

지식정보 거버넌스를 위한 메타데이터 레지스트리 통합 프레임워크

최오훈*, 임정은**, 박성공***, 나홍석****, 백두권*****

요약

지식정보 거버넌스를 위하여 국가적으로 교육학술, 과학기술, 문화 분야 등에 대한 지식정보의 통합 검색이 요구되고 있다. 그러나 분야별 지식정보는 데이터 및 서비스 특성에 따라 기관별 고유 메타데이터 형식을 개발·사용해왔다. 따라서 분산된 지식정보를 메타데이터 기반으로 통합 시, 분야별 메타데이터 이질성 문제가 발생한다. 즉, 통합 검색을 위하여, 각 전문 분야별 메타데이터의 일관성 유지가 필요하다. 본 논문에서는 분야별 메타데이터를 조사하여 각 메타데이터들을 대표할 수 있는 표준 메타데이터를 정의한다. 이를 바탕으로 기존 시스템의 메타데이터의 수정 없이 기 구축된 전문 분야별 지식정보의 통합 검색을 가능하게 하는 지식정보 거버넌스를 위한 메타데이터 레지스트리 통합 프레임워크를 제안한다. 제안된 프레임워크를 통해 기존 메타데이터의 효율적인 등록 및 관리가 가능하며, 각 메타데이터간의 의미적 연관관계를 통한 의미 확장 검색이 가능하다.

Metadata Registry Integration Framework for Knowledge Information Governance

O-Hoon Choi*, Jung-Eun Lim**, Sung-Kong Park***, Hong-Seok Na****, Doo-Kwon Baik*****

Abstract

For Knowledge Information Governance, integration search mechanism of Knowledge Information is required in various industries. But when Knowledge Information system constructs in each professional field, it developed and used its own metadata format. Accordingly, if distributed information system searches using metadata, heterogeneous problems are occurred among metadata. For integration search, it needs a consistency management of metadata. Therefore, this paper defines a standard metadata that represent standard metadata as surveying a used metadata in each field systems. Also, we propose a metadata registry integration framework for Knowledge Information Governance. It enables an integration search service without a metadata modification of existing system. Through the proposed framework, it can registries and manages efficiently the existing metadata. Also, it enables a semantically extended search using semantic relation between metadata.

Keywords : knowledge information, Metadata, Metadata registry, Information Integration

1. 서론

※ 제일저자(First Author) : 최오훈
접수일자:2007년10월22일, 심사완료:2007년10월25일
* 고려대학교 컴퓨터학과 pens@korea.ac.kr
** 고려대학교 컴퓨터학과
*** 고려대학교 컴퓨터학과
**** 한국디지털대학교 디지털정보학과
***** 고려대학교 컴퓨터학과
▣ 이 연구에 참여한 연구자는 '2단계 BK21 사업'의 지원을 받았음

국가적으로 보존 및 이용가치가 높은 지식 자원을 사용자가 유비쿼터스 환경에서 언제, 어디서나 활용할 수 있도록 과학기술, 교육학술, 문화예술, 역사, 정보통신 분야의 2억5천만건의 지식정보자원을 데이터베이스로 구축하였다[1]. 구축된 데이터는 각 분야별 전문정보센터에 의해 관리되며 인터넷을 통해 해당 정보를 사용자가 원하는 형식으로 제공하고 있다. 현재 약 900개의 전문정보센터가 운영 중이며 각 분야별 종합정보센터는 9개가 운영 중이다. 다수의 데이터를 효율적으로 검색·관리하기 위한 지식정보 거

버전을 위해서 분산 관리되는 각 정보센터와의 상호운영이 요구된다. 본 논문은 현재 운영 중인 정보센터를 분석하여 지식정보 거버넌스를 위한 통합 검색·관리 방안을 제시하고 이를 위한 시스템을 구현한다.

현재 각 정보센터는 신규 및 기존 데이터베이스간의 정보 공유를 위하여 데이터의 데이터인 메타데이터를 정의하여 사용한다. 정보센터에서 정의된 메타데이터를 통하여 간단하게는 메타검색 뿐만 아니라, 정보 공유를 위한 XML 정의에 까지 이용하고 있다. 그러나 분야별 데이터 특징 및 예전부터 사용된 데이터베이스 때문에 분야에 알맞은 메타데이터를 정보센터 구축에 사용하였다. 즉, 지식정보의 공유를 위해 구축된 정보센터는 각 분야별 데이터 특성에 효율적인 메타데이터를 사용함으로써 정보센터 간 통합 검색이 어려우며, 지식정보 거버넌스 구축을 어렵게 한다. 이에 본 논문에서는 각 분야별로 구축된 정보센터의 상호운용성을 극대화하기 위한 방법으로 메타데이터[2]를 이용한 정보통합 방법론[3][4][5][6][7][8]을 기반으로 데이터 거버넌스를 위한 확장된 메타데이터 레지스트리 기반 프레임워크를 제안한다. 분야별 작성된 메타데이터는 의미적, 표현적, 구조적 불일치성이 발생할 수 있다. 본 논문에서는 각기 다른 도메인에서 구축된 메타데이터를 통합할 수 있도록 한국과학기술정보연구원, 한국교육학술정보원, 한국문화정보센터, 국가편찬위원회, 정보통신연구진흥원, 해양수산부, 건설기술연구원, 산업연구원, 법제처 정보센터에서 사용 중인 메타데이터를 분석하여 표준 메타데이터를 제안한다. 또한 본 논문에서는 기존 정보센터의 메타데이터 및 시스템의 수정 없이 지식정보 거버넌스를 위한 표준 메타데이터 및 분야별 메타데이터 간의 의미적 관계성을 관리 할 수 있는 프레임워크를 제안하고 구현한다.

