



e-Science를 위한 지식관리 기반구조

목 차

1. 서 론
2. 관련 용어
3. 지식관리시스템 기반구조의 기능적 요구사항
4. 지식관리시스템 기반구조의 기술요소
5. 요약 및 향후 연구

최동훈 · 이윤준

(한국과학기술정보연구원 · 한국과학기술원)

1. 서 론

e-Science 환경은 거대 과학 연구에 필요한 슈퍼 컴퓨팅, 데이터, 분석 소프트웨어 등의 자원을 연구 커뮤니티에게 제공한다. 거대 과학 연구는 계산 집중적이므로, 계산 자원 중에서 적합한 자원을 발견하고 할당하기 위해 이에 관한 지식을 필요로 한다. 또한, 거대 과학에서 새로운 과학적 발견은 대량의 데이터에 대한 복잡하고 창의적인 분석 노력에 의존적이다. 연구에 필요한 데이터는 서로 이질적이고 다양한 데이터 소스에 분산되어 있기 때문에 이들 데이터의 수집은 많은 노력을 필요로 한다. 아울러 데이터에 대한 분석 소프트웨어도 연구자에게는 매우 난해하기 때문에 연구자는 분석 소프트웨어에 익숙해지기까지 많은 노력을 투자해야 한다. 이러한 노력의 낭비를 줄이고 연구자에게 연구에 집중할 수 있는 e-Science 환경을 제공하려면, 데이터 소스에 대한 연구자 관점의 의미적 메타데이터와 함께, 분석 소프트웨어에 대한 사용법, 실험자의 분석 사례와 같은 경험적 지식을 연구자에게 제공하는 것이 중요하다. 이를 통해 데이터에 대한 의

미나 분석 소프트웨어에 대한 지식이 없는 연구자도 숙달된 연구자만큼 연구의 생산성을 높일 수 있다.

영국의 e-Science 환경 중의 하나인 GEODISE의 경우, 엔지니어를 위한 MATLAB 기반의 지식 포털을 제공하고 있다[1]. GEODISE는 엔지니어가 그리드 환경에서 엔지니어링 설계의 최적화 문제를 해결할 수 있도록, 각 노드에서 독자적으로 운용하고 있는 최적화 도구, 기하학적 모형화(modeling) 및 격자(mesh) 생성 패키지, 분석(analysis) 코드, 계산 및 데이터 등과 같은 그리드 자원의 공동 활용 체제를 엔지니어에게 제공한다. 설계 최적화 문제는 데이터 및 계산 집중적이므로, 여러 데이터 및 계산 자원 중에서 적합한 자원을 발견하고 할당하기 위해 이들 자원에 관한 지식을 필요로 한다. 아울러 설계 최적화 문제는 복잡한 문제 해결 절차의 작성과 실행 환경의 설정을 요구하기 때문에 효율적으로 최적화 문제를 해결하려면, 문제 해결 절차와 실행 환경 설정에 관한 전문 지식과 경험이 필요하다. 이러한 이유로, GEODISE는 그리드 상에서

지식 관리를 위한 전형적인 어플리케이션으로 설계 최적화 문제를 선정하여 지식 포털을 개발하였다.

GEODISE 지식 포털은 MatLab 함수 및 방법에 대한 엔지니어의 문제 해결 경험(특정 함수 및 방법의 적용 사례와 입출력 데이터 등을 포함하는 로그 파일), MatLab 사용자 매뉴얼 등을 지식으로 관리하고 있으며, 이러한 지식에 대한 온톨로지(MatLab 용어 집합에 관한 택소노미와 입출력 조건을 비롯한 관련 지식의 개념화)를 구축하고 온톨로지와 지식 간의 주석을 첨부하여 양방향의 지식 탐색 기능을 제공한다. 이를 통해 초보 엔지니어는 매뉴얼을 공부하는 도중에 함수나 방법을 연습할 수 있고 전문 엔지니어의 경험 등을 참고할 수 있으며, 함수 사용 도중에 매뉴얼 설명과 전문 엔지니어의 문제 해결 경험을 참고할 수 있다. 또한 초보 엔지니어가 문제 해결 절차를 작성하거나 실행환경을 설정하고자 할 때, 온톨로지에 명시된 전문가 수준의 지식에 근거한 가이드라인을 초보 엔지니어에게 제공하여 문제 해결의 효율을 높인다.

