

국가지식정보자원관리를 위한 시맨틱웹 설계 및 정책 방향에 관한 연구*

A Study of Designing Semantic Web and Policy Directions for National Knowledge and Information Management

오 삼 균(Sam-Gyun Oh)**

초 록

이 연구의 목적은 국가지식정보자원의 효율적인 관리에 필요한 시맨틱웹 설계와 정책방향을 제안하는 것이다. 이를 위해서 구체적으로 이 논문에서 다룬 내용은 1) 메타데이터 요소, 자원, 온톨로지의 클래스와 속성들에 불변의 식별자를 부과하는 URI체계의 수립, 2) XML 네임스페이스의 적극적인 활용정책의 수립, 3) 국가통합검색을 위한 메타데이터와 응용 프로파일 표준의 설정, 4) 메타데이터의 의미적 상호운용성을 증진시키기 위한 ISO 11179 기반의 레지스트리 설계와 구축 방향의 제시, 5) W3C의 OWL과 ISO/IEC의 토픽맵(Topic Maps)을 활용한 온톨로지 구축의 필요성에 대한 논의, 6) 메타데이터 기반 지능형 검색서비스의 제공, 7) 국가지식자원관리의 미래 방향과 과제의 제시, 등에 관한 것이다.

ABSTRACT

The purpose of this study is to design semantic web and policy direction for national knowledge and information management. The paper describes all the components needed to accomplish the objective: 1) creating unchangeable and unique identifiers for metadata elements, resources, and ontology classes and properties; 2) recommending active use of XML namespaces; 3) establishing metadata and application profile standards for national integrated searching; 4) developing a metadata registry to promote semantic interoperability among metadata; 5) discussing the need of creating ontologies using W3C OWL and ISO Topic Maps; 6) providing intelligent search services based on metadata; and 7) presenting future directions and tasks of national knowledge and information management.

키워드: 메타데이터, 메타데이터 레지스트리, 시맨틱웹, 온톨로지
Metadata, Metadata Registry, Semantic Web, Ontology

* 본 논문은 2003년도 한국전산원 용역과제로 수행되었음. 2004년도 한국비블리아학회 춘계학술발표회에서 발표한 내용을 수정·보완한 것임.

** 성균관대학교 문헌정보학과 부교수(samoh@skku.edu)
논문접수일자 2004년 5월 25일 논문심사일자 2004년 6월 7일 게재확정일자 2004년 6월 23일

1. 연구의 필요성 및 목적

국제적으로 메타데이터의 중요성에 대한 인식이 점차 확산되어 다수의 메타데이터 관련 표준 활동이 활발히 진행되고 있다(Ahmed 2001). 특히 미국, 영국, 오스트레일리아, 노딕국가들이 이 분야에서 앞장서 나가고 있으며, 지식정보자원의 체계적 관리에 많은 투자를 하고 있는 실정이다. 각 분야와 매체별 메타데이터 표준 활동도 활발히 진행되고 있지만, 각 기관에서 정의하고 있는 메타데이터 스키마들 간의 상호운용에도 많은 관심을 표명하고 있다(Manola and Miller 2002). 대표적인 예로는 ISO/IEC SC32에서는 각 기관에서 정의하는 메타데이터 요소의 이름과 정의를 제대로 함으로써 데이터의 상호 운용을 증진시키기 위해서 ISO 11179 Metadata Registry(이하, MDR)에 관한 국제 표준을 제정하였고, W3C에서는 현재의 웹 자원들에 관한 메타데이터의 의미를 기계가독형으로 표현할 수 있는 시맨틱웹을 구축하기 위해서 많은 노력을 기울이고 있다(Berners-Lee 2000). 또한 이러한 노력들이 서로 통합되고 서로의 장점을 적극 도입하여 활용하려는 추세이다. 국내 지식정보자원의 경쟁력을 계속 유지하고 발전시키기 위해서는 국제 표준에 맞추어 메타데이터를 체계적으로 관리하는 체제를 구축하고 이를 지속적으로 유지하고 발전시켜야 한다.

이런 추세에 대응하기 위해서 정보통신부에서는 매년 많은 예산을 투입하여 국가지식자원을 디지털화 하고 효율적인 대 국민 검색서비스를 하기 위한 노력을 경주해왔다. 처음에는 기관 간의 메타검색에 주로 의존하여 국가통합

검색을 시도하였으나, 검색의 결과 뿐 아니라 검색속도에 있어서도 문제가 발생하여 근본적인 대책이 필요한 실정이다. 2000년도부터 검색의 질을 향상시키기 위해서 메타데이터의 구축을 적극권장하게 되었고, 국가지식자원관리를 위한 첫 메타데이터 표준이 제정되었다. 또 하나의 큰 문제는 URL을 자원의 식별자로 사용하고 있기 때문에 약 6개월이 지나면 50% 정도의 링크가 제대로 작동되지 않는다는 점을 개선할 필요성이 대두되었다. 각 기관에서 사용하고 있는 메타데이터들 간의 상호 운용을 증진시켜야 할 필요성과 국가지식정보자원을 체계적으로 분류하여 신속성과 확장성 있는 네비게이션을 제공하는 온톨로지 구축의 필요에 대한 공감대를 넓혀 가고 있다.

이러한 문제들을 해결하는데 도움을 주기 위해서, 이 연구에서는 2000년도 지식정보자원 사업으로 추진했던 메타데이터 표준의 이용 현황에 대한 조사와 분석을 수행하였고, 국외의 메타데이터 이용 사례들을 분석하여 국내 지식정보자원 관리에 도움이 될만한 지침의 방향을 설정할 수 있는 안을 제시하고, 각 기관에서 사용하고 있는 메타데이터 요소를 등록할 수 있는 ISO 11179 기반 MDR 설계의 방향, 국내 지식정보자원 관리를 시맨틱웹 기반으로 체제를 정비하기 위해서 필요한 국제 표준과 요소기술에 대해서 논함으로써 국가지식정보자원관리를 효율적으로 수행할 수 있는 방향을 제시하는 것을 목표로 하고 있다.

2. 국가지식정보자원관리 메타데이터 표준 활용 현황

국가지식정보사업으로 지원된 모든 프로젝트의 효율적인 정보검색을 도모하기 위해서 필수 메타데이터 요소를 선정하여 지식정보사업의 지원으로 구축된 2000년 모든 시스템에 적용할 것을 적극 권장하였다. 이 지식정보사업에 참여한 각 기관에 메타데이터의 이용현황에 대한 협조 자료를 부탁하였고, 그 중 3개 기관에서 자료를 제출하여 제출된 자료를 기반으로 국내의 이용현황을 분석하였다. 기관의 이름은 구체적으로 밝히지 않기로 한다.

각 기관의 요구사항에 맞추어 메타데이터 스키마를 설계할 때 먼저 국제표준 메타데이터 스키마를 적극 참조하고 채택해야 한다. 그러나 국제표준 메타데이터 요소들이 그 기관에서 필요한 메타데이터 요소를 포함하고 있지 않을 경우에는 국제표준에 없는 요소에 한하여 확장할 필요가 있다. 국제표준 메타데이터 스키마는 하나의 네임스페이스에 속하게 되고, 각 기관이 새로운 요소들을 제정할 경우에는 그 요소들에 대한 이름과 정의를 네임스페이스로 관리할 책임을 져야 한다. 이렇게 국제표준 메타데이터 요소와 기관이 새로 정의한 요소를 합하여 특정 시스템을 위해서 새로 제정한 메타데이터 셋(set)을 Application Profile(이하, AP) 이라 부른다.

2.1 A 기관

A기관은 첫째로 기술보고서에 관한 원문정보서비스를 제공하고 있다. 이 서비스를 위해서

설계한 AP를 살펴보면 지식정보사업에서 제정한 필수요소를 다 수용하였고, 그 외에도 4개의 요소를 더 추가하였다. 이 AP에서 채택한 주 요소는 Title(표제), Creator(생성자), Subject(주제), Description(요약정보), Publisher(출판사), Date(날짜), Type(자료유형), Format(포맷), Identifier(식별자), Language(언어), Relation(관계정보) 등이다.

또한 이 AP에 더블린코어 이외의 다른 주요소를 첨가하지는 않았으나, 다수의 하위요소를 확장하였다. 이 기관에서 원문정보서비스를 위해서 첫째로 Subject의 하위요소로 Classification(공식분류체계)과 Keyword(키워드)를 확장하였다. 이 기관에서 일괄적으로 사용하고 있는 표준 주제어는 Classification 요소의 값으로 입력하고, 원문의 저자가 제공한 주제어는 Keyword(키워드) 하위요소의 값으로 입력하여 주제정보의 접근성을 향상시키려는 노력을 기울였다. 또한 다른 주제어 요소를 주요소로 확장하지 않고, 국제 표준인 더블린코어의 Subject의 하위요소로 확장한 것도 메타데이터의 국제 권고사항에 준수한 것이다. 다른 기관에서 이 하위요소를 이해하지 못할 때는 그 상위 요소인 Subject 요소의 값으로 방출된다. 이러한 절차를 국제적으로 동의가 이뤄진 dubm-down 원칙이다. 다시 말하면, dubm-down 원칙은 하위요소의 의미가 국제표준이 아닐 경우에는 그것의 상위요소로 메타데이터를 방출하는 법칙이다. 하위요소의 깊이에는 제한이 없기 때문에 이런 dubm-down의 과정은 계속 위쪽으로 반복하여 발생할 수 있다.

둘째 확장은 Description의 하위요소로

toc, abstract-kr, abstract-en 등을 제정 하였다. toc는 더블링크어에서 제시하는 table-OfContents로 변경하고, abstract-kr과 abstract-en은 확장할 필요 없이 더블링크어 요소인 abstract 하나로 족하고 언어에 관한 인코딩 스킴으로 표현하면 될 것으로 판단된다. 예를 들면 <abstract xml:lang="kr">한국어 초록 내용</abstract>, 그리고 영문 초록 내용도 <abstract xml:lang="en">영문 초록 내용</abstract>으로 하면 된다. 그러나 현재 XML 파싱 기술이 불완전한 상태이어서 시스템 구축 시 XML 속성보다 요소를 선호하는 현상이 있으나, AP를 설계 할 때는 이런 원칙을 준수하고 구축 단계에서는 다양한 방법을 적용할 수 있다고 본다.