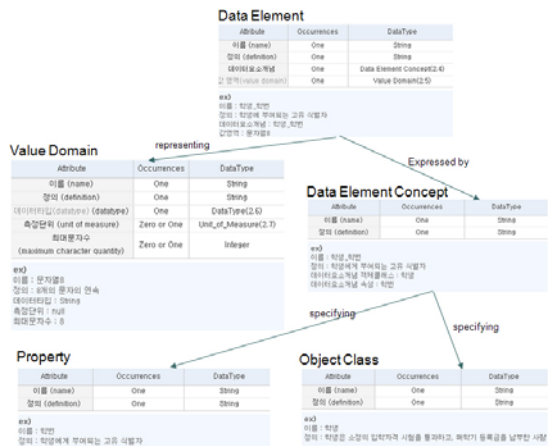
이 논문의 구성은 제 2장에서 관련 연구로서 메타데이터, 메타데이터 레지스트리와 현재 구축된 9개 분야의 메타데이터기반 시스템에 대한 문제점 및 표준 메타데이터에 대해 서술하고, 제 3장에서는 9개 분야의 정보센터를 통합 관리·검색 할 수 있는 지식정보 거버넌스를 위한 확장된 메타데이터 레지스트리 기반 프레임워크 아키텍처에 대하여 서술한다. 제 4장에서는 본 프

레이미워크의 구현을 서술하며 마지막 제 6장에서 결론 및 향후 연구에 대해 제시한다.

2. 메타데이터 관련 기술

2.1 메타데이터

메타데이터는 일반적으로 데이터에 관한 데이터로서 정보자원의 속성을 기술하는 데이터를 의미한다[2][9][12]. 즉, 메타데이터란 실제 콘텐츠는 아니면서 그에 대한 각종 정보를 가지고 있는 데이터를 말한다. 예를 들어 표제, 저자, 주제명, 분류 기호 등이 포함되어 있는 기존 도서관의 목록 레코드나 색인에 의해 생성된 데이터베이스 레코드는 이러한 의미에서 메타데이터라고 할 수 있다. 메타데이터는 여러 요소로 나누어지는데, 그 각각은 정보자원의 독특한 측면을 기술한다. 정보자원의 내용에 관한 사항, 다른 정보자원과의 관계, 정보자원의 지적속성에 관한 사항, 정보자원의 물리적 형식 및 생성 일자 등 정보자원을 식별할 수 있는 데이터 요소들로 구성된다. 이러한 요소를 메타데이터 레지스트리 데이터 모델로 표현한다.



(그림 1) 메타데이터 레지스트리의 데이터 모델

메타데이터 레지스트리의 데이터 모델은 데이터 요소 개념 (data element concept), 개념적 도메인 (Conceptual domain), 값 도메인 (value domain), 데이터 요소(data element)로 구성된다. 또한 데이터 요소는 객체분류 (Object class), 특성 (Property), 표현(Representation)의 세 가지

구성요소로 이루어진다. 데이터 요소 개념 (data element concept)은 개념적 도메인, 값 도메인과 데이터 요소의 3가지 구성요소와 결합하여 특정한 의미적 개념을 가진다. (그림 1)은 메타데이터 레지스트리의 데이터 모델을 예제를 통해 나타내고 있다.

2.2 메타데이터 레지스트리

메타데이터 레지스트리는 메타데이터를 등록하고 관리하기 위한 시스템을 의미한다. 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지 관리하며, 메타데이터의 명세와 의미의 공유를 통하여 호환성을 높이는 것을 그 목적으로 한다. 메타데이터 레지스트리는 아래 <표 1>의 ISO/IEC 11179를 기반으로 한다[2]. ISO/IEC 11179는 메타데이터 레지스트리에 관한 국제 표준으로서 메타데이터 레지스트리 구축에 관한 구체적인 시스템 아키텍처와 데이터베이스 설계 등은 명시되지 않았지만, 기술된 개념 및 논리적 데이터 모델에 따라 메타데이터 레지스트리를 구축할 수 있다.