본 논문에서는 연구자에게 친숙한 e-Science 환경을 제공하기 위한 지식관리시스템의 기반구조에 대해 논의한다. 지식관리 시스템은 사용자에게 업무 활동에 필요한 지식을 공유 및 활용할 수 있는 체계적인 수단을 제공한다. e-Science를 위한 지식관리시스템은 여러 연구 커뮤니티에 공통적인 기능과 고유 기능으로 나눌 수 있다. 고유 기능은 각 커뮤니티에 고유한 지식관리 요구사항에 밀접하게 연관되어 있으며, 특정 커뮤니티의 지식관리시스템을 타 커뮤니티와 차별화할 수 있는 부분이다. 공통 기능은 지식의 생명주기에 따라 지식의 생성, 저장, 검색 기능을 제공하는 것으로, 대부분의 지식관리시스템에 공통적이다. 우리는 이와 같은 공통 부분을 지식관리시스템 기반구조라고 부른다.

본 논문은 연구자에게 친숙한 e-Science 환경

구축을 위해 필요한 지식관리시스템 기반구조를 소개한다. 2장에서는 본 논문의 이해를 돕기 위한 기본적인 관련 용어를 설명하고, 3장에서는 지식생명주기에 근거하여 지식관리시스템 기반구조의 기능적 요구사항을 도출한다. 4장에서는 이를 위한 기술 요소를 식별하고 서술하고, 5장에서 끝맺는다.

2. 관련 용어

지식이란 특정 도메인의 전문적 경험에 관한 기록이다. 기록은 외견상으로 테이블, 텍스트, 이미지, 비디오 등 멀티미디어 데이터처럼 보이지만, 내적으로는 데이터가 표현하고자 하는 고유의 의미(meaning)를 함축하고 있다. 이러한 의미는 데이터에 관한 메타데이터, 데이터와 데이터 간의 연관 관계(association), 데이터에 함축된 의미적 개념 및 개념 간의 관계(온톨로지), 도메인에 특정된 현실 세계의 용어 및 이들 간의 연관 관계(택소노미) 등을 포함한다.

메타데이터는 구문적 메타데이터와 의미적 메타데이터의 두 가지 유형이 있다[2]. 구문적 메타데이터는 데이터에 대해 문맥과 무관한 정보를 서술하고 있으며, 저자, 작성일, 제목, 파일 형식 등과 같이 데이터의 의미에 대한 어떠한 직관적인 실마리조차도 제공하지 않는다. 반면에 의미적 메타데이터는 데이터 대해 도메인에 특정된 정보를 서술한다.

온톨로지[3]는 특정 도메인 내에서 개념의 집합을 표현하고 개념과 개념 간의 관계를 표현하는 데이터 모형으로, 도메인 내에 존재하는 객체에 대한 추론을 하는데 사용된다. 온톨로지는 각 객체를 나타내는 인스턴스, 인스턴스의 집합 또는 유형을 나타내는 클래스(또는 개념), 객체의 특성을 서술하는 속성(솔릿), 속성에 대한 제한 조건을 나타내는 퍼셋, 인스턴스와 인스턴스 간에 그리고 클래스와 클래스 간에 존재하는 관계로 구성된다. 특정 도메인 내에서 개념 간의 계

층 구조를 텍소노미라고 부르며, 텍소노미는 온톨로지 중에서 가장 단순한 형태이다.

온톨로지는 데이터의 의미(meaning), 즉 데이터에 대한 의미적 메타데이터, 데이터와 데이터 간의 연관 관계(association), 데이터에 함축된 의미적 개념 및 개념 간의 관계(온톨로지), 도메인에 특정된 현실 세계의 용어 및 이들 간의 연관 관계(텍소노미)를 위한 표현 수단으로 사용된다.

지식 관리시스템은 지식의 생산, 저장 및 관리, 활용, 피드 백(수정 또는 폐기) 등의 전체 생명주기에 걸친 사용자의 모든 활동을 포괄적으로 지원해야 한다. 사용자 입장에서 보면 지식은 누군가의 산출물이며 자신의 업무 활동을 통해 새로운 지식을 창출하기 위한 소비의 대상이다. 지식을 매개로 하여 이들 사용자 간의 관계를 살펴보면, 이들은 유사한 목적의 연구 활동을 수행하면서 공통의 지식을 공유, 생산 및 소비하는 사용자 그룹을 형성한다. 지식은 사용자 그룹 내에서 기존의 지식을 기반으로 자가 발전하는 경향을 보이고 있기 때문에, 생명 주기에 걸쳐 지식을 관리하기 위해서는 지식 그 자체뿐만 아니라 사용자 그룹 내에서 지식의 생산 및 소비에 관한 행태 정보를 관리하는 것이 필수적이다.

지식 관리 시스템 기반 구조는 사용자의 지식 관리에 필요한 기능적(지식의 저작, 저장, 검색, 동시성 제어, 버전 관리 등) 요구사항을 지원하는 구성 요소로 구성되며, 기존의 상업용 데이터베이스 관리 시스템과 웹 콘텐츠 관리 시스템을 포함하지만, 데이터의 의미(meaning)를 관리하기 위한 온톨로지 관리 시스템, 의미 기반의 질의 처리를 위한 지식 중재자 시스템 등 새로운 기술적 요소를 포함한다.

