했다.

이후 1990년 말에 NeXT라는 컴퓨터에서 동작하는 최초의 WWW 소프트웨어가 발표되었다. 이 소프트웨어는 컴퓨터 전문가들만이 사용할 수 있는 특수한 명령어에 의해 이용되고 있었던 인터넷상에서 하이퍼텍스트로 된 정보를 전송하고 보여주는 기능을 새롭게 선보였다. 그리고 이 소프트웨어를 이용한 시연회가 CERN 위원회에서 개최한 세미나와 1991년도에 개최된 하이퍼텍스트 학술대회에서 발표되었다.

2. 검색엔진의 등장

이후 웹(WWW)은 92년 미국의 슈퍼컴퓨팅센터(NCSA: National Center for Supercomputing Applications)에서 '모자이크(Mosaic)'이라는 브라우저를 만들어 무료로 배포하면서 인터넷환경의 대표적 서비스로 자리매김하게 되었다.

월드와이드 웹 탄생 이후, 각각의 사용자들이 자신의 홈페이지를 구축하여 다양한 정보를 제공하면서 전세계 인터넷 정보량은 기하급수적으로 많아지게 되었고 최근 들어서면서 매년 두 배의 양적 팽창이 가속화되고 있는 실정이다. 이러한 거대규모의 정보와 자료들 중에서 꼭 필요한 정보를 보다 빠르고 쉽게 찾아내기 위해 등장한 도구가 바로 검색 엔진인 것이다.

검색엔진은 로봇 에이전시(Robot Agency)라는 프로그램이 정보를 제공해 주는 모든 서버들을 찾아다



니면서 그 안의 여러가지 정보를 수집·분류하게 되며, 이렇게 수집된 정보중에서 사용자가 요청한 키워드와 관련된 정보를 추출하여 보여주는 것이다. 즉, 이와 같은 역할을 수행하는 프로그램을 가지고 있는 사이트 자체를 검색엔진이라 부르기도 한다. 1990년대 중반은 웹의 폭발적인 성장기였던 만큼 수많은 검색엔진들이 등장했다.

3. 야후(Yahoo)의 등장

한때 인터넷의 대명사였던 'Yahoo'는 1994년, 스탠포드의 대학원생이었던 David Filo와 Jerry Yang이 자신들이 가지고 있던 웹 주소들을 검색 가능한 디렉터리 형태로 정리하여 만들어 냈다.

현재 야후는 여러 성능 좋은 검색엔진들을 사들여 검색포털 서비스를 제공하고 있으나, 아직까지도 사람이 직접 분류하는 디렉터리 서비스 수준을 벗어나지 못하고 있다. 1997년 이후 검색엔진의 지평은 조금씩 변하기 시작했다. 인터넷 정보량이 폭발적으로 증가하면서 알타비스타(Altavista)는 그 한계를 드러내기 시작했다. 하나의 질의어에 대해서 너무나 많은 수의 결과를 보여주기 때문에 사용자 입장에서는 오히려 검색결과로 제공된 수백 수천의 페이지들을 입입이 확인해야 하는 지루한 과정을 수반해야만 했다.

4. 구글(Google)의 출현

1998년, 혼란스럽던 검색엔진 분야에 새로운 강자가 등장했다. 스탠포드 대학의 선산학 전공 대학원생이었던 Larry Page와 Sergey Brin은 검색엔진이 보다 효율적으로 작동하기 위해서는 새로운 개념이 필요하다고 생각했다.

'Yahoo'와 같이 사람이 직접 유지·관리하는 목록은 보편적 주제를 효과적으로 검색할 수 있다는 장점이 있지만, 주관적이기 쉽고 구축·유지비용이 매우

높은데다가, 실시간 수정·보완이 어려워 정보증가속도를 따라갈 수 없으며 다양하게 조합되는 특이한 주제를 수용할 수 없다는 단점이 있다.