셋째 확장은 관계요소를 표현하기 위한 Relation의 하위요소로 year와 level로 제정하였는데, 이에 대한 예제가 없어서 정확한 의미와 용도에 대한 판단이 어려웠다. 그러나 Relation 정보의 하위요소를 확장할 때도 그것이 Relation 정보로 dubm-down이 되어도 의미가 성립해야 한다는 점, 또한 Relation이나 Relation의 하위요소의 값은 URI에 해당하는 것들만 올 수 있다는 것도 유념해야 할 사항이다.

또한 이 기관에서는 건설공사정보에 관한 데이터베이스도 보유하고 있다. 이 데이터베이스를 서비스하기 위해서 제정한 AP는 12개의 더블링크어 주요소(Title, Creator, Subject, Description, Publisher, Date, Type, Format, Identifier, Language, Relation, Rights)를 채택하였고, 1개의 주요소를 확장하였다. 확장한 주요소 projectOrganization은 건

설공사정보에 아주 중요한 정보로 더블링크어의 어떤 요소와도 매핑이 되지 않아서 확장한 것으로 판단된다.

이 기관에서 건설공사정보 서비스를 위해서 확장한 하위요소들을 살펴보면, 첫째로 Creator의 하위요소로 personalName(개인명)과 corporateName(회사명)으로 확장하였다. 둘째로 Subject의 하위요소로 combined(통합주제)와 specific(개별주제)으로 나누었다. 셋째로 Description의 하위요소로 다시 description 하위요소를 두었는데 이 요소의 이름은 상위요소와 같아서 변경하는 것이 바람직하다고 본다. 넷째로 Publisher의 하위요소로 personalName과 corporateName을 확장하였다. 다섯째로 Relation의 하위요소로 isBasedOn과 isBasisFor를 첨가하였다. 마지막으로 새로 추가된 주요소 projectOrganization의 하위요소로 owner, contractor, designer, constructionManager를 추가 하였다.

2.2 B 기관

B기관은 학술지, 학술지 권호데이터, 학술논문 서비스를 하기 위해서 하나의 AP를 제정하였고, 이 AP에서 13개의 더블링크어 주요소(Title, Creator, Subject, Description, Publisher, Date, Type, Format, Identifier, Language, Relation, Source, Rights)를 채택하였다. 이 AP를 설계하는데 있어서 주요소는 확장하지 않았고 하위요소만 확장하였다. 이 기관이 확장한 것은 첫째 Creator의 하위요소로 alternative(대체저자명)과 affiliation

(저자소속), 둘째 Description의 하위요소로 abstract.alternative(대체초록), 셋째 Publisher의 하위요소로 alternative(대체발행처)와 place(발행국), 넷째 Contributor의 하위요소로 alternative(대체기여자), affiliation(기여자소속), role(기여자역할), 다섯째 Date의 하위요소로 metadataCreated(메타데이터 작성일)과 metadataModified(메타데이터 수정일), 여섯째 Format의 하위요소로 pages(수록페이지), 일곱째 Source의 하위요소로 alternative(대체소스명), type(소스유형), volume(권호사항) 등으로 확장하였다. 더블링크어의 정의에 의하면 Source의 값으로는 URI만 가능하기 때문에 이 기관의 Source의 확장은 dumb-down원칙에 위배된다고 판단된다. Source에 이와 같은 정보를 입력하고자 하는 요구가 반영되어서 더블링크어에서 Identifier의 하위요소로 bibliographicCitation이라는 하위요소를 표준으로 제정하였다. Source에 대한 하위요소의 확장을 더블링크어의 bibliographicCitation의 하위요소를 변경할 것을 권한다.

2.3 C 기관

C기관의 AP는 두 개를 제정하였고, 첫 AP는 연구보고서를 위한 것이고, 두 번째 AP는 학술서지를 위한 것이다. 첫 AP에서는 표준 더블링크어의 10개의 주요소(Title, Creator, Subject, Description, Date, Format, Identifier, Language, Rights, Audience)와, 더블링크어 하위요소 중 6개(created, issued, extent, rightsAccess, educationLevel, medi-

ator)를 채택하였다. 두 번째 AP에서는 8개의 더블링크어 주요소(Title, Creator, Subject, Description, Date, Type, Source, Language)와 2개의 하위요소(created, issued)를 채택하여 사용하고 있다.

3. 국외 메타데이터 활용 사례

W3C를 중심으로 전개되고 있는 시맨틱웹의 핵심은 메타데이터의 의미를 제대로 표현하는데 있다(Berners-Lee 2000). 지식포털 사이트들도 메타데이터를 제대로 개발하여 이를 토대로 모든 서비스를 제공하는 추세이다. 국외 메타데이터 사례는 너무 많아서 이 논문에서 다 포함할 수 없고 그 중에서 대표적이고 특색 있는 사례들만 소개하고자 한다. 미국국립과학디지털도서관 프로젝트는 Open Archive Initiative(이하, OAI) 프로토콜을 광범위하게 실제 상황에서 적용한 것으로 OAI를 사용해서 얻을 수 있는 장점들을 파악하는데 도움이 될 것이다. 또한 MusicBrainz는 시맨틱웹 온톨로지의 핵심개념인 Resource Description Framework(이하, RDF)를 적극 활용한 프로젝트로 RDF를 적용하여 어떻게 정보서비스를 향상시킬 수 있는지에 대해 참조할 수 있는 사례이다.

3.1 미국국립과학디지털도서관(NSDL 프로젝트)

NSDL은 미국 과학재단(NSF)이 지원 하에 과학, 기술, 공학 및 수학(STEM) 교육의 질을

지속적으로 향상시키고 촉진시키기 위해 개발된 국가과학전자도서관이다(Lagoze 2002). 교육자, 과학자, 콘텐츠 제공자 간의 협력을 통해 개발된 NSDL은 세계에서 가장 광범위한 전자 교육 네트워크 개발을 목적으로 한다. 궁극적으로, NSDL은 전자 도서관을 통한 과학 교육 분야에서 가장 선도적인 역할수행을 목표로 하고 있다. 2000년 이후 현재까지 NSF가 지원하는 144여개의 하부 프로젝트가 NSDL의 4대 핵심 연구 분야인 핵심 통합(core integration), 자료 수집(collections), 서비스 개발(services) 및 집중연구(targeted research)를 수행하여, 수집된 자료를 NSDL를 통해 통합되어 서비스되고 있다. NSF는 2001년 핵심 통합을 위한 기금을 연구대학연합(UCAR)에 속해 있는 콜롬비아 대학과 코넬 대학에 주어 NSDL 시스템을 구축하도록 하였다.

다양한 자료간의 상호운용은 메타데이터를 통합하려는 노력의 주요 테마 중의 하나이다. NSDL을 통해 서비스되는 자료는 다양한 데이터 타입, 메타데이터 표준, 프로토콜 인증 체계와 비즈니스 모델을 수용하며, 수많은 기관들에 의해서 관리되고 있다. 이러한 다양한 자료를 지원하는 상호운용 체계의 설립을 위해 기술적, 내용적, 조직적 합의를 도출하였다. NSDL 전자도서관의 상호운용을 위한 모델로 연합(federation), 메타데이터 수집(harvesting) 및 자료 수집(gathering) 중에서 Harvesting 방법을 채택하였다. NSDL은 OAI (Open Archives Initiative) 프로토콜을 활용하여 참여 기관들 간의 메타데이터 방출을 용이하게 하였다. 참여기관은 자신이 소유한 메타데이터를 더블린코어로 매핑하는 것은 의

무적으로 수행해야 하며, 다른 메타데이터는 XML 스키마를 같이 제공해야 한다. 이런 방법으로 과학 분야의 디지털 자원을 공유할 수 있는 Z39.50에 비해 구축하기가 용이한 가벼운 모델을 성공적으로 이룬 프로젝트로 인식되고 있다.

현재 국가통합검색시스템에서 질의를 신청할 경우 종합정보센터와 전문정보센터에 메타검색을 요청하므로 소요되는 시간이 너무 길고 검색의 정확도에 있어서도 많은 문제점을 안고 있어서, NSDL 모델이 국가지식정보사업의 통합검색의 효율성을 높이기 위해서 적용할 수 있는 좋은 모델로 생각된다. OAI 프로토콜을 채택하여 각 기관이 OAI 표준에 맞춰 메타데이터를 방출하는 것을 의무화하여 이렇게 수집된 메타데이터를 중앙 Metadata Repository(이하, MR)에 관리할 필요가 있다. 국가 지식정보자원의 메타데이터를 중앙 MR 한 곳에 저장하거나, 각 종합정보센터에 MR을 두어 산하전문정보센터의 메타데이터를 통합해서 저장해야 검색시간을 단축하고 검색의 질을 높일 수 있다고 본다. 현재 국가통합검색시스템의 검색속도를 높이고 질 높은 검색을 제공하기 위해서는 NSDL과 같은 MR을 구축할 것을 권한다. 현재 국가지식자원관리 사업으로 지원되는 모든 사업은 메타데이터 구축을 의무화해야 하고, 구축된 메타데이터는 종합정보센터를 거쳐서 중앙 MR에 자동적으로 방출되어야 한다. 이러한 MR이 구축되지 않으면 국가통합검색시스템이 메타검색으로 인한 검색속도 지연을 해결하기 어려울 것이다. 각 기관이 메타데이터를 MR에 방출할 때는 국제적으로 널리 활용되고 있는 OAI 프로토콜을 채택할

것을 권장한다.

3. 2 오스트레일리아 디지털 공학도서관 (AVEL 프로젝트)

AVEL 프로젝트는 양질의 공학 및 정보기술을 공학도와 그 외의 전문가에게 제공하기 위해 개발된 인터넷 포털이다. AVEL은 1998-2002년에 오스트레일리아 연구협회(ARC)의 지원 하에 개발되었다. 그 이후, AVEL-SD(Sustainable Development)를 거쳐 현재 AVEL-SKN(Sustainability Knowledge Network)이 AVEL의 차세대 인터넷 포털 서비스로서 제공되고 있다. AVEL-SKN에서 강조하는 지속성(sustainability)은 경제적(기술적 경제성을 포함), 생태학적, 그리고 사회적 지속성을 의미한다. 즉, 공학적 생산품은 기술적으로 견고하고 비용적인 측면에서 효율적이어야 하며 생태학적 그리고 사회적 지속성을 유지해야 한다(AVEL-SKN 2003).