3. 지식관리시스템 기반구조의 기능적 요구사항

지식 생명주기는 지식의 생산, 저장 및 관리, 확산, 활용, 피드 백(수정 및 폐기)에 의한 순환

사이클을 말한다. 지식의 생산은 업무 활동 중에 발생하는 지식의 소요 제거로부터 시작된다. 지식의 소요는 새로운 지식의 생산, 기존의 지식의 갱신 또는 폐기를 포함한다. 새로운 지식의 생산은 소요의 공식화, 내부의 학습 활동 및 외부로부터 수집 활동에 의한 지식의 획득, 획득된 지식의 평가 등 일련의 과정을 통해 조직 내의 도메인 전문가의 공동 저작에 의해 이루어진다. 획득된 지식의 평가는 지식의 효용성을 평가하는 것을 의미하며, 전문가 집단뿐만 아니라 업무 활동을 직접 수행하는 실무자가 함께 참여한다.

생산된 지식의 확산 및 활용을 극대화하기 위해 콘텐츠에 포함된 텍스트의 주제에 따라 지식을 분류하여 카테고리를 결정한다. 다른 데이터에 대한 연관성을 찾아내고, 텍스트, 이미지, 오디오 등 콘텐츠에 대한 의미를 메타 데이터로 표현하며, 콘텐츠에 포함된 개념을 분석하여 온톨로지를 갱신 또는 생성한다. 의미적 메타데이터 간의 의미적 혼란을 방지하기 위해 의미적 메타데이터를 정규화한다. 정규화는 동음어 및 동의어에 대해 통일된 단어를 적용하거나 단어 간의 매핑을 이용하여 의미적 혼란을 해소하는 것을 말한다. 이렇게 생성된 데이터에 대한 의미는 사용자 관점에 근거한 지식 모형에 따라 지식 저장소에 저장된다.

의미 기반의 지식 검색을 위해 의미적 메타데이터와 콘텐츠 간에 연결이 필요하며, 이러한 연결의 생성 과정을 주석 첨부라고 한다. 주석 첨부를 통해 콘텐츠와 의미 간의 연결이 되고 지식에 대한 사용자의 다양한 관점을 지원한다. 주석은 지식 모형에 따라 지식 저장소에 저장된다.

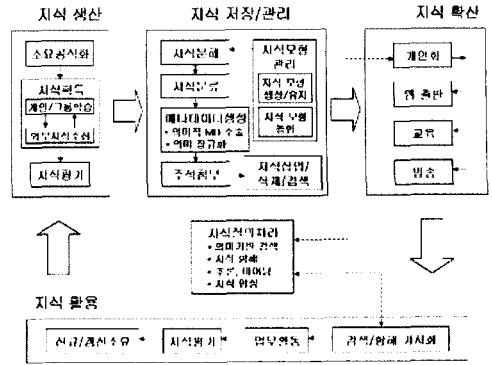
지식 모형은 콘텐츠의 저장 및 관리에 필요한 의미의 개념적 구조를 말한다. 콘텐츠의 의미는 메타데이터, 연관관계, 온톨로지, 텍소노미를 포함하며, 지식 모형은 이들 의미에 대한 사용자 관점의 외부 모형과 지식의 저장 관리에 사용되는 내부 모형을 포함한다. 내부 모형은 여러 외부 모

형을 하나로 통합하여 생성되는 것으로, 시스템 내부에는 실제로 존재하지만 관리자를 제외한 사용자에게는 제공되지 않는다. 반면 외부 모형은 지식을 활용하는 사용자의 관점에 따라 다양하기 때문에 여러 가지가 존재하며 사용자에게 제공된다. 지식 모형은 초기 생성 이후에도 지식의 저장, 갱신, 폐기에 따라 지속적으로 변경 및 진화된다. 따라서 이러한 지식 모형은 지속적으로 일관성 있게 관리되어야 하며, 다양한 지식 모형 간의 의미적 일관성을 유지하기 위해 이들 간의 매핑을 유지하는 것이 필요하다. 지식은 온톨로지에 의해 표현 가능하고 온톨로지는 고유의 모형을 갖기 때문에, 본 연구에서는 지식 모형을 온톨로지 모형과 상호 호환적인 의미로 사용한다.

지식의 업무 활용도를 높이려면, 사용자가 원하는 지식을 찾아 원하는 형태로 지식을 출판하여 활용할 수 있는 기능이 필요하다. 이를 위해 웹 브라우저를 통해 지식을 쉽게 탐색하고, 원하는 지식을 쉽게 합성하여 여러 사용자에게 출판하여 공유할 수 있어야 한다. 이를 위한 기능으로 개인화와 웹 출판이 있다. 그리고 사용자에게 새로운 지식에 대해 교육이나 방송을 통해 가이드를 제공하는 것도 지식 확산에 매우 중요하다.