한편 키워드 매칭(Keyword Matching)을 기반으로 하는 알타비스타(Altavista)와 같은 자동화된 검색엔진은 대체적으로 검색결과가 너무 많고 정확도가 낮은 문제가 있었다. 이러한 문제점을 파악한 구글은 검색결과와 정확도를 높이기 위해서 웹에 민주주의적인 방식을 도입했다. 해당 페이지의 중요도를 파악하기 위해 다른 웹페이지에서 해당 페이지를 가리키는 인바운드 링크(Inbound Link)의 수를 세고 이를 바탕으로 설정된 페이지 랭크라는 개념을 도입한 것이다. 생물학에 관한 중요한 정보가 있는 페이지라면 생물학에 관련된 다른 웹페이지에서 그 웹페이지를 링크로 연결하는 확률이 높을 것이라고 본 것이다. 마치 과학 연구에서 중요한 논문일수록 인용하는 횟수가 높은 것처럼 말이다. 이들은 이 아이디어를 야후의 창업자 중 한명인 Filo에게 제안했으나, 구체적이지 않다는 이유로 거절당했다.

그러나 독자적인 회사를 설립하여 서비스를 시작한 이들의 아이디어는 적중했고 기존의 검색엔진들이 만들어낸 수백 페이지의 '쓰레기 검색 결과(Junk results)'에 시달리던 사람들은 '구글(google)'에서 불과 1-2페이지 안에 자신이 원하는 정보를 찾게 되자 환호성을 질렀다. 이후 7년 이상 지난 지금까지 구글은 검색엔진의 절대강자로 군림하고 있다.

5. 검색엔진의 한계

구글의 CEO인 에릭슈미트는 캘리포니아에서 열린 실리콘밸리 컨퍼런스에서 다음과 같은 얘기를 하였다. "검색엔진의 가장 중요한 요소는 보다 정확한 검색결과를 찾아주는 것에서 개인화로 이동하고 있다." 에릭슈미트의 얘기에서도 표현되어 있지만 구글은 개인화 기술에 깊은 관심을 가지고 있으며, 최근에 개인



별 맞춤형 검색기술을 개발해 온 칼틱스(Katix)란 회사를 인수하기도 하였다. 주목할 만한 사실은 '정확한 검색결과'라는 표현에 대한 부정적인 견해이다. 정확하다는 사실은 주관적인 판단이 필요하다. 즉 절대적인 정확함이란 것은 존재하지 않는다.

왜냐하면 동일한 키워드를 입력하더라도 사람마다 원하는 정보는 모두 다르기 때문이다. 검색을 하는 이유는 정답을 찾기 위해서가 아니라 정보를 찾기 위한 용도가 더 크다는 사실이 중요하다. 현재까지 개발되어 서비스되고 있는 검색엔진들은 대부분 확률에 기반을 두어 만들어졌다. 키워드 반복 횟수라던가 자신을 링크하고 있는 페이지 개수를 분석하는 방식은 모두 확률에 기반을 둔 이론들이 뒷받침된 기술들이라고 할 수 있다. 하지만 확률에 기반을 둔 기술을 가지고는 개인화된 검색 결과를 제공해 줄 수가 없다.

이러한 기존기술의 문제점들을 혁신적으로 개선할 수 있는 방법이 '의미기반 검색 방식'이다. 의미기반 검색기법은 최근에는 '시맨틱웹'이란 표현으로 많이 사용되고 있다.

지식공유 기술의 새로운 패러다임 시맨틱웹

시맨틱웹은 문서의 의미 정보와 문서간의 연관정보를 표현하는 RDF와 제약조건 그리고 로직을 표현하도록 확장된 언어인 온톨로지를 사용해 표준용어를 정의하고 용어들 간의 계층구조와 연관관계를 정하여 지식을 표현하는 기술이다.

시맨틱웹은 월드와이드웹(WWW)의 아버지이자 창조자인 팀 버너스리의 두번째 작품이다. 현재의 HTML로 이루어진 웹 개발 후 "나름에 들지 않는다"며 새로운 프로젝트에 돌입. 탄생된 것이 시맨틱웹이다. W3C(World Wide Web Consortium)에서 제정한 HTML(Hypertext Markup Language), HTTP(Hyper Text Transfer Protocol)와 URI(Universal Resource Identifier)는 웹을 구성하는 핵심 표준으

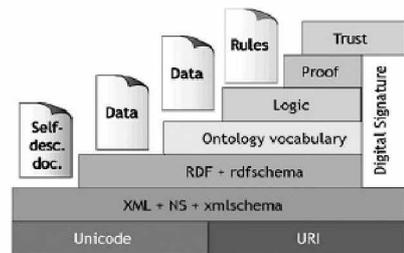
로, 팀 버너스 리(Tim Berners Lee)에 의해 처음으로 설계되었다. 이러한 웹 표준으로 인터넷이 연결된 어느 곳에서나 정보를 자유로이 교환할 수 있는 정보유통, 공유 및 활용이 가능한 가상공간이 생성되었다.