초창기의 AVEL은 퀸스랜드 대학을 주축으로 8개의 대학협회 및 정부 기관이 공동으로 개발하였다. AVEL 포털의 차기 버전으로 제안된 AVEL-SD는 기존의 AVEL 포털에 포함되어 제공되다가 AVEL-SKN으로 통합되었다. AVEL-SKN은 단순히 정보를 제공하는 기존의 포털 개념에서 탈피하여, 인터넷을 통하여 공학, 신기술 및 지속성에 관련한 정보의 제공과 동일한 목적을 가진 사용자를 한데로 묶는 공간을 제공한다. AVEL-SKN은 지속적으로 개선되어 왔으며, 메타데이터 수집(harvesting)은 사이트의 두 번째 개발 단계의 중요한 요소이다. 메타데이터 수집 기능을

통해 AVEL-SKN이 보유, 처리할 수 있는 자료의 양은 빠른 확장을 보이고 있다. 현재 AVEL가 지원하는 메타데이터 요소로는 총 19개가 있다.

3. 3 MusicBrainz 프로젝트

MusicBrainz(<http://www.musicbrainz.org/>)는 음악 레코드에 관한 메타데이터 데이터베이스 표준이며, 이 표준은 더블링크어 메타데이터 표준을 확장한 것이다. 음악 메타데이터는 앨범에 나타나는 예술가, 앨범 이름, 오디오 트랙 리스트 등을 나타내는 정보이다. MusicBrainz는 MP3, Ogg/Vorbis 파일에도 적용된다. 현재 6만 명 이상의 예술가, 10만개 이상의 앨범, 100만개 이상의 오디오 트랙에 관한 메타데이터를 저장하여 서비스하고 있다. MusicBrainz는 음악에 관한 메타데이터 검색서비스를 제공하는 것이지 음원 자체를 제공하는 것은 아니며, 음악에 관한 다양한 정보를 깊이 있게 검색할 수 있는 서비스를 제공하기 위해서 노력하고 있다.

MusicBrainz는 MusicBrainz 메타데이터(이하, MM)라는 자체 개발한 메타데이터 집합을 가지고 있고, 현재 버전 2.1이 개발된 상태이다. MM은 디지털 오디오나 비디오 트랙에 관련한 메타데이터를 저장 또는 교환하는 수단을 제공한다. 즉, MM은 인터넷상에서 오디오나 비디오 트랙의 콘텐츠를 기술하는 모델이다. MM은 오디오나 비디오 메타데이터를 교환하는 메커니즘으로 RDF/XML을 사용하고 있으며, 실제로 운영되고 있는 애플리케이션에서 RDF가 활용되는 한 사례이다. 앞으로

RDF에 기반을 둔 서비스가 점점 증가할 것으로 전망된다. MM표준은 더블링크어 메타데이터를 사용하는 RDF 네임스페이스를 지원하며, MusicBrainz 서버가 제공하는 모든 데이터는 MM과 더블링크어의 형식으로 기술되어있다.

3. 4 노르딕 국가 메타데이터 프로젝트

스칸디나비아 반도에 위치하는 북유럽 국가들 간의 도서관 상호연계대출(Inter Library Loan)과 문서전달 서비스를 보다 용이하게 하기 위해, 공유할 수 있는 메타데이터를 생성할 시스템에 대한 필요가 대두되었다. 이러한 요구에 부응하여 시작된 것이 노르딕 국가 메타데이터 프로젝트이다. 노르딕 메타데이터 프로젝트 I은 첫 국제 더블링크어 프로젝트로써 NORDINFO(<http://www.nordinfo.helsinki.fi>)의 재정적 지원 하에 노르웨이의 Bibsys, 덴마크 도서관 센터, 헬싱키 대학 도서관, Lund 대학 도서관, SICS, 스웨덴의 Netlab, 아이슬란드 국가 대학도서관이 공동으로 참여하여 1996년 10월 시작되어 1998년 6월에 완료되었다. 이 기간에 더블링크어 메타데이터 집합에 기반을 둔 메타데이터 생성과 이용시스템을 개발하였다. 또한, 현존하는 메타데이터 표준들(더블링크어, IAF A, MARC, SOIF) 등을 평가하고, 더블링크어 표준을 확장하는 방법을 채택하였다. 그리고 더블링크어를 노르딕 MARC 형식으로 전환하는 방법, 더블링크어 메타데이터 구문의 생성, 사용자 환경에 대한 연구, 메타데이터 검색 서비스를 통한 노르딕 인터넷 문서의 검색과 발견에 관한 연구도 병행하였다.

노르딕 메타데이터 프로젝트의 성공에 힘입어 60개국 이상에서 노르딕 프로젝트 사이트를 방문하였으며, 28개국 이상에서 이 프로젝트에서 개발한 메타데이터 템플릿을 이용하여 더블링크어 레코드를 생성할 수 있게 되었다. 1999년 1월에 노르딕 메타데이터 프로젝트 II가 NORDINFO의 지원 하에 시작되었고, 이 프로젝트 II를 통하여 15개의 더블링크어 요소와 함께 개별 메타데이터 요소를 추가하는 방법의 개발과 XML 등의 새로운 문서 양식을 더블링크어로 표현하는 방법을 강구하였다. 이외에, 프로젝트 I에서 개발된 더블링크어-MARC 변환 시스템을 개선하였고, 개선된 더블링크어 사용자 가이드, 메타데이터 템플릿 및 메타데이터 수집(harvesting)하고 색인 할 수 있는 도구들을 개발하여 제공하고 있다. 도너(donor) 프로젝트(<http://www.konbib.nl/donor/index-en.html>)는 메타데이터 생성기를 평가하여, 프로젝트 내에서 메타데이터를 생성하는데 노르딕 메타데이터 생성기를 사용하기로 하였다. 현재, 노르웨이 국가 도서관은 웹 문서에 NBN을 부착함으로써 URN 체제를 도입하였고, 스웨덴과 핀란드의 국가 도서관들도 이러한 서비스를 시작하였다.

3. 5 영국 e정부 메타데이터 표준(e-GMS) 프로젝트

영국의 e정부 메타데이터(e-GMS)는 공공기관 간의 메타데이터 일관성을 유지하기 위해서 제정된 표준이다. e-GMS는 e정부 상호운용 프레임워크(e-Government Interoperability Framework, e-GIF)의 한 부분이고

현재 버전 2.0까지 개발되어 있다. e-GMS는 포괄적인 표준이어서 영국의 공공 분야에서 필요한 모든 요소와 요소들의 확장을 포함하고 있다. 특정 시스템이 이 모든 요소를 필요로 하지는 않을 것이다. 각각의 도메인의 필요에 따라 각 시스템의 요구에 따라 표준을 정하고, e-GMS의 요소 중 그 도메인에 필요한 요소만을 채택하여 사용하면 된다. e-GMS는 1.0은 간단한 더블링크어 요소들로 구성되어 있다. 더블링크어는 e-GMS 2.0 표준에서도 가장 중요한 요소이며, 더블링크어의 주요소, 주요소의 확장(refinements), 인코딩 스킴 등이 더블링크어의 정의와 주석과 함께 최대한 채택되었다. e-GMS 버전 2.0에서는 또한 정보와 레코드 관리를 위한 새로운 요소들이 추가 되었다. e-GMS의 각각의 요소는 정의, 의무사항(의무사항, 적용 시 의무사항, 권장사항, 선택사항), 요소의 목적, 주석, 확장, 예제, HTML 구문, 인코딩 스킴, 다른 표준으로의 사상(mapping, 寫像) 정보(더블링크어, AGLS, GILS, IEEE LOM) 등이 기술되었다. 각각의 e-GMS 요소는 독립적이고 간단하며, 기타 영국의 정부 표준과 부합하도록 제작되었다. 또한, 정보 환경의 변화에 적용할 수 있도록 제작되었으며 안정적이고 확장 가능하다.

e-GMS의 상위 표준인 e-GIF 표준은 인터넷으로 상호 연결된 정부를 만드는데 중요한 선결요소들을 정의하며 영국의 e-정부 전략의 가장 중요한 초석을 제공하고 있다. e-GIF 표준은 두 부분으로 이루어져 있다. 첫 부분은 프레임워크로서 정책에 관련한 문안, 관리와 구현에 대한 표준 부합성에 관한 것을 기술한다. 두 번째 부분은 기술적 정책과 명세에 관

한 것으로서 기술적 정책, 명세표, 어휘 및 약자 등에 관한 내용을 포함하고 있다. e-GIF는 정부와 공공 분야 간의 정보 교환과 흐름에 관한 최소한의 기술적 정책에 관한 자세한 내용을 담고 있다. e-GIF는 상호연결, 데이터 통합, e-서비스 접근과 콘텐츠 관리 등에 관한 기술을 제공하고 있고, 또한 2001년 5월에 공표된 e-정부 메타데이터 프레임워크에 상술된 정책들을 포함한다. 영국 정부는 e-GMS와 e-GIF 내에 정의된 메타데이터 스키마 및 정책의 빠른 확장과 수용을 위해 영국 내각 사무실이 지원하는 UK govTalk사이트를 설립하였다(<http://www.govtalk.gov.uk/>). 이 사이트는 각종 지침서, 정책 문서, 도구, XML 스키마, e-GMS, e-GIF 등에 관한 관련 문서들을 제공하고 있다.

4. 국가지식정보자원 메타데이터 표준화 발전방안의 수립

지식정보자원을 효율적이고 효과적인 활용을 위해서는 메타데이터의 체계적인 정책과 전략이 필요하다고 본다. 2000년도 정통부와 전산원의 주관으로 종합정보센터, 전문정보센터 관계자, 메타데이터 전문가로 구성된 메타데이터 실무 작업반에서 7 개의 메타데이터 필수요소를 제정하였다. 각 기관에서 이 표준을 성실히 적용하고 있는 것으로 판단되고 있으며, 그 요소들이 국가지식포털 사이트에서도 검색요소들로 활용되고 있다.