지식질의처리 기능은 원하는 지식을 손쉽게 찾을 수 있는 기능을 제공한다. 사용자는 지식질의처리 기능을 통해 의미 기반의 검색을 수행하거나 지식 모형에 따라 항해하여 원하는 지식을 찾을 수 있다. 경우에 따라 지식 질의 처리는 연상 작용에 의한 추론이나 지식 마이닝을 포함하기도 한다. 사용자 간에 지식을 확산하는 방법으로 웹 출판 이외에도 조직 내의 방송이나 교육을 통한 방법도 있다.

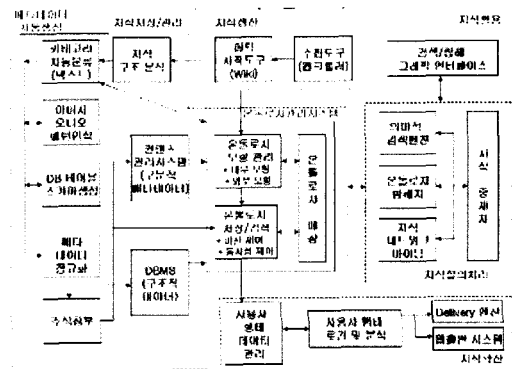
사용자는 업무 적용 경험을 바탕으로 지식을 평가할 수 있다. 지식의 평가는 기존 지식의 갱신이나 새로운 지식의 생산에 근거가 된다. 지식 생명주기에 따른 지식관리시스템 기반구조의 기능적 요구사항을 표현하면 (그림 1)과 같다.



(그림 1) 지식관리시스템 기반구조의 기능적 요구사항

4. 지식관리시스템 기반구조의 기술요소

이 절에서는 지식관리시스템 기반구조의 실현에 필요한 기술 요소를 기능적 요구사항으로부터 도출하여 서술한다. 기능적 요구사항에 근거하여 지식관리시스템 기반구조의 기술 요소를 식별하면 (그림 2)와 같다.



(그림 2) 지식관리시스템 기반 구조의 기술 요소

4.1 지식의 생산을 위한 기술 요소

지식 생산은 해당 분야의 여러 전문가의 공동 노력을 요구한다. 이들은 지식 생산에 필요한 정보 수집은 물론 공동 저작 및 토론을 거쳐 사용자의 소요 제기에 적합한 지식을 생산한다. 이를 위한 기술 요소로, 인터넷 상에서 필요한 정보를 수집할 수 있는 웹 크롤러, 공동 저작 및 토론 공간을 제공하는 위키와 같은 협력 저작 도구가 있다.

4.2 지식의 저장 및 관리를 위한 기술 요소

생산된 지식을 저장 관리하려면 주제에 따라 콘텐츠를 분류하고 이에 적합한 의미를 생성하여야 한다. 의미는 메타데이터를 비롯, 연관관계와 개념화 온톨로지를 포함하며, RDF 모형의 온톨로지로서 표현된다. 의미를 생성하기 위해, 지식의 구조 분석, 텍스트 부분에 대한 카테고리 자동분류 및 메타데이터 생성, 이미지 오디오에 대한 메타데이터 생성, 테이블 형태의 데이터 부분에 대한 스키마 정의 등이 필요하다. 또한, 의미를 온톨로지로서 표현하고 관리하려면, 온톨로지 저장 및 검색뿐만 아니라, 온톨로지 모형 관리, 온톨로지 간의 매핑을 지원하는 온톨로지 관리 시스템이 필요하다.

4.2.1 지식 구조 분석

멀티미디어 콘텐츠를 포함하는 지식을 효율적으로 저장 관리하기 위해 각 데이터 유형에 적합한 분석 기술을 적용하여 메타데이터를 생성한다. 이를 위해 콘텐츠의 구조를 분석하여 각 유형의 데이터를 분리해 내는 것이 선행되어야 한다. 지식 구조 분석은 멀티미디어 콘텐츠를 분석하여 구조를 찾아내고 텍스트, 이미지, 오디오, 테이블 등의 데이터 요소로 분리한다.

4.2.2 의미적 메타데이터의 자동 생성

분리된 각 데이터 요소에 대해 카테고리 자동분류(텍스트 마이닝), 패턴 인식, 테이블 스키마 인식 등의 기술을 적용하여 메타데이터를 생성한다. 이 과정에서 데이터의 내용에 근거한 의미적 인덱싱과 주석첨부 프로세스, 내용 기반의 인식과 메타데이터 자동 생성 등을 위한 자동화 도구는 전체적인 효율을 지배하는 매우 중요한 필수 요소이다. 각 카테고리는 이미 정의된 메타데이터를 가지고 있으며, 이 정보를 활용하면 이미 지나 오디오 등의 패턴 인식을 통한 메타데이터 생성의 정확도를 높일 수 있다. 다양한 형태의

멀티미디어 데이터는 의미적 메타데이터와 하나의 객체로 통합되어 사용 가능성이 높은(usable) 지식으로 생성된다.