<표 1> 메타데이터 레지스트리 관련 국내 표준

규격 번호	국제 규격번호	규격명
TTAS.IS-11179.1	ISO/IEC 11179-1	정보기술 - 데이터 요소에 대한 규격과 표준화 - 데이터 요소에 대한 표준화 프레임워크
TTAS.IS-11179.2	ISO/IEC 11179-2	정보기술 - 데이터 요소에 대한 규격과 표준화 - 도메인 식별을 위한 개념 분류
KS X 1517-3	ISO/IEC 11179-3	데이터 요소 - 제3부 : 데이터 요소의 기본속성
KS X 1517-4	ISO/IEC 11179-4	데이터 요소 - 제4부 : 데이터 요소의 정의를 위한 규격과 지침
KS X 1517-5	ISO/IEC 11179-5	데이터 요소 - 제5부 : 데이터 요소를 위한 명명 및 식별 원칙
KS X 1517-6	ISO/IEC 11179-6	데이터 요소 - 제6부 : 데이터 요소의 등록

메타데이터 레지스트리는 다양한 기관에 산재된 메타데이터를 데이터 모델로 정의하고, 서비스하는 메타데이터 중앙관리소의 역할을 하는 시스템으로 메타데이터를 등록받고, 인증하고, 관리하고, 서비스하는 체계를 제공하는 표준이다. 인터넷을 통한 정보 관리와 교환을 위하여,

특히 정보의 상호운용성 확보, 의미 파악 등을 위하여 중요한 정보를 제공하는 역할을 한다. 현재 미국 환경청 (EPA), 법제처 등에서 메타데이터를 구현하였으며, 국내에서는 한국과학기술정보연구원 (KISITI)과 한국데이터베이스진흥센터 (DPC)에서 메타데이터 레지스트리를 구현하였다 [9][11][12].

2.3 분야별 메타데이터 및 표준 메타데이터

<표 2> 지식정보자원관리 필수 메타데이터 요소

필수 요소명	정의	사용 가능한 인코딩 스킴	요소에 대한 설명
Title	표제		- 정보자원에 부여된 제목 - 정보자원을 공식적으로 지칭하는 이름
Creator	제작자		- 정보자원의 내용에 주된 책임을 갖고 있는 개체(사람 또는 단체)의 이름
Subject	주제	- KDC - DDC - LCC - UDC - LCSH - MeSH - 자체 주제 분류	- 정보자원의 내용을 간결하게 기술한 정보자원의 주제 - 정보자원의 주제를 표현하는 키워드, 구절, 또는 분류코드로 나타내어지는 데, 통제 어휘나 공식적 분류 스킴에 의한 값을 사용하는 것이 바람직함
Publisher	발행처		- 정보자원을 현재의 형태로 이용 가능하게 만든 데 대한 책임을 갖고 있는 개체 (개인 또는 단체)의 이름
Date	날짜	- ISO 8601 - W3C-DTF - DCMI Period	- 정보자원이 존재하는 동안 어떤 변화가 발생한 경우의 해당 날짜 - 대체로 정보자원의 생성일이나 정보자원이 현재의 형태로 이용 가능하게된 날짜와 관련됨
Type	자료 유형	- DCT1 - 자체 유형 분류	- 정보자원의 범주 또는 유형 - 내용의 일반적인 범주, 기능, 유형, 또는 집성 수준을 표현하는 말로서, 통제 어휘의 사용이 권장됨
Identifier	식별자	- URI - DOI - ISBN - ISSN	- 정보자원을 식별하기 위한 고유한 문자열 또는 지시자

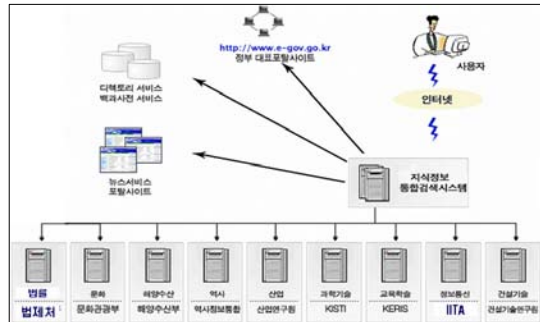
9개 분야의 지식정보센터에서 사용하고 있는 메타데이터는 내용상 크게 16개 메타데이터로 구분되며 이들은 각각 최대 10개 이상의 상세 정보를 위한 메타데이터를 사용하고 있다. IITA 정보센터에서는 Relation이란 메타데이터 하위에 isPartOf, hasPart, isVersionOf, hasVersion, isFormatOf, hasFormat, isReferenceBy, references, isRequiredBy, requires, isReplacedBy, replaces, conformsTo와 같은 13가지 메타데이터를 상세

정보를 위하여 정의하여 사용한다. 한편 국사편찬위원회 정보센터는 하위에 isPartOf, requires 라는 2가지 상세 메타데이터만을 사용하고 있다. 각 분야별 메타데이터의 수 및 그 정의, 표현 방식이 다르다. 따라서 본 논문에서는 9개 분야의 메타데이터를 모두 조사하고 그 의미 및 표현 구조를 분석하여 의미적, 구조적 충돌 없이 사용할 수 있는 표준 메타데이터를 선정하였다. 선정된 표준 메타데이터는 간단하게 기술할 수 있는 15개 요소로 구성된 더블린코어를 기반으로 제정하였다. 지식정보 거버넌스를 위한 표준 메타데이터 요소 집합으로 Title, Creator, Subject, Publisher, Date, Type, Identifier의 7개 요소를 필수 요소로 선정한다[13]. 지식정보 거버넌스를 위한 표준 메타데이터 요소에 대한 설명은 다음의 <표 2>와 같다. 7개의 필수 요소를 제외한 나머지 요소들은 향후 지식정보 거버넌스를 위한 통합 메타데이터 검색에서는 이용되지 않지만, 각 기관 또는 분야별 메타데이터를 구축할 때 정보자원에 대한 보다 풍부한 의미를 기술할 수 있도록 하기 위해 사용될 수 있다. 즉, 이들 중 일부는 분야별 확장 메타데이터 요소 선정 시 필요에 따라 특정 분야의 필수 요소로 포함될 수도 있다. 각 분야별 혹은 기관별 요구사항에 따라 표준 메타데이터 요소 집합에 정의되지 않은 별도의 메타데이터 요소나 한정어가 필요한 경우에는 분야별 혹은 기관별로 새로운 한정어를 추가하거나 자체 메타데이터 요소를 정의하여 사용한다.