2000년도에 제정한 메타데이터 요소를 점검하는 과정에서 필수요소의 개념을 코어요소의

개념으로 전환하는 것이 필요하다고 판단되었다. 필수요소는 요소의 값을 빠짐없이 입력해야 되는 강제조항을 뜻하고, 코어요소는 기술 대상 자원이 그 요소의 값을 가질 수 있는 경우에는 한 꼭 입력해야 한다는 의미이다. 지식정보사업에서 필수요소를 선정하는 것보다 코어요소를 선정하는 것이 약간의 융통성을 부여하면서 메타데이터의 질을 일관성 있게 유지할 수 있는 방안이라고 본다. 또한 지식정보사업이 많은 분야에 걸쳐서 다양한 매체에 대한 국가지식관리사업을 지원하게 되는데 각 매체별로 코어요소를 표준으로 정하고 각 기관에 도움이 될 수 있는 선택요소들을 권고하는 것도 바람직하다고 본다.

4. 1 국가지식정보자원관리 코어 메타데이터의 수립 방안

국가지식정보자원의 코어 메타데이터 수립은 이전에 제정된 필수요소를 기반으로 각 기관에서 사용해본 경험을 반영하고, 각 국가에서 선정한 메타데이터 코어요소들을 참조할 것을 권장한다. 국가 지식정보사업 코어 메타데이터 요소의 선정은 지금까지 가장 널리 활용되어온 메타데이터 스키마를 기반으로 하는 것이 바람직하다고 본다. 이 범주에 해당하는 것은 더블링크어, KOMARC, 그리고 미국 의회 도서관에서 MARC와 호환성을 유지하면서 보다 보편적으로 메타데이터를 기술하기 위해서 개발한 MODS(Metadata Object Description Schema, URL: <http://www.loc.gov/standards/mods/>)를 참조하는 것도 권장한다. MODS Lite은 가볍게 MODS의 표준을

사용하고자 할 때 고려할 수 있는 MODS의 코어요소라고 볼 수 있기 때문에 MODS Lite를 정밀 분석할 필요가 있다. 현재 메타데이터 작업반이 수차례의 회의와 검토를 통해 완성한 국가지식정보자원 메타데이터 표준안 2.0의 초안은 다음과 같다.

4. 2 국가지식정보자원 분야 및 매체별 코어요소의 표준화 방안

매체별 특성을 고려하여 매체별 코어요소를 선정할 필요가 있다고 본다. 국가지식정보사업의 관리대상인 모든 매체들을 조사하고 각 매체별 국제 표준의 유무를 먼저 확인할 필요가 있다. 그 분야의 국제 메타데이터 표준이 존재할 경우에는 이를 적극 활용하고 국내 자원의 특수성 때문에 국제 표준의 적용이 불가능한 경우에는 새로운 요소를 정의하여 네임스페이스와 URI를 부과하여 사용할 것을 권한다. 이렇게 함으로써 국내에서 개발된 메타데이터 스키마와 응용 프로파일의 국제적으로 활용될 수 있는 하부구조를 조성해야 한다. 이를 위해서는 각 매체별로 가장 많은 서비스를 해본 경험이 있는 기관과 업체들, 국제 메타데이터 표준 현황에 대한 전문가, 시맨틱웹 기술의 전문가 등으로 구성된 매체별 Working Group을 조성할 필요가 있다. 이러한 결과물이 국제 표준 규격에 맞는 메타데이터의 유지와 관리 작업이 병행되면 메타데이터의 국제 표준에서 선도적인 역할을 할 수 있는 기반을 점차적으로 마련하게 될 것이다.

(표 1) 국가지식정보자원관리 메타데이터 표준안 2.0 초안

주요소	하위요소	필수	인코딩스킴	카디날리티	비고
1 title(제목)	mainTitle(대표제목)	yes		1..1	
	alternative(부가제목)	no		0..n	부제목, 번역제목, 등
2 creator(생성자)		yes		1..n	
3 subject(주제)		yes		1..n	통제어, 시소러스 사용 권장
4 type(자료유형)		yes		1..n	통제어 사용 권장
		yes	URL	1..1	
5 identifier(식별자)		no	UCI	0..1	
		no	DOI	0..1	
6 publisher(출판사)		yes		1..1	
7 date(날짜)	issued(발행일)				
	created(생성일) modified(갱신일)	yes	W3C-DT F		셋 중 하나는 필수
8 description(요약정보)	abstract(초록)				
	tableOfContents(목차정보)	yes			둘 중 하나는 필수
9 format(포맷)	extent(소요시간/파일크기)	no		0..1	
	medium(파일형식)	yes	IMT	1..1	
10 metaMetadata(관리메타데이터)	mdCenter(메타데이터센터명)	yes		1..1	
	recordUnit(메타데이터단위)	no		0..1	책, 기사, 보고서, 등

4.3 국가지식정보자원 분야 및 매체별 선택요소의 선정 방안

각 매체별로 코어요소를 결정한 후에 기관에서 사용할 수 있는 선택요소들을 매체별로 제정하는 것이 유익할지는 확실치 않다. 매체별 선택요소를 선정함으로써 얻는 유익은 선택요소들 간의 상호운용을 증진시킬 수 있다는 것이다. 또 하나의 방법은 매체별 선택요소는 자유로이 사용할 수 있도록 허락하지만 그 요소들의 이름과 상세한 정의를 중앙 Metadata Registry(이하, MDR)에 등록하는 것을 의무화하는 방법이다. 이런 방법을 통하여 제대로

정의된 요소가 현장에서 자연스럽게 활용되도록 유도하는 방식을 취할 수도 있다. 매체별 선택요소를 제정하고 새로운 요소를 기관에서 필요에 따라서 정의하고 등록하는 것을 동시에 의무화하는 방안이 가장 효율적이고 효과적이라고 생각되나, 이해 관계자들과의 협의가 필요한 상황이라고 판단된다.

4.4 더블링크어 네임스페이스와 URI 정책

국내에서도 시맨틱웹의 기반을 견고히 다지기 위해서는 무엇보다도 각 기관에서 네임스페이스와 URI 정책을 조속히 수립해야 한다. 이

러한 정책이 시행되면 기관 간의 메타데이터 방출은 네임스페이스와 URI 기반으로 될 것이기 때문에 의미충돌을 대부분 방지할 수 있으며 메타데이터의 요소가 기계가독형으로 표현되는 큰 장점을 갖추게 된다. 시맨틱웹 기반 네임스페이스와 URI 정책을 가장 모범적으로 적용하고 있는 것은 더블린코어 스키마이다(<http://www.dublincore.org/documents/dcmi-terms/>). 현재 더블린코어는 3 개의 네임스페이스를 사용하고 있다(Powell et al, 2001). 첫 번째 네임스페이스는 더블린코어의 주요소를 위한 것이다. 두 번째 네임스페이스는 한정어와 인코딩스킴을 관리하기 위한 것이고, 마지막 네임스페이스는 자료유형의 값을 정의하기 위한 것이다. 상세한 네임스페이스 정보는 아래와 같다.

- 주요소 네임스페이스: <http://purl.org/dc/elements/1.1/>
- 한정어, 새로 정의된 요소, 인코딩스킴의 네임스페이스: <http://purl.org/dc/terms/>
- 자료유형 네임스페이스: <http://purl.org/dc/dcmitype/>

주요소 네임스페이스에 속한 더블린코어 요소는 널리 알려진 15개 요소로 국한되어 있다. 더블린코어에 'audience' 라는 주요소가 최근에 첨가되었는데 이것이 주요소의 네임스페이스로 배정되지 않았고, 한정어와 인코딩스킴의 네임스페이스(<http://purl.org/dc/terms/>)로 배정되었다. 여기에서 우리가 파악할 수 있는 사실은 더블린코어에서 앞으로 새로 제정하는 더블린코어 요소는 주요소나 한정어에 관계

없이 <http://purl.org/dc/terms/> 네임스페이스에 할당하겠다는 정책을 세운 것으로 판단된다. 다음은 네임스페이스별로 더블린코어 요소와 인코딩스킴, 자료유형의 값들을 정리한 것이다.

(<http://purl.org/dc/elements/1.1/>)

네임스페이스 속하는 요소들의 URI

- 표제 URI: <http://purl.org/dc/elements/1.1/title>
- 생성자 URI: <http://purl.org/dc/elements/1.1/creator>
- 주제 URI: <http://purl.org/dc/elements/1.1/subject>
- 출판처 URI: <http://purl.org/dc/elements/1.1/publisher>
- 요약정보 URI: <http://purl.org/dc/elements/1.1/description>
- 기여자 URI: <http://purl.org/dc/elements/1.1/contributor>
- 날짜 URI: <http://purl.org/dc/elements/1.1/date>
- 출처 URI: <http://purl.org/dc/elements/1.1/source>
- 식별자 URI: <http://purl.org/dc/elements/1.1/identifier>
- 범위 URI: <http://purl.org/dc/elements/1.1/coverage>
- 이용조건 URI: <http://purl.org/dc/elements/1.1/rights>
- 언어 URI: <http://purl.org/dc/elements/1.1/language>
- 포맷 URI: <http://purl.org/dc/element>

s/1.1/format

- 자료유형 URI: <http://purl.org/dc/elements/1.1/type>
- 관계정보 URI: <http://purl.org/dc/elements/1.1/relation>

위와 같이 네임스페이스와 URI를 활용하면 메타데이터 요소의 식별이 기계가독형으로 정의되고 메타데이터 요소의 이름과 의미가 충돌이 없이 질서정연하게 사용될 수 있다. 모든 메타데이터 요소에 네임스페이스가 부착되기 때문에 어느 기관에서 정의한 요소라는 것이 분명하게 밝혀지고, 또한 이 URI는 고유성과 영속성이 보장되기 때문에 시맨틱웹의 기반을 제공하게 된다. 같은 맥락에서 다음은 더블린 코어의 하위요소와 새로이 정의되는 요소들의 URI를 소개한다.