그러나 팀 버너스 리와 W3C는 웹을 한 단계 더 발전시키기 위해 장대한 목표를 제시하였다. 그 목표들 중의 하나가 시맨틱웹이다. 시맨틱웹은 컴퓨터 스스로가 웹에 연결된 정보의 의미를 인식하고 사용자가 필요로 하는 정보를 검색하며 검색된 정보에서 지식을 유추할 수 있는 기능을 제공한다.

즉, 시맨틱웹은 사람이 웹 정보에 대하여 의미를 파악하고 의미에 따라 필요한 정보를 선택하듯이 컴퓨터가 웹 정보의 의미를 이해하고 의미에 따라 선택적으로 정보를 획득, 통합, 가공할 수 있는 웹 환경을 제공할 수 있다.

1. 시맨틱웹의 구조

시맨틱웹을 실현하기 위한 다양한 접근방법이 현재 제시되어 있다. 하지만 HTML을 기반으로 한 현재의 웹을 개선하는 기본 취지에서 보면 시맨틱웹을 달성하기 위해 웹 프로토콜과 같은 하위 레벨의 개념을 정의하고 이 하위레벨을 이용하여 다음 레벨의 개념을 정의하는 계층구조(layered structure)를 설정하는 것이 일반적인 연구 방향이다. 시맨틱웹의 계층구조는 그림과 같다. 이 계층구조에서 보면 가장 하위레벨의 웹 프로토콜에서 자원을 지칭하기 위한 주소지정 방법인 URI가 밑받침되고 이를 기반으로 기본적인 표현레벨에서 XML과





NameSpace. 문법적 기반을 제공하는 RDF와 RDF스키마, 지식표현을 위한 온톨로지 순서로 연구가 진행되고 있으며, 그 위의 계층인 Logic에 대해서는 인공지능의 추론연구를 밑받침으로 일부 연구가 시작되었다. 또한 보다 더 상위 계층인 Proof와 Trust는 시맨틱웹 정보의 신뢰성과 보안에 관한 내용으로서 아직 개념정도만 이야기되고 있으며 차후 연구과제로 제시되고 있다. 시맨틱웹은 현재 자동화된 웹 에이전트, 지능화된 검색엔진, 실시간 그리드 컴퓨팅 등 광범위한 분야에서 그 응용 가능성이 모색되고 있다.

2. 시맨틱웹의 목표

시맨틱웹(Semantic Web)의 궁극적 목표는 컴퓨터도 이해할 수 있는 지식의 원천으로서의 웹을 만드는 것인데, HTML 형태의 문서들로 이뤄진 현재의 웹은 사람에게 정보를 주는 역할은 하고 있지만 컴퓨터 프로그램이 각 문서의 내용을 정확히 파악할 수 없다는 문제의식에서 출발한다. 시맨틱웹의 중심에 확장성표기언어(XML)를 기반으로 하는 RDF(Resource Description Framework)와 DAML-OIL을 발전시킨 OWL(Ontology Web Language)이 있는 것은 이러한 이유다.

시맨틱웹은 현재의 보여주는(display) 방식인 HTML 언어의 한계를 해결하기 위해 고안된 XML언어 기반의 차세대 웹이다. 지금까지 웹 정보가 사람을 위한 것이었다면 시맨틱웹 정보는 컴퓨터를 위한 정보로 변해야 한다. 즉 웹에서 폭증하는 정보를 이제 더 이상 사람이 처리할 수 없기 때문에 컴퓨터(기계 또는 에이전트)에 의미를 이해할 수 있는 기능을 부여하여 자동화하자는 것이다. 시맨틱웹을 가능하게 하는 기술로서 온톨로지와 온톨로지 표현언어(ontology representing language)가 있다.