3. 메타데이터 레지스트리 통합 프레임워크 설계

최근 각 분야별로 분산된 지식정보센터의 유통 및 검색에 메타데이터가 적용되고 있다. 그러나 지식정보의 가공 시 분야별로 상이한 방식과 형태로 메타데이터가 구축되어 각 분야별 메타데이터가 제정되어 이들 간 지식정보의 유통 및 메타데이터의 통합이 어렵다. 기존 메타데이터 레지스트리는 특정 도메인의 메타데이터를 등록, 인증, 관리하는 것을 범위로 삼는다. 즉 ebXML 레지스트리, UDDI 레지스트리와 같이 기 구축된 다른 레지스트리와 정보통합을 위해서는 각

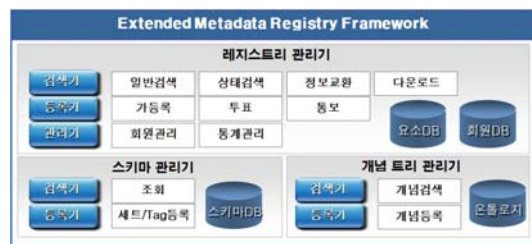
레지스트리의 데이터 포맷을 1:1 매핑[14]으로 변환하거나, 의미적 유사성을 통한 변환[3][5], 변환규칙을 통해 변환[3]해야 한다. 본 논문에서는 각기 기 구축된 메타데이터 레지스트리에서 관리되는 각 분야별 메타데이터를 통합 관리할 수 있는 확장된 메타데이터 레지스트리를 통하여 다른 레지스트리와 정보 공유를 위한 모델을 제시한다. (그림 2)는 제안된 지식정보 거버넌스를 위한 아키텍처를 나타내고 있다.



(그림 2) 지식정보 거버넌스를 위한 통합검색 아키텍처

확장된 메타데이터 레지스트리 기반 지식정보 관리 시스템은 각 분야별로 구축된 정보센터의 메타데이터를 등록, 인증, 관리하며, 정보센터에서 사용하는 XML 문서와 메타데이터의 의미적 관계를 적용할 수 있는 시스템을 제공한다.

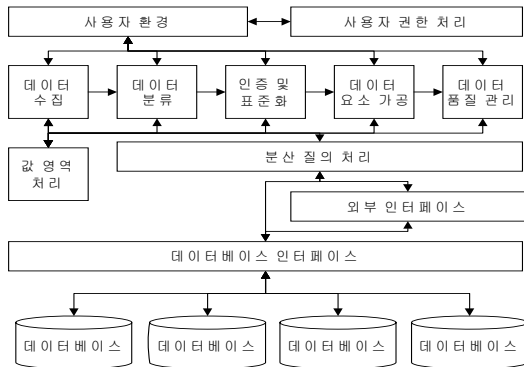
본 논문에서는 이 시스템을 레지스트리 관리기, 스키마 관리기, 개념트리 관리기의 세 부분으로 나누어 제안하며, 구현하였다. (그림 3)은 각 관리기의 기능을 블록 다이어그램으로 표현하였다.



(그림 3) 확장된 메타데이터 레지스트리 프레임워크

3.1 레지스트리 관리기 설계

레지스트리 관리기는 ISO/IEC 11179의 표준을 준수하며 기반 기술은 기 연구된 컴포넌트를 사용한다[9]. 메타데이터를 등록하고자 하는 사용자는 등록하고자 하는 메타데이터가 기 등록되어 있는지 검색 가능하며, 필요시 해당 메타데이터를 등록 할 수 있다. 인증 과정은 표준 인증 프로세스를 따른다[2][9].



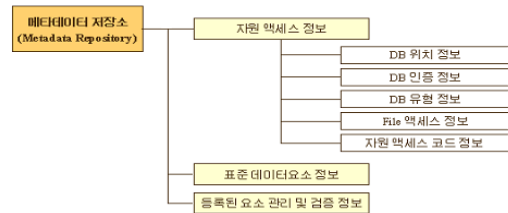
(그림 4) 메타데이터 레지스트리의 구성 요소 및 관계

레지스트리 관리기 각 사용자는 사용자 환경을 통해 메타데이터 레지스트리의 서비스를 이용하게 된다. 이 때 각 사용자는 사용자 권한 처리부에 의해 자신의 권한을 부여받고, 사용자 권한 처리부는 사용자가 이용할 수 있는 서비스에 제한을 가하여 메타데이터 레지스트리의 안정성을 부여한다. 메타데이터는 데이터 수집, 데이터 분류, 인증 및 표준화, 데이터 요소 가공과 데이터 품질 관리 요소들을 통해 표준으로서의 가치를 부여받게 된다. 이때 모든 과정에서 값 영역 처리부와와의 정보 교환을 통해 더 정제된 품질을 유지하게 된다. 데이터 품질 관리 처리부는 지속적으로 각 메타데이터의 품질의 관리한다. 사용자 및 모든 요소들의 데이터베이스 접근은 분산질의 처리를 통해서 처리된다. 이는 메타데이터 레지스트리의 데이터베이스가 논리적인 데이터베이스이고 물리적으로는 원격의 데이터 레지스트리이기 때문에 분산질의 처리부를 통해서 물리적 데이터베이스에 접근한다. 분산질의 처리부는 원격의 데이터베이스를 접근할 때는 외부 인터페이스와 데이터베이스 인터페이스를 통해 물