(<http://purl.org/dc/terms/>) 네임스페이스 속하는 요소들의 URI

- 대체포제 URI: <http://purl.org/dc/terms/alternative>
- 목차정보 URI: <http://purl.org/dc/terms/tableOfContents>
- 초록정보 URI: <http://purl.org/dc/terms/abstract>
- 제작일 URI: <http://purl.org/dc/terms/created>
- 유효일 URI: <http://purl.org/dc/terms/valid>
- 이용가능일 URI: <http://purl.org/dc/terms/available>
- 발행일 URI: <http://purl.org/dc/terms/>

issued

- 수정일 URI: <http://purl.org/dc/terms/modified>
- 수용일 URI: <http://purl.org/dc/terms/dateAccepted>
- 지적재산권취득일 URI: <http://purl.org/dc/terms/dateCopyrighted>
- 제출일 URI: <http://purl.org/dc/terms/dateSubmitted>
- 서지정보 URI: <http://purl.org/terms/bibliographicCitation>
- 전체/부분관계 URI: <http://purl.org/dc/terms/isPartOf>
- 전체/부분관계 URI: <http://purl.org/dc/terms/hasPart>
- 버전관계 URI: <http://purl.org/dc/terms/isVersionOf>
- 버전관계 URI: <http://purl.org/dc/terms/hasVersion>
- 포맷변환관계 URI: <http://purl.org/dc/terms/isFormatOf>
- 포맷변환관계 URI: <http://purl.org/dc/terms/hasFormat>
- 참조관계 URI: <http://purl.org/dc/terms/isReferencedBy>
- 참조관계 URI: <http://purl.org/dc/terms/references>
- 의존관계 URI: <http://purl.org/dc/terms/isRequiredBy>
- 의존관계 URI: <http://purl.org/dc/terms/requires>
- 대체관계 URI: <http://purl.org/dc/terms/isReplacedBy>

- 대체관계 URI: <http://purl.org/dc/terms/replaces>
- 준수관계 URI: <http://purl.org/dc/terms/conformsTo>
- 공간적 범위 URI: <http://purl.org/dc/terms/spatial>
- 시간적 범위 URI: <http://purl.org/dc/terms/temporal>
- 지적재산권 접근 URI: <http://purl.org/dc/terms/rightsAccess>
- 교육수준 URI: <http://purl.org/dc/terms/educationLevel>
- 대상자 URI: <http://purl.org/dc/terms/mediator>

다음은 더블린코어에서 활용하고 있는 모든 인코딩 스킴의 이름들에 대한 URI들을 소개한다. 인코딩 스킴명의 네임스페이스는 하위요소와 새로 정의되는 요소들의 네임스페이스와 동일하다. 결론적으로 말하면, 더블린코어에서는 모든 메타데이터 요소는 주요소와 하위요소를 동일한 네임스페이스에서 관리하기로 결정했고, 또한 인코딩 스킴도 요소의 네임스페이스 안에서 같이 관리하기로 정책을 정했다는 것이다.

(<http://purl.org/dc/terms/>) 네임스페이스 속하는 스킴들의 URI

- URI 스킴의 URI: <http://purl.org/dc/terms/URI>
- 주제 DDC 스킴 URI: <http://purl.org/dc/terms/DDC>
- 주제 LCC 스킴 URI: <http://purl.org/dc/terms/LCC>

- 주제 LCSH 스킴 URI: <http://purl.org/dc/terms/LCSH>
- 주제 MeSH 스킴 URI: <http://purl.org/dc/terms/MeSH>
- 주제 UDC 스킴 URI: <http://purl.org/dc/terms/UDC>
- 포맷 IMT 스킴 URI: <http://purl.org/dc/terms/IMT>
- 자료유형 DCMIType 스킴 URI: <http://purl.org/dc/terms/DCMIType>
- 지명 TGN 스킴 URI: <http://purl.org/dc/terms/TGN>
- 국가명 ISO3166 스킴 URI: <http://purl.org/dc/terms/ISO3166>
- 공간영역 DCMI Box 스킴 URI: <http://purl.org/dc/terms/Box>
- 공간영역 DCMI Point 스킴 URI: <http://purl.org/dc/terms/Point>
- 시간범위 DCMI Period 스킴 URI: <http://purl.org/dc/terms/Period>
- 날짜 W3CDTF 스킴 URI: <http://purl.org/dc/terms/W3CDTF>
- 언어 ISO639-2 스킴 URI: <http://purl.org/dc/terms/ISO639-2>
- 언어 RFC3066 스킴 URI: <http://purl.org/dc/terms/RFC3066>

또한 네임스페이스와 URI는 메타데이터 요소를 구분하기 위해서만 사용하는 것이 아니라, 온톨로지 구축에도 사용되어야 한다. 각 통제어 리스트나 온톨로지는 하나의 네임스페이스를 할당하는 것이 바람직하다고 본다. 앞으로 많은 온톨로지가 구축될 것으로 기대되고

있기 때문에 각 온톨로지는 고유의 네임스페이스가 할당될 것으로 기대된다. 국내에서도 온톨로지를 구축할 때 고유의 네임스페이스를 적용하여 어느 기관에서 혹은 어느 분야에서 개발한 온톨로지인지와 책임 여부를 투명하게 관리하여 Best Practice가 점점 확산되도록 기반을 마련해야 한다.

(<http://purl.org/dc/mitypes/>) 네임스페이스 속하는 자료유형 값들의 URI

- Collection URI: <http://purl.org/dc/dcmitype/Collection>
- Dataset URI: <http://purl.org/dc/dcmitype/Dataset>
- Event URI: <http://purl.org/dc/dcmitype/Event>
- Image URI: <http://purl.org/dc/dcmitype/Image>
- InteractiveResource URI: <http://purl.org/dc/dcmitype/InteractiveResource>
- MovingImage URI: <http://purl.org/dc/dcmitype/MovingImage>
- PhysicalObject URI: <http://purl.org/dc/dcmitype/PhysicalObject>
- Service URI: <http://purl.org/dc/dcmitype/Service>
- Software URI: <http://purl.org/dc/dcmitype/Software>
- Sound URI: <http://purl.org/dc/dcmitype/Sound>
- StillImage URI: <http://purl.org/dc/dcmitype/StillImage>
- Text URI: <http://purl.org/dc/dcmitype/Text>

pe/Text

지금까지 언급한대로 각 기관에서 모든 메타데이터 요소에 고유의 URI과 XML 네임스페이스로 관리를 할 경우에 시맨틱웹을 구축하기 위한 하부구조를 이루는데 큰 도움이 되는 것은 사실이다. 그러나 각 기관에서 다양하게 설계되고 있는 모든 메타데이터 스키마에 대해서 다른 기관에서 참조할 수 있게 함으로써 동일한 의미의 메타데이터 요소가 다르게 정의되는 것을 방지하고 상호 운용을 증진시키기 위해서는 ISO 11179에 기반을 둔 국가 MDR을 구축할 필요가 있다.

5. 국가지식정보자원관리를 위한 MDR 설계 및 구축 방안

각 기관에서 사용하고 있는 메타데이터 요소가 제대로 관리되고 있지 않기 때문에 정보의 공유와 데이터 통합 기능을 수행하는데 많은 장애를 받고 있는 실정이다. 이런 손실을 막기 위해서 국제적으로 ISO 하부위원회(SC 32)에서 1998년부터 모든 기관에서 설계하는 메타데이터의 의미(semantics)를 체계적으로 관리할 수 있는 표준을 제정하였고, 이에 기반을 둔 메타데이터 등록기의 구축이 확산되고 있다. 가장 대표적인 ISO 11179 기반 메타데이터 등록기로는 미국 환경청에서 구축한 Environmental Data Registry(<http://www.epa.gov/edr/>), 영국 UKOLN에서 개발한 MDR(<http://www.ukoln.ac.uk/metadata/>), 더블린코어 MDR 등이다. 이렇게 국제 표준에 맞는 형식으로 국

내의 메타데이터를 관리할 수 있는 중앙등록기의 구축이 시급하다고 본다. 다음은 ISO 11179에 대한 핵심내용을 소개한다.

5. 1 MDR 국제 표준

ISO/IEC 11179 MDR 국제표준은 데이터의 의미(semantics)를 표현, 등록, 관리, 교환, 공유하는데 목적이 있다. MDR은 특히 전 세계적 데이터 교환을 가능케 하는 방법 부재로 인한 어려움, 데이터 요소(data element)를 식별하는데 쓰이는 식별자(identifier) 문제, 데이터의 자동 공유, 추출, 교환을 용이하게 하는 데이터 요소 특성을 체계적으로 문서화하는 문제, 데이터 요소와 값의 영역(value domain)에 대한 식별의 문제, 개발과 서술에 대한 일관된 지침의 부재로 인한 문제, 무수히 많은 데이터 요소나 값의 영역 중에서 특정 데이터 요소나 값의 영역을 추출하는 방법의 부재와 불충분으로 인한 문제, 데이터 요소나 값의 영역을 체계화를 위한 보편적 방법의 부재로 인한 문제, 기관과 기업간의 데이터 표준의 부재로 인한 문제, 표준화된 데이터의 부재로 인해 조직과 기업간의 데이터 교환 시 새로운 형태의 데이터 교환 표현 방식의 출현 문제, 데이터의 재사용에는 부적합한 데이터 정의와 기술, 10) 데이터 요소나 값의 영역에 대한 기술이 부적당하여, 현재 XML을 사용하여 진행되고 있는 글로벌 전자 데이터 교환이 원활하지 못한 문제 등을 해결하는 것을 목표로 하고 있다.

정보의 의미 표현과 공유를 가능케 하는 MDR은 등록된 정보간의 상호운용성의 확보에 중요한 개념적 기반을 제공하며, 실제적으

로 동일한 정보의 중복과 구문적, 문자적 차이를 식별할 수 있게 함으로써, 의미적으로 동일한 요소들 간의 통합이 이뤄질 수 있게 한다. MDR은 다음의 6개의 하위 표준 문서로 구성되어 있다. 1장은 전반적인 MDR 표준에 관한 프레임워크, 2장은 데이터요소의 분류, 3장은 메타모델과 기본속성, 4장은 관리대상 개체들의 정의를 기술할 때 따라야 할 지침, 5장은 관리대상의 명명과 식별체계, 6장은 등록체계 등으로 구성되어 있다. 이 ISO 11179 표준의 목적은 각 기관에서 정의하여 사용하고 있는 모든 메타데이터 요소들 간의 의미적 상호 운용을 극대화하기 위한 것이다.