4.2.3 테이블 스키마의 반자동 생성

콘텐츠 내의 테이블은 헤딩을 가지고 있으며, 이미 존재하는 몇몇 특정 데이터베이스의 테이블과 연관되어 있을 가능성이 크다. 관련 데이터베이스로부터 테이블에 대한 스키마 정보를 얻어서 이 테이블을 데이터베이스 내에서 관리하고 사용자 간에 공유할 수 있도록 하는 것이 스키마 생성의 목적이다. 스키마의 생성은 전반적인 자동화가 불가능하나, 스키마에 대한 구문적 및 의미적 정보와 스키마 간의 유사성에 근거한 추론을 통해 스키마 추천 기능은 자동화가 가능하다. 즉, 스키마 추천 자동화 도구가 유사한 스키마를 추천하면, 이것을 데이터베이스 관리자가 정제하여 스키마 생성을 완성한다. 이렇게 하여 스키마 생성은 반자동으로 이루어진다.

4.2.4 메타데이터의 의미적 정규화

의미적 정규화는 다양한 데이터 소스에 연관된 의미적 이질성을 다룰 때 매우 중요하다. 메타데이터의 정규화는 두 가지 방법이 있다. 첫째 방법은 데이터 소스와 포맷에 상관없이 동일한 도메인에 속하는 콘텐츠에 동일한 메타데이터를 연관시킨다는 것이다. 뉴스 메일의 기사와 웹사이트 기사가 '조류독감'에 관한 것이라면, 조류독감에 특정된 동일한 유형의 메타데이터가 두 기사에 동일하게 연관되어야 한다. 다른 방법은 동일한 개념에 대한 다양한 이름을 표준 이름으로 통일시키는 것이다. 예를 들어 조류독감 바이러스를 "Avian Flu Virus" "AVF Virus"로 명명했을 경우, 전자에 의한 키워드 검색은 후자를 포함하지 못한다. 이 문제는 표준 이름을 매핑시켜서 해결하며, 이름 간의 매핑은 상당히 요긴한 정규화 기법이다.

4.2.5 주석 첨부

주석이란 온톨로지의 개념과 그 개념을 포함하는 콘텐츠 간의 쌍방향 연결을 말하며, 이 연결을 통해 지식의 의미와 콘텐츠 간에 물리적인 참조가 가능해진다. 모든 온톨로지의 개념에 대해 이러한 주석을 생성하는 것을 주석 첨부이라고 하며, 주석 첨부은 상당히 전문적인 노력을 요구한다. 일반적으로, 온톨로지 생성 자동화 도구는 주석 첨부를 지원한다.

4.2.6 온톨로지 관리 시스템

지식의 의미는 온톨로지와 온톨로지 모형으로 표현 가능하다. 온톨로지 모형은 모든 사용자의 관점을 통합한 내부 모형과 각 사용자의 관점에 따른 외부 모형으로 구성된다. 내부 모형은 지식을 효율적으로 저장 및 관리하기 위한 것이며, 외부 모형은 사용자에게 익숙한 관점으로 지식의 의미를 서비스하기 위한 것이다.

온톨로지 관리를 위한 통합적 방법과 도구는 엄청난 양의 지식 자원을 다루는 어플리케이션의 개발에서 매우 중요하다. 온톨로지 관리 시스템은, 온톨로지 스미카의 설계부터 온톨로지 저장 및 검색, 논리적인 추론, 온톨로지 진화, 온톨로지 간의 상호 운용성을 고려한 온톨로지 매칭까지, 온톨로지의 생성과 활용에 관련된 다양한 기능을 지원해야 한다. 이러한 온톨로지 관리 시스템은 콘텐츠에 관한 의미를 일관성 있게 관리하고 지식의 상호운용성을 보장하는 기능을 담당하는 것으로, 지식관리시스템 기반구조의 핵심 요소이다.

온톨로지 모형의 개발 과정은 도메인에 특화된 전문 지식뿐만 아니라 궁극적으로 어플리케이션의 요구사항에 대한 이해를 필요로 한다. 이 과정을 반자동화하기 위해 클러스터링 기법이 초기에 사용될 수 있으나, 전과정을 자동화할 수 있는 일반적인 방법은 존재하지 않는다. 온톨로지 모형은 의미의 활용이 필요한 어플리케이션

을 위한 실세계의 의미를 반영하므로, 온톨로지의 중요한 부분이 된다. 온톨로지 모형과 온톨로지는 모두 의미 기반의 추론의 토대가 된다.