리적 데이터베이스에 접근한다.

각 구성 요소에 해당하는 서비스를 제공하고 있다는 점을 감안한다면, 각 구성 요소들은 메타데이터 레지스트리를 구성하고 있는 기본적인 모델을 만드는 가장 기본 요소로 작용한다. 즉, 응용 분야별 또는 표준의 수정으로 인해 메타데이터 관련 레지스트리에 추가 또는 삭제될 기능들은 각 구성 요소의 상호 연관성을 변형시키기보다 각 구성 요소들이 제공해야할 서비스들의 변형을 가져오는 수준에서 해결 될 수 있다.

메타데이터 저장소는 통합 검색, 데이터 요소의 등록 및 관리 등과 관련된 추가적인 메타 정보들을 관리한다. 통합 검색과 관련해서, 각각 분산 구축되어 관리되는 데이터베이스들에 대한 위치정보, 액세스 정보, 테이블의 스키마 등에 대한 정보를 지니게 된다. 또한 데이터 요소의 등록과 관리를 위해 필요한 위원 정보, 등록된 데이터 요소, 채택된 요소 정보 등을 지니게 된다.

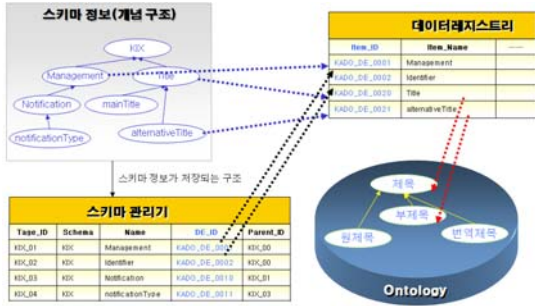


(그림 5) 메타데이터 저장소의 구성

메타데이터 저장소는 크게 사용자 정보, 사용자 인증 권한 정보, 자원 액세스 정보, 표준 데이터 요소 정보, 스키마 정보 등으로 구성되어 있으며, 이들 정보는 각각 통합 검색, 데이터 요소 등의 관리를 위해 필요한 정보들이다. 또한 이들 정보들은 통합 검색 및 메타데이터 요소 관리를 위해 상호 연관성을 지니게 된다.

3.2 스키마 관리기 설계

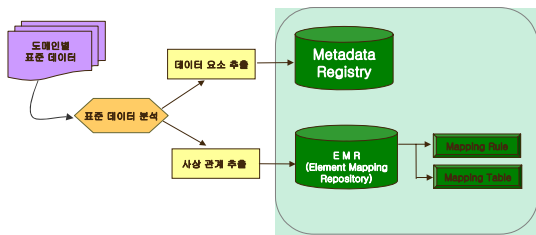
스키마 관리기는 각 정보센터에서 사용되는 XML 문서의 XML tag를 표준 메타데이터와 매핑 시킬 수 있다. 메타데이터 레지스트리에서 관리하는 메타데이터는 다양한 분야에서 사용되는 공통적인 메타데이터만을 관리하므로 분야별로 특수한 메타데이터나 XML tag등이 등록 되지 않을 수 있다. 따라서 스키마 관리기를 통해 표



(그림 6) 메타데이터와 스키마 정보간의 매핑

준으로 등록되지 못한 분야별 메타데이터 및 XML tag는 매핑 표준 메타데이터와 매핑 관계가 있으므로 그 의미를 제공받을 수 있다. (그림 6)은 스키마 관리기에 저장된 메타데이터와 XML tag의 스키마 정보간의 매핑을 보여준다.

(그림 7)은 요소 매핑 레포지토리인 EMR을 나타내고 있다. EMR은 표준 데이터 요소들과 구축 대상이 되는 각 분야의 데이터 요소들 간의 사상관계를 저장한다. EMR은 Mapping Rule, Mapping Table의 2가지 요소로 구성되어 있다.



(그림 7) 요소 매핑 레포지토리 (Element Mapping Repository)

<표 3>은 정보센터 메타데이터 요소와 표준 데이터 요소들 사이의 사상에 발생할 수 있는 이질성에 대한 분류이다. 각 항목은 하나의 정보센터 메타데이터 요소가 적용될 수 있는 사상 규칙에 대해서 나타낸다. Mapping Rule를 적용한 정보센터 메타데이터 요소와 표준 요소들과의 사상 관계를 테이블 형태로 저장한다. 두 표준 요소들 간의 데이터 요소 교환을 위하여 Mapping Table에 입력된 동일한 의미를 갖는 정보센터 메타데이터 요소 정보를 검색한다.