5. 2 더블린코어 MDR

더블린코어 MDR은 이용자가 더블린코어 메타데이터의 어휘를 검색할 수 있게 하는 애플리케이션이다. 이용자는 손쉽게 용어와 정의를 검색할 수 있고, 이들 용어간의 관계를 볼 수 있다. 이 MDR의 목적은 이미 정의되어 있는 요소의 의미 발견, 재사용과 확장과 새로운 어휘의 생성을 용이하게 하는데 있다. 더블린코어 MDR은 2000년 9월 Phase I 개발이 종료되어 <<http://dublincore.org/dcregistry/index.html>>에서 서비스되고 있다. 현재 23개의 언어를 지원하며, 개발 소스 코드는 <http://wip.dublincore.org>를 통하여 서비스되고 있다. Phase I은 인간가독형이고 다중언어 인터넷 인터페이스를 제공하여 더블린코어 용어를 검색하고 네비게이션할 수 있는 기능을 제공하고 있다. Phase I MDR은 RDF 스키마와 XML 파일을 입력으로 받는 Jena 도구를 이

용한 자바 애플리케이션이며 PostgreSQL 데이터베이스를 이용하고 있다. 출력은 XSLT를 이용하여 XHTML로 변환된 XML 형식을 취하고 있다. Phase II 개발은 어휘관리 도구, 레지스트리 간의 인터페이스, 애플리케이션 API, 번역 추가, 인터페이스 스타일 분리 등의 개발에 중점을 두었다. Phase II는 2003년 5월부터 서비스되고 있다.

더블린코어 MDR은 다양한 서블릿(servlet)으로 이루어져 있다 - 초기화, 쿠키 관리, 변환 서비스 등을 제공하는 MDR 서블릿, 상요자의 선호사항을 관장하는 서블릿, 검색과 네비게이션을 위한 질의 서블릿, 질의 모드를 위한 서블릿이 있다. 이용자 인터페이스는 표준인터페이스와 RDF 인터페이스를 지원하며, 검색과 네비게이션 기능, 관련 용어의 링크 기능, 영어 번역기능으로의 링크 등을 제공한다. 검색 기능은 전문 검색, 요소 및 한정자로의 링크, 용어간의 관계에 기반을 둔 네비게이션 등이 제공되고 있다. 현재 더블린코어 MDR 개발에 있어서 가장 중요한 것은 분산 레지스트리의 개발에 있다. 분산 레지스트리는 현재의 중앙집중식 레지스트리 모델에 기반을 두어 개발된 것이다. 분산 레지스트리는 조직에 중요한 정보의 접근을 개선시켜, 특정 도메인과 지역에 국한된 용어, 확장어, 애플리케이션 프로파일, 문서, 분류 스키마 등에 접근할 수 있게 한다는 것이다.

5. 3 국가지식정보자원관리를 위한 MDR 설계 방안

지식자원이 양적으로 증가하고 유형이 다양

화하고 있는 현 정보사회에서는 그 지식자원들의 단순한 축적과 수집과 니열을 넘어 보다 효율적으로 사용하고 관리하는 활용성의 측면이 부각되고 있다. 국가 MDR의 구축을 통해 지식정보자원의 효율적 관리는 지식사회를 지향하는 모든 국가의 시급한 과제로 부상되었고, 지식정보자원 MDR에 관한 연구는 차세대 정보 선진국으로의 도약에 중요한 과제로 도출되었다. 위에서 언급했듯이 이런 필요성에 대응하기 위하여, 국제표준협회 산하의 32 하부위원회(ISO/IEC JTC1 SC32)에서 연구한 결과로 제시된 ISO 11179 표준은 데이터 구체화를 위한 것으로 열린 정보 시스템에서 데이터의 공유를 가능하게 하는 기본적인 특징들을 구체화하기 위한 노력이다. 데이터가 공유된다는 것은 데이터의 이용자와 소유자가 그 의미와 표현, 식별에 대한 공통의 이해를 가지고 있다는 것을 의미한다. 이런 의미에서 ISO 11179 표준 문서는 데이터요소(data element)의 식별, 명명, 정의, 개념, 값, 분류 등에 중점을 두고, 교환될 데이터의 이용자가 의미, 표현형식, 및 식별에 대한 공통된 이해를 갖기 위한 메타데이터 등록(metadata registration)과 유지를 원활히 수행하기 위한 지침을 제공하고 있다.

ISO 11179와 더불어 W3C의 시맨틱웹의 활동은 산재한 정보자원을 기계가독형으로 기술할 수 있는 기반을 조성함으로써 웹상의 다양한 정보를 지식화 하기 위한 움직임이다. 이 시맨틱웹의 기반은 모든 자원, 클래스, 속성 등에 불변의 고유 식별자를 부과하는 URI 개념, 이런 식별자들을 군집화해서 관리하는 네임스페이스 개념과 상이한 시스템에 저장된 정보의

통합을 용이하게 처리할 수 있는 XML 기술이 그 기반을 이루고 있다. 여기에 메타데이터 간의 의미적 상호운용성을 증진시키기 위해서 개발된 RDF, RDF 스키마, 웹온톨로지언어(OWL) 등의 새로운 접근 방법들이 현존하는 다양한 메타데이터 레지스트리 간의 의미적 상호운용을 증진시키기 위해 개발되고 있다. 미래의 MDR은 URI와 XML 네임스페이스 개념이 적극 도입될 것으로 전망되고 있다.

ISO 11179 표준은 각 기관에서 사용하고 있는 모든 데이터 요소의 의미를 체계적으로 관리하기 위한 국제표준이기 때문에 이 표준은 국내 지식정보자원관리에도 중요한 의미를 지니고 있다고 본다. 이 표준은 상세한 의미관리 체계를 제시하고 있으나 이 표준의 모든 부분으로 다 적용해야 한다는 의무조항은 없다. ISO 11179에 제시된 모든 데이터 요소의 정의, 명명, 등록하는 의무를 각 기관에 부여하는 것은 현재 국내의 인력구조로는 어려운 일이라 판단된다. 국내 MDR 설계는 융통성과 확장성을 고려하여 국제 표준을 따르면서도 국내 실정을 반영하는 전략을 세워야 한다. 먼저 ISO 11179에서 핵심 부분을 도출하여 등록 절차를 간소화 하여 기관과 업체의 부담을 줄이면서 MDR을 활용해서 가능한 한 빠른 시일 내에 국내지식자원의 데이터요소 의미관리 체계를 수립하는 것이다. 또한 미래의 확장성을 고려하여 MDR 시스템은 ISO 11179의 모든 요소를 수용할 수 있도록 설계되어야 하고, 각 기관에서 사용하는 메타데이터 요소를 등록할 때 어떤 사항을 의무적으로 할 것인가에 대해서는 선별적 기준을 적용해야 한다. 이러한 방식으로 각 기관에서 사용하고 있는 데이터가 정의

되고 관리된다면 데이터요소에 관한 참조와 재사용이 용이해지기 때문에 데이터 상호 운용을 크게 증진시킬 수 있을 것이다. 현재 미국에서는 이 표준에 근거한 메타데이터 등록시스템이 분야별로 개발되어 운영되기 시작하는 단계에 있고 더욱더 활성화될 전망이다. 특히, 국가지식정보자원관리를 위한 MDR의 구축은 영국의 DESIRE MDR과 더블링크어 MDR의 구조가 이미 ISO 11179에서 핵심 요소를 추출해서 구축한 예이기 때문에 이런 시스템을 기반으로 할 것을 권한다.

시맨틱웹 기반 MDR의 설계를 위해서는 ISO 11179 MDR 표준과 W3C의 OWL과 ISO13250의 Topic Map은 통합시키는 방법을 생각해 볼 수 있다. ISO 11179 MDR을 통한 인간가독형 데이터요소의 의미관리와 OWL과 Topic Map의 기계가독형 의미 표현력(semantic expressiveness)의 결합은 지식정보자원 MDR 간의 의미적 상호 운용을 위한 기술적 기반을 제공할 것으로 기대되고 있다.

6. 시맨틱웹 기반 지능형 검색 서비스

MDR에 등록된 정보를 시맨틱웹 기술을 적용하여 기계가독형으로 변환하면 다양한 지능형 검색을 지원할 수 있을 것으로 기대된다. 모든 클래스들 간의 상·하 관계와 유사관계, 요소들 간의 상·하 관계와 유사관계, 모든 클래스와 요소들의 기계가독형 정의, 모든 요소들이 취할 수 있는 값들의 영역과 범위에 대한 체계적 정의 등을 활용할 수 있을 것으로 판단된다.

6. 1 MDR을 활용한 검색 서비스

MDR의 이용은 인터넷과 기타 정보 시스템 내에 기술된 다양한 종류의 정보를 통합하는 기반을 제공한다. MDR에 저장되는 메타데이터는 단일 메타데이터, 복합 메타데이터, 혼성 메타데이터의 3가지로 구분될 수 있다. 단일 메타데이터 방식은 의미 기술에 사용되는 공통어휘를 제공하는 하나의 메타데이터 집합을 제공한다. 모든 정보는 이 단일 메타데이터에 의해 기술되어야 한다. 단일 메타데이터 방식은 도메인에 대한 같은 관점을 제공하는 정보자원의 통합이나 정보자원에 대한 제한된 기술만이 요구되는 환경에 유용하게 쓰일 수 있다. 그러나 단일 메타데이터 방식은, 특정 정보 자원이 메타데이터에서 지원하지 않는 정보나 관점을 표현할 때나, 정보 기술의 상세성(granularity)의 수준이 단일 메타데이터에서 지원되는 범위를 넘어서는 때 효과적이지 못하다. 또한 단일 메타데이터 방식은 레지스트리에 참여하는 각 기관이 단일 메타데이터에 맞게 기존의 메타데이터를 수정해야 한다.

복합 메타데이터 방식에서는, 각 기관이 필요에 따라 최적의 메타데이터를 정의하거나 이미 표준으로 사용되는 메타데이터를 사용한다. 복합 메타데이터를 사용하면, 각 기관의 메타데이터는 다른 메타데이터와는 별도로 개발되거나 표준화된 메타데이터를 사용할 수 있다. 이 방식을 이용하면 정보의 변화에 유연하게 대처할 수 있다. 즉, 정보의 변화가 특정 소스 메타데이터에 국한된다. 그러나 실제로, 이 방식은 공통된 메타데이터의 부재 때문에 소스 메타데이터 간의 비교는 매우 어려운 작업이

며, 이를 극복하기 위한 메타데이터간의 사상(mapping) 방법이 고려되어야 한다.