온톨로지는 특정 분야에서 서로 달리 사용하는 정보 간 의미적 상호운용성을 극대화하기 위한 용어·

개체·개념 등의 표준화, 이들 간 의미적 관계, 사용규칙 및 제한점 등을 정의한 개체·개념의 집합이라 할 수 있다. 온톨로지 표현언어는 의미적 관계로 표현된 용어·개체 등의 집합을 모델링하여 컴퓨터가 이해할 수 있도록 표현해 주는 도구로서, ISO 표준인 토픽맵과 W3C 표준인 RDF·RDF(S)·OWL 등이 있다. 현재는 W3C 표준 중심으로 활발한 연구와 시장이 형성되고 있다. 그러나 이 두 표준은 경쟁 관계라기보다는 상호보완 관계로 보는 게 적절할 것 같다

지식공유를 위한 온톨로지 토픽맵

본고는 시맨틱웹을 구현하기 위한 온톨로지 표현언어 중에서 토픽맵 표준과 토픽맵의 우수성을 알리고 토픽맵 구현 사례 등 최신 기술 동향을 소개한다.

토픽맵은 1991년에 소프트웨어 문서를 위해 SGML DTD 표준을 개발하는 일을 시작했던 Davenport 그룹을 시작으로 길고 복잡하게 얽힌 역사를 갖고 있다. 이 그룹은 상당히 빠르게 CAPE(Conventions for the Application of Hy Time)이라고 불리는 과생물을 분리 신설했다. 그들의 작업중에 하나는 컴퓨터화된 책의 뒤 색인을 위한 어플리케이션을 설계하는 것이었다. 자동적으로 그것들을 통합하는 것이 가능해야 했기 때문에, 이 색인들은 소설 형태를 갖도록 하려고 했었다. 이 어플리케이션에 숨겨진 아이디어들이 결국은 토픽맵이 된 것이었다. 토픽맵은 많은 어플리케이션을 가지지만, 주된 어플리케이션들 중에 하나는 정보의 발견가능성 문제를 해결하는 것이다. 즉, 어떻게 거대 정보군 속에서 원하는 정보를 발견할 수 있는가를 해결하는 것이다.

토픽맵은 또한 지식관리를 위해, 웹포탈개발, 콘텐츠관리, 그리고 엔터프라이즈 어플리케이션 통합(EAI)을 위해 사용될 수 있는 기술로서, 시맨틱웹을 가능하게 하는 기술로서 평가된다. 아래는 여러 기술들의 사용 영역을 비교하는 표이다.



Area	TM(Topic Map)	RDF	DAML	OIL
Findability	Yes	Yes	Yes	Yes
Portals	Yes	Yes	Yes	Yes
Content management	Yes	Yes	Yes	Yes
EAI	Yes	-	-	-
E-commerce	Yes	Yes	Yes	Yes
KM	Yes	Yes	Yes	Yes
Semantic web	Yes	Yes	Yes	Yes

각 방법은 제안자가 의도한 대로, 기술의 응용 영역이 동일하다는 것을 분명히 알 수 있다. 유일한 차이는 RDF, DAML,과 OIL은 EAI 기술로서 제출되지 않았다는 것이다.

그러나, 이 사실은 그 기술들이 EAI 목적을 위해 사용될 수 없다는 것을 의미하는 것은 아니다. 요컨대, 이 영역중 어느 것에서 정보를 교환하고자 하고 이 표준들중 하나를 사용하는 사람은 교환파트너가 다른 기술중 하나를 선택했다는 것을 발견하게 된다.

1. 토픽맵 구조

토픽맵 모델의 핵심요소는 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 남겨있는 자원의 주소(URL로 표현된다)나 지식 데이터 자체를 가리킨다

그리고 토픽들 사이에는 토픽타입과 토픽간의 상하관계와 토픽들 간의 연관성을 표현하는 어소시에이션

이 있다. 온톨로지를 구축하기 위해서는 지식을 설명할 수 있는 용어(term)를 정의하고 용어들 사이의 관계를 정의한다.

용어들 사이의 관계는 분류 체계를 표현하는 텍사노미(taxonomy)와 개념간의 연관관계를 표현하는 릴레이션(relation)으로 나뉜다. 온톨로지가 정의되면 용어를 토픽으로, 텍사노미를 토픽 타입과 토픽간의 관계로, 그리고 릴레이션을 토픽들 사이의 어소시에이션으로 매핑 하여 초기 토픽맵을 생성한다.