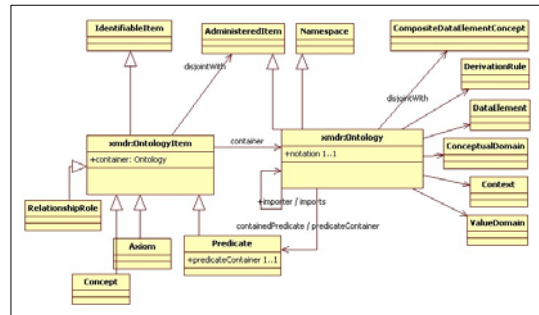
<표 3> 데이터 이질성 분류 및 변환

Data Heterogeneity		Definition
Semantic	Substitution	하나의 표준 메타데이터 요소가 다른 메타데이터 요소로 대체
	Decomposition	하나의 표준 메타데이터 요소가 다른 메타데이터 요소들의 집합으로 구성
Schematic	Composition	하나의 표준 메타데이터 요소가 다른 메타데이터 요소를 구성하는 요소들 중 하나로 위치
	Rearrangement	하나의 표준 메타데이터 요소와 구조를 이루는 순서가 다름

3.3 개념트리 관리기 설계

개념트리 관리기는 메타데이터에서 제공하지 않는 개념을 저장할 수 있다. 온톨로지는 노드(요소, 데이터)와 아크(관계)로 구성된다. 따라서 온톨로지 파일 (OWL, RDF 등)을 저장하며 그 안에서 제공되는 관계성을 파악하여 연관관계에 추가할 수 있다. 노드는 메타데이터로 정의하여 메타데이터 레지스트리 관리기를 통해 관리할 수 있다. 본 개념트리 관리기는 OWL Lite에서 정의된 관계성을 기본으로 제공하여, 메타데이터 간, XML tag 간, 메타데이터와 XML tag간의 관계성을 정의한다. 이를 통해 지식정보통합 관리기에 저장된 메타데이터를 기준으로 관계성을 유추하여 추후 개념트리 기반 메타데이터 레지스트리를 구축하는 초석을 제공한다.

(그림 8)은 확장된 메타데이터 레지스트리에 개념트리를 관리할 수 있도록 한 표현 모델이다.



(그림 8) 확장된 메타데이터 레지스트리의 개념트리 표현 모델

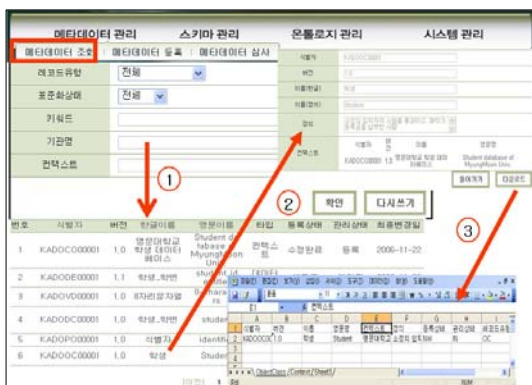
4. 메타데이터 레지스트리 통합 프레임워크 구현

앞에서 지식정보 거버넌스를 위한 확장된 메타데이터 레지스트리 기반 프레임워크 개념과 필수 컴포넌트에 대해 기술했다. 이 장에서는 앞서 정의한 지식정보 거버넌스 프레임워크의 구현 및 동작 과정을 시나리오에 기반하여 기술한다. 9개 분야의 지식정보 센터의 구현은 이미 구축되었으므로 본 과정에서는 배제되었다. 지식정보 센터에서 사용하는 메타데이터 및 XML 스키마는 지식정보 거버넌스를 위한 확장된 메타데이터 레지스트리에 등록 및 검색 데이터로 사용된다.

4.1 표준 메타데이터 조회

확장된 메타데이터 레지스트리 기반 프레임워크를 효율적으로 이용하기 위해서 각 정보센터에서 사용하는 메타데이터 및 XML 스키마가 표준으로 사용되고 있는지 조회과정이 필요하다. 만약 비슷한 의미의 연관된 표준이 사용되고 있다면 사용자에게 그 결과를 반환해야 한다. 조회시 결과값이 없다면, 해당 정보센터에서 사용하는 메타데이터 및 XML 스키마를 확장된 메타데이터 레지스트리에 등록하여 표준 심사를 거쳐 표준 메타데이터로 승인 받아 사용할 수 있다.

(그림 9)는 표준 메타데이터 조회과정을 나타내고 있다. ①은 키워드, 기관명, 컨텍스트(메타데이터가 의미를 갖는 범위)에 따라서 표준 메타데이터가 등록되어 있는지를 조회한다. ② 해당 키워드가 포함된 표준 메타데이터가 제공되며, 사용자의 선택에 따라 해당 상세 정보를 나타낸다. ③ 사용자의 요구시 별도의 파일로 저장하여 이용할 수 있다.