혼성 메타데이터 방식은 복합 메타데이터 방식의 기반위에 공통 어휘를 사용한다. 혼성 메타데이터의 장점은, 새로운 정보 자원이 공유 어휘나 사상(mapping) 방법의 변화 없이도 쉽게 더해질 수 있다는 것이다. 또한, 이 방식은 메타데이터의 발전(evolution)을 지원하며, 공통어휘의 사용으로 소스 메타데이터간의 호환성이 높다. 메타데이터 레지스트리의 이용과 혼성 메타데이터 방식은, 각각의 정보 자원이 도메인에 최적의 메타데이터에 의해 기술될 수 있게 함은 물론, 이종 메타데이터 간의 의미 비교를 통한 정보 통합을 가능케 한다. 모든 메타데이터 요소가 XML 네임스페이스를 활용한 URI를 부착하여 MDR에 등록되어 있으면 각 기관들 간의 상호운용성이 크게 진전될 것이고, 검색질의의 축소와 확장서비스도 다양하게 제공될 것이다. 다양한 메타데이터 요소들간의 관계가 분명하게 정의되기 때문에 데이터 마이닝에도 큰 도움이 될 것으로 판단된다.

6. 2 시맨틱웹 기반 온톨로지 구축 방안

시맨틱웹은 인터넷 정보의 기하급수적 증가와 다량의 정보 분석을 위한 하부구조의 부재로 시스템 간의 커뮤니케이션이 원활하지 못함을 극복하기 위한 여러 기술과 노력을 지칭한다. 시맨틱웹은 자원의 의미를 체계적으로 정의하여 컴퓨터와 사람의 협력적 운용을 유도하기위해 정보의 명확한 정의와 연결에 기반을 두어 효율적 검색, 자동화, 통합, 재사용을 시

도한다. 이를 위하여 시맨틱웹은 상용화된 태깅 스킴(tagging scheme)을 정의하기위한 XML의 능력과 데이터를 표현하기 위한 RDF 기술 기반 위에 구현될 것이다. RDF는 자원에 대한 메타데이터를 기술할 때 특정 메타데이터 스키마에 구애 받지 않고 일반적으로 기술할 수 있게 하는 프레임워크이어서, 각 기관에서 독자적으로 자원에 대한 기술을 할지라도 상호운용성을 증진시킬 수 있는 방법이다. 여기에 각 메타데이터의 스키마의 클래스와 속성을 정의하는데 필요한 어휘를 갖추기 위해서 RDF 스키마가 제정되었다.

시맨틱웹은 시맨틱웹의 핵심 개념(URI, XML 네임스페이스)과 이 개념들에 기반을 둔 시맨틱웹의 기반구조 기술인 RDF, RDF 스키마, Web Ontology Language(이하, OWL)로 이루어져 있다. 이 중 RDF, RDF 스키마와 OWL은 기계가독과 사용되는 용어의 의미적 상호운용을 위해 개발되었다. OWL은 웹문서에서 사용된 용어의 의미를 형식적으로 기술할 수 있는 온톨로지 언어이고, 이 언어는 RDF 스키마보다 의미 표현력 측면에서 더욱 정교하다 할 수 있다. 온톨로지는 지식을 표현하는 방법이고, XML 구문으로 표현되는 것과 같은 문서교환을 위한 메시지 포맷이 아니다. OWL은 시맨틱웹의 핵심 기술로서 OWL 온톨로지에 관련한 추론 도구의 개발이 활발히 진행되고 있다. OWL은 부가적, 의미적 어휘를 제공함으로써 XML, RDF, RDF 스키마가 기술하는 웹 콘텐츠를 기계가 처리할 수 있는 능력을 제공한다. OWL은 문서 안에 정의된 어휘, 즉, 용어의 의미를 보다 더 분명하게 표현하고 그 용어간의 관계를 표현하기위

해 사용되며, 이러한 용어의 표현과 그 상호관계를 온톨로지라 한다. OWL은 XML, RDF, RDF 스키마 보다 의미를 더 잘 표현하도록 함으로써, 웹에서 기계가 읽을 수 있는 콘텐츠를 표현하는 능력에 있어 다른 언어보다 뛰어나다(Stevens 2003; Smith 2002).

시맨틱웹의 목적인 지능적 인터넷을 구현하기 위해서는, 용어를 기계가 가독할 수 있고, 정의된 용어의 의미와 용어간의 관계가 명확히 도출되어 애플리케이션이 의미에 기반을 둔 추론과 적용이 가능한 기술을 적용해야 한다. OWL은 이러한 목적에 적합한 기능을 제공한다. 또한 OWL은 메타데이터 온톨로지 간의 상호운용성을 촉진하고 이러한 온톨로지를 기반을 둔 새로운 지능적 정보 서비스의 기반을 제공한다는 관점에서 중요한 기반 기술이다.

6. 3 온톨로지를 활용한 지능형 검색서비스의 개발 방향

기존의 인터넷 검색 서비스는 색인 된 키워드에 기반을 둔 검색이 주종을 이루었다. 사용자는 데이터베이스에 키워드나 여과된 값, 혹은 블리언 연산자를 이용하여 정보 소스에 대한 검색을 실행한다. 인터넷에 정보량이 많아짐에 따라 텍스트의 사상에 거의 의존하는 키워드 기반 검색은 검색되는 양을 너무 많아졌고 정확율은 현저하게 저하 되는 현상이 뚜렷하여 이에 대한 개선이 시급한 실정이다.

미래의 검색 방향은 온톨로지를 활용한지능형 검색을 추구하는 것이다. 온톨로지를 이용하면, 사용자는 보다 명확하고 지능적인 질의를 기술할 수 있고, 온톨로지 클래스 정의와

추론 메커니즘(속성 상속 등)을 이용하여 사용자에게 데이터의 의미와 상호 관계를 보다 풍부하고 명확하게 제시할 수 있다(Garshol 2004). 도메인별로 정교한 온톨로지가 개발되면 다양한 맞춤형 정보서비스가 가능해질 것이다.

6. 4 분야별 온톨로지 설계 및 구축 방안: OWL과 Topic Maps 적용

지식자원이 양적으로 증가하고 유형이 다양화하고 있는 현 정보사회에서는, 그 지식자원들의 단순한 축적과 수집, 나열을 넘어, 보다 효율적으로 사용하고 관리하는 활용성의 측면이 부각되고 있다. 토픽맵(Topic Map)은 정보를 정교하게 분류 및 연결하고 배분, 이용할 수 있게 해 줌으로써 최적의 정보 이용을 가능케 하는 기술이다. 적절히 개발된 토픽맵은 기관 또는 회사에서 생성되는 유용한 지식과 정보에 다양하게 접근하여 정보객체들에 포함된 콘텐츠, 데이터를 주제별로 연결하여, 연관된 정보를 서로 긴밀히 연결된 상태로 볼 수 있게 한다. 또한 동일한 자원에 대해서도 다수의 토픽맵을 적용함으로써 이용자의 인지적 구조를 반영한 차별형 지식체계를 제공할 수도 있다. 현재, 토픽맵은 국제표준(ISO 13250)으로 제정되어 미연방정부 에너지부, 국제청, 유럽정부(노르웨이의 과학기술 분야 연구정보 국가 지식포털 사이트, IT교육, 소비자정보, 음식정보, 부동산정보, 정당정보) 등에서 정보사이트에 주체중심의 검색 및 지식관리에 핵심적인 도구로 등장하고 있다(Pepper and Garshol 2004). 토픽맵은 다양한 애플리케이션 분야에 적용할

수 있다.

시맨틱의 핵심 기술로 부각되고 있는 W3C의 OWL과 ISO 표준인 Topic Map은 서로 경쟁관계로 보는 것보다는 보완관계로 봐야 한다. 두 기술이 모두 XML 구문구조를 사용하고 있으며 네임스페이스와 URI 개념을 적용하고 있기 때문에 서로의 장점을 취하는 통합에 관한 연구도 활발히 진행되고 있다. 국내 지식정보사업에서도 이 두 기술에 대한 균형된 관심과 지원이 필요하다고 본다. Topic Map에 기반을 둔 온톨로지의 구축으로 혜택을 받을 분야를 선정하여 체계적으로 온톨로지를 구축하여 다양한 지능형 검색서비스를 시도할 수 있는 기반을 마련해야 한다.

OWL의 장점은 종전의 인공지능에서 추구하던 논리적 추론(inference)을 강하게 지원하면서 그 표현에 대한 구문구조를 국제표준으로 제정했다는 것이다. 메타데이터 요소들 간의 관계, 온톨로지 개념들간의 관계, 카디날리티 제어기능, 논리적 표현력 등을 활용하여 다양한 형태의 추론을 가능케 한다. 그러나 정보와 지식관리의 차원에서 그러한 기능이 얼마나 요구되는지는 별로 아는 것이 없는 상태이며, 앞으로 많은 연구와 이용자 검증을 거쳐야 할 부분이라고 판단된다.

이에 반하여, Topic Map은 이미 유용성이 입증된 책 색인정보를 전자화하기 위한 노력으로 시작되었고, 그 기본적인 기능을 바탕으로 필요한 요소들은 추가하여 ISO 13250 표준을 제정하기에 이르렀다. Topic Map은 이미 개발된 통제어, 텍사노미, 시소러스 등을 한곳으로 통합하면서도 신축성과 확장성을 제공하기 때문에 문헌정보학에 아주 중요한 도구라고 생

각한다. 정보와 지식을 분류하는 일은 한번에 완성하고 종료되는 작업이라 생각되지 않고, 지속적으로 이용자 연구에 의해서 수정과 보완 작업을 거쳐야 되는데, 종전의 방식은 이러한 수정과 보완 작업을 원활히 지원하지 못했다. Topic Map은 토픽(개념)과 토픽들 간의 관계를 직접 연결하고 토픽의 인스턴스까지도 다 연결된 상태이기 때문에 토픽을 재배치하면 그 인스턴스가 함께 따라가게 되어 있고 필요하다면 인스턴스 수정도 물론 가능하다. 이러한 기능은 Topic Map을 수정하면 새로운 서비스를 제공하는 결과를 가져오기 때문에, 유럽과 미국의 여러 기관과 회사에서는 지식관리와 지식포털 사이트를 Topic Map 기반으로 구축하는 곳들이 등장하고 있다. 이런 관점에서 본다면 우리 분야가 Topic Map 표준에 더 많은 관심을 기울여야 하고, 온톨로지 구축이 문헌정보학의 중요한 연구 분야 중의 하나가 되어야 한다.