온톨로지 매칭[4]은 두 개의 온톨로지 간에 의미적 매핑을 찾는다. 가상적으로 다수의 온톨로지를 포함하는 어떤 응용에서도 이들 간의 상호운용성을 보장하기 위해 의미적 매핑이 정의되어야 한다. 이러한 응용은 지식 관리, 정보 수집, 생물 정보학, 웹 서비스 등 상당히 많은 도메인에서 나타나고 있다. 이렇게 폭넓은 보급에도 불구하고, 온톨로지 매칭은 아직도 주로 수동으로 수행되고 있다. 수동 매칭은 많은 노력이 들어 가면서도 오류가 나기 쉬운 문제점이 있으며, 대규모 정보관리 시스템의 구축에 주요 병목으로 작용하고 있다. 이러한 이유로 온톨로지 매칭 프로세스를 위한 자동화 지원 도구의 개발은 다양한 분야의 정보 공유에 매우 중요하다.

4.3 지식의 확산을 위한 기술 요소

지식의 확산에서 개인화는 사용자가 원하는 지식을 제공하여 활용도를 높인다는 차원에서 매우 중요하다. 개인화는 사용자가 주로 접근하는 지식을 파악하여 이에 관련된 지식이나 가이드를 배송하는 것을 말한다. 이를 위한 기술 요소로 사용자 행태 로깅 및 분석, 행태 데이터 관리, 배송(Delivery) 엔진 등을 나열할 수 있다. 지식은 웹을 통해 탐색된 여러 멀티미디어 콘텐츠를 조합하여 생성될 수 있기 때문에 웹 출판 시스템을 클라이언트에서 사용 가능한 기술 요소로 포함한다.

4.3.1 사용자 행태 데이터 관리

사용자가 지식을 사용할 때마다, 유일한 세션이 생성되고 사용자의 모든 행위가 사용자 세션 내에서 로그 파일에 남는다. 각 사용자 세션에 대해 로그 파일 내용을 파싱하여 누가 무엇을 사용하였는지 트리플 형식의 데이터 <주어, 술어,

목적어)로 생성하여 저장한다.

이 밖에 사용자 행태 데이터는 사용자, 항목 등과 같은 객체의 요약 정보, 객체의 위치, 객체 내용의 인덱싱, 객체 간의 관계를 나타내는 네트워크를 포함하며, 이후 이들 정보에 대한 사용자 접근 기록도 포함한다. 트리플의 술어는 네트워크 내에서 노드 간의 구조적 연결을 형성한다. 이러한 데이터는 사용자와 지식 간의 활용 행태에 대한 패턴을 찾거나 지식 네트워크를 찾기 위해 사용되며, 이렇게 찾아낸 패턴은 개인화의 근거가 되며 사용자 커뮤니티 성장 상태의 모니터링을 통해 지식 확산 활동의 촉진에 활용된다.

4.3.2 사용자 행태 로깅 및 분석

사용자가 지식에 대한 접근을 할 때마다, 이러한 행태를 하나의 레코드로 로그 파일에 저장한다. 사용자 행태 데이터 관리자는 이것을 파싱하여 데이터를 생성하여 관리한다. 이 데이터는 OLAP, 데이터 마이닝 등의 소프트웨어에 의해 개인화를 위한 사용자 행태 분석에 활용된다.

4.3.3 웹출판 시스템

인터넷 환경에서는 지식의 양이 폭발적으로 증가하고 있다. 서로 관련이 있는 기존 형태의 지식을 사용자가 직접 탐색하고 합성하여 새로운 형태의 지식으로 출판하는 기능은, 맞춤형 지식을 사용자 커뮤니티 내에서 신속하고 효율적으로 공유한다는 면에서 중요성을 갖는다. 최근 들어 기존의 데스크 탑 출판 시스템이 웹 환경으로 확장 발전되었고 사용자가 함께 동의하여 구축한 온톨로지는 지식 간의 의미적 혼란을 해소할 수 있다. 이러한 이유로, 사용자에게 편리한 웹 출판 시스템은 탐색된 지식을 합성하여 원하는 지식을 출판할 수 있는 매우 현실적인 기술 요소이다.

4.3.4 배송 엔진

배송(Delivery) 엔진은 지난 십 수년 동안 웹

사이트의 콘텐츠를 관리하기 위한 수단으로 많이 활용되어 왔다. 한 조직 내에 여러 웹 사이트가 동일한 웹 페이지를 포함하고 있는 경우, 특정 웹 사이트의 웹 페이지가 갱신되면, 관련 있는 다른 웹 사이트의 동일 웹 페이지도 함께 갱신되어야 마땅하다. 이러한 웹 페이지를 한꺼번에 갱신하기 위해, 배송 엔진은 갱신된 웹 페이지를 이들 웹 사이트에 전송하여 웹사이트를 갱신하도록 한다. 배송 엔진은 워크플로우 엔진을 포함하고 있어서, 업무 프로세스 중에 사용자가 웹 페이지를 갱신하면 즉시 다른 사이트로 전송하여 갱신을 유도한다.