(그림 9) 표준 메타데이터 조회과정

4.2 표준 메타데이터 등록

조화과정에서 합당한 결과를 얻지 못한 경우, 정보센터는 메타데이터를 등록하기 위한 표준 메타데이터 등록과정을 진행한다. 표준 메타데이터 등록은 컨텍스트, 객체클래스, 속성, 데이터 요소 개념, 데이터 타입, 측정단위, 값영역, 데이터 요소의 단계에 따라 해당 값을 등록한다. (그림 10)과 (그림 11)은 표준 메타데이터 등록 과정을 나타내고 있다. ① 컨텍스트는 메타데이터가 의미를 갖는 범위를 의미한다. 즉 컨텍스트는 비즈니스 도메인, 데이터베이스 파일, 데이터 모델이나 표준문서일 수 있다. ② 객체클래스는 데이터를 수집하고 저장하고자 하는 대상이 되는 사물 혹은 개념을 의미한다. 객체지향모델에서는 클래스에 포함되어 있는 개념과 연관되고, 개체-관계 모델에서는 개체와 연관된다. 차, 사람, 가정, 학교 등이 그 예이다. ③속성은 객체클래스의 모든 구성원들이 지니는 공통적인 특징이다. 속성은 객체지향모델이나 개체-관계 모델에서의 속성에 포함된 개념과 연관이 있다. 예를 들어 객체클래스 “사람”의 속성은 성별, 나이, 수입, 주소 등을 들 수 있다. ④ 데이터 요소의 사상을 이루는 특정 값 영역이 지정되지 않는 개념을 의미한다. 즉, 데이터 요소 개념은 객체클래스와 속성으로 이루어진다. 학생_학번, 학생_성명, 학생_학과 등이 그 예이다. 또한 데이터 요소 개념 등록 과정에서는 연관 데이터 요소 개념을 등록할 수 있다. 즉 등록하는 메타데이터가 표준 메타데이터와 비슷한 의미로 사용될길 원할 경우 “is-a”, “part-of”의 관계로 연관관계를 설정할 수 있다.



(그림 10) 표준 메타데이터 등록과정 I

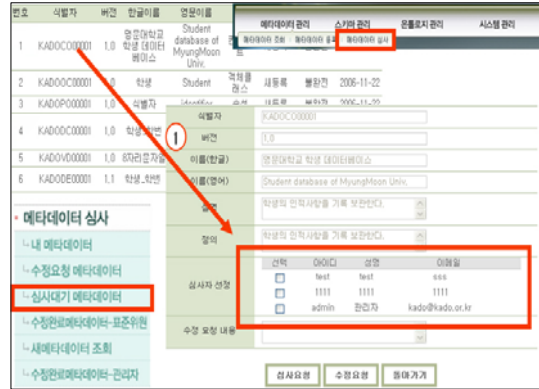


(그림 11) 표준 메타데이터 등록과정 II

⑤ 데이터 타입은 특정 데이터 요소의 허용값을 표현하는 방식이다. 문자열 7자리, 정수 6자리 등이 그 예이다. ⑥ 측정단위는 값이 측정되는 실제 단위로서 Cm, mile, km 등이 사용된다. ⑦ 값 영역은 어떤 데이터 요소의 허용 가능한 값의 집합으로, 0~10 혹은 ISO 3166등과 같은 표준이 사용될 수 있다. ⑧ 데이터 요소 등록은 ①부터 ⑦과정이 모두 완료된 후 메타데이터를 등록하는 과정이다. 데이터 요소는 어떤 조직이 생성, 관리, 삭제하는 데이터의 기초 단위로 데이터 요소 개념과 표현 부분으로 구성된다. 데이터 요소 개념은 객체 클래스와 속성으로 구성되며, 표현은 값영역, 데이터 타입, 측정단위로 구성된다. 학생_학번_문자코드, 학생_성명_한글, 학생_학과_영문 등이 그 예이다.

4.3 표준 메타데이터 등록

표준 메타데이터 심사 과정은 각 정보센터에서 등록한 메타데이터를 심사 운영위원회를 통해 표준 메타데이터에 연관시켜 줄 것인가의 여부를 결정짓는다. 또한 다수의 정보센터에서 표준 메타데이터로 사용하기 원하는 메타데이터는 수정 및 심사 과정을 통하여 확장된 메타데이터 레지스트리의 표준 메타데이터로 선정할 수 있다. (그림 12)는 표준 메타데이터 심사과정을 간략하게 나타내고 있다.



(그림 12) 표준 메타데이터 심사과정

4.4 XML 스키마 등록 및 조회

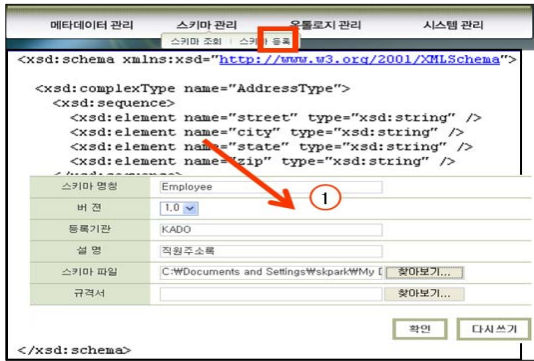
XML 스키마 등록 및 조회는 정보센터에서 사용 중인 메타데이터 등록할 경우 4.2처럼 그 과정이 별도의 교육이 필요할 정도로 복잡하다. 따라서 XML 스키마 형식으로 메타데이터를 제공받아, 확장된 메타데이터 레지스트리에 등록 및 조회 할 수 있는 환경을 제공한다. (그림 13)은 확장된 메타데이터 레지스트리에서 XML 스키마 및 관련 메타데이터를 조회할 수 있는 화면이다.