7. 결 론

다음은 지금까지 논의한 내용 중에서 지식자원관리센터, 종합정보센터, 전문정보센터의 역할과 미래의 방향에 대한 핵심 사항을 정리한 것이다.

- 국가지식정보자원의 메타데이터를 체계적으로 관리를 위해서 ISO 11179 표준에 근거한 중앙 MDR을 구축한다. 만약 각 종합정보센터가 자체 MDR을 구축하고자 할 경우에는 ISO 11179의 표준을 준수해야 하고, 종합정보센터 MDR에 등록
- 된 것은 중앙 MDR에도 자동으로 등록되도록 해야 한다.
- 국가통합검색이 현재 검색속도가 너무 느린 단점을 극복하기 위해서는 메타데이터를 통합적으로 저장할 수 있는 MR이 필요하다. 국가지식정보사업으로 지원되는 과제로 생성되는 모든 메타데이터는 종합정보센터의 MR에 방출하는 것을 의무화해야 한다. MR에 저장된 메타데이터는 어느 기관에서든지 자유로이 사용하여 서비스를 개발할 수 있는 것을 원칙으로 한다.
- 현재 국제적으로 사용되고 있는 메타데이터 관련 모든 네임스페이스와 응용 프로파일을 중앙 MDR을 통해서 각 기관이 참조할 수 있도록 해야 한다.
- 국가지식정보사업 참여 기관은 적어도 하나의 네임스페이스 확보하고, 그 기관에서 새로 정의하는 메타데이터 요소는 네임스페이스와 연결하여 고유의 URI를 부과시켜야 한다.
- 응용 프로파일을 설계할 때, 국제 표준 메타데이터를 수용하고 한국 자료의 특성 때문에 국제 표준그룹이 고려하지 못한 요소에 대해서는 새로운 요소를 정의할 수 있으며, 각 기관이 사용하고 있는 모든 메타데이터 요소는 중앙 MDR이나 종합정보센터 MDR에 등록하는 것을 의무화한다. 새로운 요소의 확장은 우선 기존 요소의 하위요소로 확장할 것을 먼저 고려하고, 그것이 불가할 경우에는 새로운 주요 요소를 제정한다.
- 새로 정의한 요소가 국제 표준 메타데이터 요소와 어떤 관계가 있는 것들은 중앙

MDR이나 종합정보센터 MDR에 등록할 때, 이 관계 정보(상·하, 동의어, 역, 교집합, 합집합)를 명시해야 한다.

- 국가지식자원관리센터는 참여기관에 중앙 MDR에 등록된 모든 정보를 공개하고 이용할 권한을 부여하여, Best Practice가 자연적으로 확산되도록 제도적 장치를 마련한다.
- 국가지식자원관리센터는 중앙 MDR에 등록된 정보는 이미 시맨틱웹의 기반을 다진 것이고, 보다 등록된 정보의 의미를 명확하게 표현하기 위해서 시맨틱웹 기반으로 변환할 연구를 수행하고, 점차로 시맨틱웹 기반 서비스를 제공하는 것을 목표로 한다. 이 때 적용할 시맨틱웹 핵심기술은 W3C의 RDF/OWL 접근과 ISO의 Topic Map을 사용한 접근을 고려해야 한다. 이 두 기술은 경쟁관계에 있기도 하지만, 상호보완적으로 보아야 하고 결국 통합되어야 한다.
- 마지막으로 가장 힘들지만 중요한 작업이 기계가독형 온톨로지를 생성하는 것이다.

각 분야의 전문가들과 온톨로지 전문가들이 한 팀이 되어 온톨로지 생성 작업을 수행해야 한다. 온톨로지 없이는 고급 지능형 검색은 어렵다고 본다. 지식정보사업 참여 기관은 공동의 노력으로 시맨틱웹 기반 온톨로지를 구축하는 일에 협조할 필요가 있고, 분야를 나누어서 온톨로지를 구축하는 것이 바람직하다. 온톨로지의 생성도 RDF/OWL을 사용할 수도 있고, Topic Map을 사용할 수도 있다. 둘 다 기계가독형으로 기술되기 때문에 검색엔진이 활용하는 데는 아무런 지장이 없을 것이다.

- 분야별로 온톨로지가 구축이 되어야 시맨틱웹 기반으로 기술되는 모든 자원의 의미기술에 있어서 완성도를 높일 것이고, 또한 이렇게 기술된 메타데이터는 지능형 검색을 지원하는 훌륭한 하부구조를 갖추게 될 것이다. 이러한 온톨로지의 구축은 정부의 지원 없이는 어렵다고 판단된다. 지식정보사업이 기계가독형 온톨로지의 구축을 지원할 필요가 있다고 본다.

참 고 문 헌

- 신문수, 류광열, 전무영. 2002. DAML+OIL을 이용한 모바일 에이전트 기반 협상에서의 정보 모델링. 『대한산업공학회』, 추계학술대회.
- 오삼균. 2002. 디지털도서관에서의 메타데이터의 역할. 『정보과학회지』, 20(8): 45-57.
- 오삼균. 2002. 시맨틱웹 기술과 활용방안. 『정보관리학회지』, 19(4): 297-320.
- 오삼균. 2003. 기본의미등록기의 RDF/OWL 연계 방안에 관한 연구. 『정보관리학회지』, 20(3): 241-260.
- 이현실, 이두영. 2003. 온톨로지 기반 한의학 처방 지식관리시스템 설계에 관한 연구.

- 『정보관리학회지』, 20(1): 341-371.
- 장태우, 신기태, 박진우. 2002. 온톨로지 기반의 e-비즈니스 통합 프레임워크. 『한국지능정보시스템학회지』 545-552.
- 정도현, 김태수. 2003. 시소러스를 기반으로 한 온톨로지 시스템 구현에 관한 연구. 『정보관리학회지』, 20(3): 155-176.
- Ahmed, K. et al. 2001. Professional XML Meta Data. Birmingham, UK: Wrox.
- AVEL-SKN(Substantiality Knowledge Network). 2003.
<<http://avel.edu.au/>>
- Berners-Lee, T. et al. 2000. "Semantic Web Development: Technical Proposal."
<<http://www.w3.org/2000/01/sw/DevelopmentProposal>>
- Bray, T., D. Hollander, and A. Layman. 1999. "Namespaces in XML."
<<http://www.w3.org/TR/REC-xml-names/>>
- Connolly, D., F. Hamelen, I. Horrocks, D. McGuinness, P. Patel-Schneider, L. Stein. 2001. "Annotated DAML+OIL Markup."
<<http://www.w3.org/TR/daml+oil-walkthru/>>
- Costello, R. 2002. "XML Schema Tutorial."
<<http://www.xfront.com/#schema>>
- Dean, M., D. Connolly, F. Harmelen, J. Hendler, I. Horrocks, D. McGuinness, P. Patel-Schneider, and L. Stein. 2002. "Web Ontology Language(OWL) Reference Version 1.0."
<<http://www.w3.org/TR/owl-ref/>>
- Denker, G., et al. 2001. Accessing Information and Services on the DAML-Enabled Web. Semantic Web Workshop 2001. Hongkong, China.
- The Dublin Core Metadata Initiative. 1999. "Dublin Core Qualifiers in RDF."
<<http://purl.org/dc/terms/>>
- Environmental Data Registry.
<<http://www.epa.gov/edr/>>
- Fallside, D. 2001. "XML Schema Part 0: Primer."
<<http://www.w3.org/TR/xmlschema-0/>>
- Garshol, LM. 2004. Metadata? Thesaurus? Taxonomies, Topic Maps?: Making sense of it all. XML 2004 Conference and Exposition. April 19-21, 2004 Amsterdam, Netherland.
<<http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>>
- Gomez-Perez, Asuncion. 2001. Evaluation of ontologies. International Journal of Intelligent Systems, 16(3): 391-409.
- Hamelen, F., I. Horrocks, and P. Patel-Schneider. 2001. "A Model-Theo-

- retic Semantics for DAML+OIL (March 2001).”
 <<http://www.w3.org/TR/daml+oil-model>>
- Lagoze, Carl. 2002. Core Services in the Architecture of the National Digital Library for Science Education (NSDL). Joint Conference in Digital Libraries.
 <<http://arxiv.org/ftp/cs/papers/0201/0201025.pdf>>
- Manola, F. and E. Miller. 2002. “RDF Primer.”
 <<http://www.w3.org/TR/rdf-primer/>>
- Miller, E. 1998. “An introduction to the Resource Description Framework.”
 <<http://www.dlib.org/dlib/may98/miller/05miller.html>>
- MusicBrainz.
 <<http://www.musicbrainz.org/>>
- Pepper, S. and L. Garshol. 2004. Toward seamless knowledge: integrating public sector portals in Norway. XML 2004 Conference and Exposition. April 19-21, 2004 Amsterdam, Netherland.
- Pepper, S. and L. Garshol. 2002. The XML Papers: Lessons Learned on Applying Topic Maps. XML 2002 Conference and Exposition. December 8-13, 2002. Baltimore, Maryland. USA.
- Powell, A., H. Wagner, S. Weibel, T. Baker, T. Matola, and E. Miller. 2001. “Namespace Policy for the Dublin Core Metadata Initiative (DCMI).”
 <<http://dublincore.org/documents/2001/10/26/dcmi-namespace/>>
- Smith, M., D. McGuinness, R. Volz, and C. Welty. 2002. “Web Ontology Language(OWL) Guide Version 1.0.”
 <<http://www.w3.org/TR/owl-guide/>>
- Stevens, Roberts, et al. 2003. Building ontologies in DAML+OIL. Comparative and Functional Genomics, 4(1): 134-141.

к с і