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

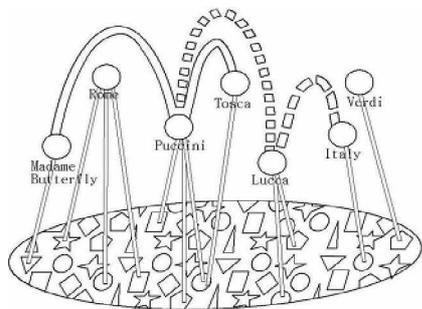
2. 토픽맵 표준

토픽맵은 정보 리소스들의 복잡한 구조에 대한 지식을 표현하고 이를 토픽과 그 관계들로 조직화되는 일종의 가상 맵이다. 기존의 온톨로지 언어들은 속성



과 값으로 표현되므로 온톨로지를 구축하기 위해서 기존 정보들을 변환하여야 하지만, 토픽맵은 위치 정보를 이용하여 형태를 변환하지 않고도 정보 리소스들을 통합할 수 있다. 또한, 토픽맵은 서로 다른 토픽맵을 쉽게 병합할 수 있으므로 많은 양의 이종 정보 리소스를 통합할 필요가 있을 때 매우 유용하다.

토픽맵의 표준으로는 2000년 1월 HyTM으로 표현된 ISO/IEC 13250 Topic Maps이 최초로 제안되었다. 그리고, 2001년 2월 XML 구문체계를 이용한 XML Topic Maps(XTM) 1.0 이 Topicmaps.org에서 제안되었는데, 10개월후 ISO/IEC 13250(HyTM)과 XTM 1.0 이 통합되어 ISO/IEC 13250:2002에서 공동(HyTM, NYM 1.0)으로 존재한다. 표준화된 토픽맵의 근간은 토픽(Topic), 연관관계(Association) 및 자원(Occurrence)으로 이루어져 있다.



예를 들어 오페라 토픽맵을 보면, 토픽은 "Tosca", "Madame Butterfly", "Rome",

"Italy", "Puccini", "Lucca" 등의 주제를 표현한다. Puccini는 Lucca에서 태어났고 Tosca를 작곡하였으므로, "Puccini"와 "Lucca" 토픽간에는 "born-in"이라는 연관관계가 존재하고, "Puccini"와 "Tosca" 토픽간에는 "Compose"라는 연관관계가 존재하게 된다. 이 토픽에 대한 자원은 <http://www.topic.net/opera.htm>, <file:///C:/topicmaps/opera/occurs/gallery.htm> 등의 주소 형태 또는 1924/1/29, 1858/12/22 등의 텍스트 데이터 형태를 갖는다. 토픽은 실제계의 주제를 기술하고, 주제는 "기계에 의해 주소화가 가능한 어떤 것"

이며, "사람"과 같이 주소화 할 수 없는 것 또는 "음악"과 같은 추상적 개념도 토픽이 될 수 있다. 다시 말해서 토픽은 사람, 엔터티, 개념, 의미 등 어떤 대상이라도 될 수 있다.

일반적으로 실제 세계에서 "thing"이라고 표현되는 것들을 의미하며, 토픽맵의 표준은 이를 서브젝트(subject)라고 정의한다. 즉, 특정 문서의 토픽은 그 문서의 작성자가 나타내고자 하는 주제를 표현할 수 있는 단어들로 구성된다. 연관관계는 토픽 집합들 간의 상호관계를 기술하고 토픽들 간의 "관계"를 명시하거나 또는 각 관계들 간에 작용하는 "역할"을 정의한다. 그리고, 연관관계는 계층적인 일반적 개념을 표현하는데 사용되거나 "회의", "미팅"과 같이 더 복잡한 일단의 관계를 표현할 수 있다. 연관관계 구성시 정보를 추가할 수 있다. 다시 말해서, 연관관계는 토픽 간의 상하 관계뿐만 아니라 다양한 의미적 관계성을 표현할 수 있게 한다. 연관된 토픽들은 연관관계의 "멤버"가 되며, 각각의 멤버는 관계 내에서 "역할"을 가진다.

자원은 토픽과 관계되는 세부 정보자원이다. 자원은 정보자원(또는 정보 그 자체)과 토픽 또는 정보자원간의 관계를 명시하는데 사용할 수 있고, 사람 자체를 표현할 수도 있으며 그 사람의 홈페이지 URL인 자원정보를 가질 수도 있다. 다시 말해서, 자원은 토픽과 정보 리소스들 간의 루트를 표현하는데, URI로 토픽맵 외부에 있는 리소스의 위치를 가리킬 수도 있고, 리소스가 가지는 텍스트 데이터로 직접 나타낼 수도 있다. 토픽이 참조할 수 있는 정보 리소스는 문서 파일, 이미지 파일, 비디오 파일, 데이터베이스내 특정 레코드, 텍스트 데이터 등 다양한 형태를 가질 수 있다. 또한, 한 토픽에 대한 모든 자원은 토픽에 의해 표현되는 모든 속성을 공유한다.