4.4 지식 질의 처리를 위한 기술 요소

4.4.1 지식 중재자(Knowledge Mediator)

지식에 관한 포괄적인 질의가 들어왔을 때, 검색, 항해, 지식 네트워크 마이닝 중에서 무엇을 할 것인지, 온톨로지 저장/검색 시스템, 콘텐츠 관리 시스템, DBMS 등을 어떻게 조합하여 사용할 것인가에 대해 결정하고, 이들의 결과를 합성하는 역할을 수행한다. 여기서 합성이란 온톨로지에 저장된 의미를 이용하여 부분 결과 간의 이질성을 해소하고 다양한 형태의 콘텐츠를 하나의 콘텐츠로 묶는 것을 말한다. 분산 환경의 경우, 이질적인 지식 소스로부터 필요한 지식을 추출하는 래퍼(wrapper)의 역할도 포함한다. 지식 중재자는 온톨로지 관리시스템, 의미적 검색 엔진, 온톨로지 항해자, 지식 네트워크 마이닝 등을 이용하여 콘텐츠와 의미에 관한 포괄적인 질의를 처리하고 콘텐츠를 합성한다.

4.4.2 의미적 검색 엔진

현재의 검색 엔진은 질의의 문맥을 고려하지 않기 때문에, 의미가 다른 동음어를 구분하지 못한다. 어떤 경우에는 검색이 기술 카테고리에 제한될 수도 있으나, 복잡한 질의의 경우, 현재의 검색 엔진은 키워드에 의존하기 때문에 키워드

간의 관계에 의한 의미를 고려하지 않아 의미에 부합하지 않는 검색 결과를 상당히 포함한다. 데이터에 연관된 의미적 주석 또는 메타데이터는 검색은 물론 브라우징과 개인화에도 더욱 강력한 기능을 제공한다.

의미적 검색 엔진[5]은 추론을 통해, 사용자가 명시적으로 요구하지는 않았더라도 의미적으로 관련있는 정보를 사용자에게 제시해야 한다. 예를 들어 'H5N1'을 키워드로 하여 의미적 검색을 실행하는 경우, '조류독감이 사람에게 미치는 영향에 관한 보고서'를 연관시키는 것은 매우 중요하다. 왜냐하면 'H5N1'은 조류독감 바이러스 중에서 '사람'에게 감염될 수 있고 감염되면 '치명적'일 뿐만 아니라 감염된 사람을 매개로 하여 '사람' 간에 전파'될 수 있기 때문이다. 이와 같은 검색 결과를 제공하려면, 검색 엔진은 이들 간의 의미적 연관 관계가 있다는 것을 추론할 수 있어야 한다.

4.4.3 온톨로지 기반의 지식 항해자

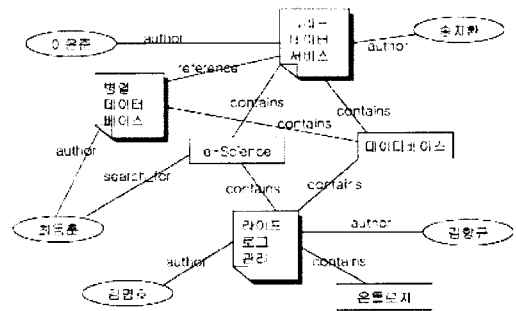
사용자는, 키워드 검색에 의존하는 대신에, 특정 주제에 대한 폭넓은 관심을 가지고 지식을 항해하면서 브라우징을 한다. 따라서, 야후와 같이 이미 주어진 지식 맵이나 텍소노미에 따라 주제 중심의 항해를 하여 원하는 지식을 탐색한다.

여러 사용자는 지식 탐색에 대한 서로 다른 관점을 가지고 있다. 항해를 위한 지식 맵이 지원되더라도 사용자의 관점과 다르거나 용어가 서로 달라 의미적 혼란이 가중될 수 있고, 지식 맵의 그래픽 인터페이스가 좋지 않으면 항해 중에 방향을 잃을 수도 있다. 이러한 이유로 온톨로지 기반의 의미적 항해를 통한 브라우징이 가능해야 한다.

4.4.4 지식 네트워크 마이닝

지식을 구성하는 객체는 서로 참조 연관관계를 가지고 있다. 이들 간의 참조 연관관계는 그래프 구조로 표현될 수 있으며 이것을 지식 네트

워크라고 한다. 예를 들어 연구자와 연구자 간의 관계, 연구자와 콘텐츠 간의 참조 관계를 그래프로 나타내면 (그림 3)과 같은 지식 네트워크가 된다. 이러한 참조 연관관계는 트리플 데이터 모형으로 표현되어 온톨로지에서 저장 및 관리할 수 있다.