3. 토픽맵 적용 사례

토픽맵을 이용하여 다양한 시스템을 제작 또는 구



축할 수 있는데, 대표적으로는 인간적 사고 흐름을 반영한 직관적 네비게이션 인터페이스 제공 및 지식정보간의 다양한 의미적 연관정보를 제공하는 “의미기반 포털시스템 <예: 노르웨이 포털 사이트(청소년 기술정보 포털, 소비자정보 포털, 농업부 포털, 보수경당 포털, 사법부 포털, 문화재청 포털, 세무성 포털, 수상관지 포털, 통계국 포털, 원자력 프로젝트 포털, 문화부 포털), 미국의 국제청 세금 정보제공 포털>”, 지식자원들을 상호 연관성에 따라 연결 및 조직하여 지식구조를 기술할 수 있도록 토픽맵 모델을 기반으로 하는 “지식관리시스템 <예: empo.is의 e:KMS, mondeca의 ITM, Bravo 지식관리Tool, ISOGEN International의 SemanText, Paderborn의 K-discovery>”, 토픽맵 애플리케이션을 이용하여 콘텐츠들의 연관관계를 설정하고 구조화하여 콘텐츠의 관리를 편리하게 하는 기능 제공 “콘텐츠 관리시스템 <예: 미국 에너지부 기밀문서분류 가이드 및 관리, Opensource ZTM, Tu-berlin의 ivS>”, 토픽맵의 강력한 정보 통합 기능을 이용하여 분리된 다양한 종류의 정보 저장소로부터 다양한 조직 구성원들에 의해 동시에 여러 개의 토픽맵들이 구축될 수 있으며, 구축된 토픽맵들은 유기적으로 연결되어 조직 전체의 맥락에서 정보를 파악할 수 있도록 하는 하나의 통합된 뷰를 제공하는 “애플리케이션 통합(EAI) <예: StarTeam 소프트웨어 개발/보수를 위한 공동작업 환경>”, 학생들이 배운 개념과 주제를 기술하고 서로 연결하여 개개인의 토픽맵을 작성하면, 선생님은 이를 브라우징하여 학생 개개인에 대한 학습 이해도 및 성취도, 학습 유형 등을 평가해 주는 “전자학습(e-Learning) <예: 독일 교재 검색 애플리케이션, 노르웨

이 BrainBank 학습지원 시스템>”, 토픽맵 기반의 “웹 애플리케이션 개발 <예: 노르웨이 ITU 교육프로젝트용 웹사이트, Patrimoine 금융관계 문서용 인트라넷, 남아프리카CSIR iWorks Ideabank 연구소의 지식아이디어 공유, Google의 하이와이어 프레스>” 등이 있다.

맺음말

시맨틱 웹의 응용 분야는 매우 다양하다. 이질적이고 분산돼 있는 정보들을 통합해야 하는 응용분야, 예를 들어 각 연구기관에 산재해 있는 국가통합지식정보, 기업의 EAI나 포털의 정보 통합, 전자정부, 전자상거래 등 모든 분야에 활용되고 있다. 국내에서는 지식정보를 취급하는 진문기관 및 전자정부 차원에서 시맨틱 웹 기반의 정보제공 서비스를 준비하고 있으며, 그 외 일반기업 차원에서도 토픽맵 같은 시맨틱 웹 구조의 솔루션을 개발하기 위하여 박차를 가하고 있다. 국내 기업중 인프로보마는 이미 토픽맵 에디터와 웹프레임워크를 시제품 단계까지 완성도를 보이고 있으며, 자체 토픽맵기반의 여행허브 포털사이트를 5월 중 오픈 목표로 구축중에 있다.

IT의 핵심은 정보를 어떻게 활용하느냐에 달려 있다. 인간은 데이터를 분석해 정보를 만들고, 정보의 이해를 통해 지식을 생산하게 된다. 이 과정을 자동화하기 위한 기술이 시맨틱웹이고 토픽맵과 같은 온톨로지다. 앞으로 웹은 시맨틱웹으로 진화할 것이고, 이 진화의 핵심은 온톨로지를 포함한 시맨틱웹 기술이 남당할 것이다 **표준**