(그림 3) 지식 네트워크의 예

지식 네트워크는 객체의 인스턴스와 인스턴스 간의 참조 연관관계에 근거하며 이러한 연관관계는 고정되어 있지 않고 진화하게 마련이다. 따라서 이러한 지식 네트워크를 찾아 내는 것은 인스턴스 간의 참조 연관관계에 관한 패턴을 찾아 내는 것[6]으로, 지식 네트워크의 효율적인 탐색은 참조 연관관계에 관한 트리플 데이터의 자동 생성 및 효율적인 마이닝 기법에 전적으로 의존한다.

지식 네트워크는 매우 복잡한 구조를 갖기 때문에 그래픽 인터페이스를 제공하지 않을 경우 이해하기 매우 어렵다. 그래픽 인터페이스는 사용자가 제시한 질의의 검색 결과를 지식 네트워크로 쉽고 즉각적으로 이해할 수 있도록 표현한다. 지식 네트워크는 키워드와 그에 관련된 검색 결과 간의 관계뿐만 아니라 검색 결과 간의 관계도 표현한다. 이밖에 지식 네트워크는 사용자가 관심 있는 주제에 대한 전반적이고 개괄적인 정보를 제공하여 원하는 정보를 더욱 직관적으로 쉽고 편리하게 검색할 수 있도록 한다.

5. 요약 및 향후 연구

우리는 e-Science에서 필요한 지식관리시스템의 기반구조가 무엇인지, 이를 위한 기능적 요구사항과 기술적 요소에는 무엇이 있는지 소개하였다. 이를 구현하기 위한 기술 요소는 대부분 공개 소프트웨어나 상용 소프트웨어로 나와 있으나, 온톨로지 관리 시스템과 지식 중재자를 비롯한 몇 가지 기술적 요소는 사용 가능한 소프트웨어가 존재하지 않는다. 우리는 이러한 기술적 요소에 대한 기술 연구 및 개발을 수행할 계획이다.

e-Science에서 지식관리를 적용하려면 사용자의 요구사항에 따른 시나리오를 정의하는 것이 매우 중요하나, 연구 커뮤니티에 따라 요구사항이 매우 유연하다. 의생물학 분야의 조류독감 연구 e-Science 환경에 대한 우선 적용을 목적으로 사용자 시나리오를 정의할 예정이며, 장기적으로 이를 위한 지식관리시스템을 지식 포털 형태로 개발할 것이다.

참고문헌

- [1] L. Chen, N. Shadbolt and C. Goble, "A Semantic Web-based Approach to Knowledge Management for Grid Applications," IEEE TKDE Vol.19, No.2, 2007, pp.283-296
- [2] A. Sheth, "Semantic Content Management for Enterprises and the Web," IEEE Internet Computing, 2002, pp.80-87
- [3] <http://en.wikipedia.org/wiki/Ontology>
- [4] A. Doan, "Reconciling Schemas of Disparate Data Sources: A Machine Learning Approach," Proceedings of the ACM SIGMOD Conference, 2001
- [5] R. Guha, "Semantic Search," Proceedings of the 12th International Conference on World Wide Web, 2003
- [6] H.D. Green, N.S. Contractor and Y. Yao, "CI-KNOW: Cyberinfrastructure Knowledge Networks on the Web," <http://sonic.ncsa.uiuc.edu>, 2006

저자약력



Seok Dong-un

1981년 2월 서울대학교 계산통계학과 졸업(학사)
 1983년 2월 한국과학기술원 전산학과 졸업(석사)
 1986년 6월 Northwestern University 전산학과 졸업(박사)
 1983년 2월 ~ 1986년 8월 한국증권전산(주) 과장대리
 1989년 8월 ~ 1992년 2월 한국국방연구원 선임연구원
 1992년 3월 ~ 1999년 2월 동덕여자대학교 전산학과
 부교수(1992년 ~ 1999년)
 2004년 6월 ~ 2005년 1월 한국과학기술원 초빙교수
 2005년 2월 ~ 현재 한국과학기술정보연구원 책임연구원



Lee Eun-jeon

1977년 2월 서울대학교 계산통계학과 졸업(학사)
 1979년 2월 한국과학기술원 전산학과 졸업(석사)
 1983년 6월 프랑스 ENSIMAG-INPG 전산학과 졸업(박사)
 1984년 3월 ~ 현재 한국과학기술원 전산학과 교수
 관심분야 : 데이터베이스 시스템, 인터넷, 웹 데이터베이스,
 분산 시스템