① 사용자는 스키마 조회를 통하여 키워드를 기반으로 메타데이터를 조회하면 해당 메타데이터가 속한 스키마 정보가 표현된다. ② 트리 구조로 표현된 스키마 정보는 해당 태그 정보가 선택된 경우 상세한 정보를 나타낸다.



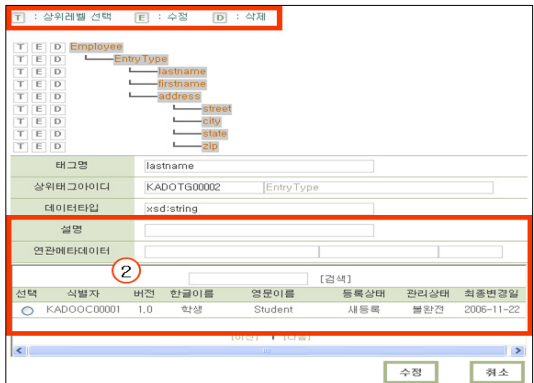
(그림 13) XML 스키마 및 관련 메타데이터 조회 과정

스키마 등록은 정보센터에서 사용되는 메타데이터에 관한 XML 스키마를 입력받아, 이를 파싱하여 사용자에게 효율적으로 메타데이터를 등록할 수 있도록 한다. (그림 14)과 (그림 15)는 XML 스키마 등록 과정을 나타내고 있다. ①은 XML 파일을 확장된 메타데이터 레지스트리에 업로드 한다.



(그림 14) XML 스키마 등록 과정

② 파싱을 통해 XML 태그값을 추출하여 트리 구조로 사용자에게 제공한다. 제공된 메타데이터를 하나씩 클릭하며 표준 메타데이터와 연관 시키거나, 원하는 내용을 수정할 수 있다.

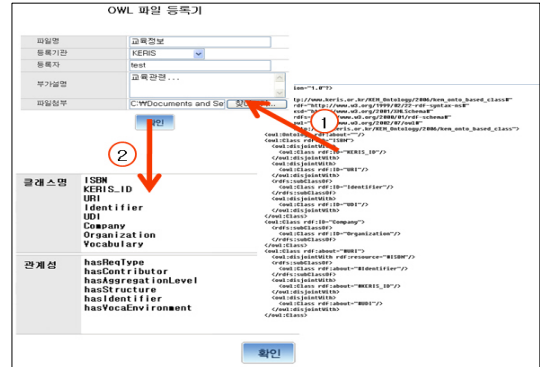


(그림 15) XML 스키마 수정과정

4.5 OWL 파일 등록

OWL 파일 등록은 4.4의 스키마 등록기와 비슷한 기능을 제공한다. OWL 파일을 등록받았 그 파일을 파싱하여 클래스는 메타데이터로, 프로퍼티는 관계성으로 추출하여 시스템에 등록한다. ①은 정보센터에서 사용하는 메타데이터

대한 XML 스키마를 업로드 하는 과정이다. ② 파싱과정을 통해 클래스와 관계성이 추출되어 표준메타데이터와 연관할 수 있도록 EMR에 임시 저장된다.



(그림 15) OWL 파일 등록 및 추출 과정

5. 결론 및 추후 연구

지식정보사회가 고도화되면서 구축 및 제공되는 지식정보가 증가하고 있으며, 이러한 지식정보를 효과적으로 통제, 관리하기 위하여 특정 분야별로 메타데이터를 분야별 요구사항 및 제약 조건에 따라 제정하여 정보센터를 구축하였다. 그러나 이러한 메타데이터는 해당 분야 또는 시스템에 특성화되고 있기 때문에, 타 분야에 사용하기 어렵다. 또한 같은 분야라고 해도 상세 사용에 따라 지역화된 메타데이터가 계속 생겨나고 있기 때문에 실제적으로 다른 분야나 다른 영역의 메타데이터를 정확히 이해하고 데이터를 적절하게 이용하기는 쉽지 않다. 즉, 메타데이터의 효율적인 관리를 통하여 고도화된 지식정보를 제공하며, 국가적으로 가치있는 과학기술, 교육학술, 문화, 예술, 역사, 정보통신 분야에 대한 지식정보 거버넌스가 요구되고 있다.

따라서 본 논문에서는 지식정보 거버넌스를 위한 메타데이터 레지스트리 통합 프레임워크를 제안하였다. 제안된 프레임워크는 표준 메타데이터 및 특정 분야 메타데이터의 검색, 등록, 심사를 위한 레지스트리 관리기와 각 분야별로 사용되는 XML 스키마를 표준 메타데이터와 매핑시킬 수 있는 스키마 관리기, 표준 메타데이터와 XML tag간의 의미적 연관관계를 매핑 시

킬 수 있는 개념트리 관리기를 제안 및 구현하였다. 이를 통하여 기 구축된 9가지 정보센터들에 대한 재구축작업 없이 통합 검색 및 관리가 가능하다. 또한 정보센터들에서 사용하는 메타데이터를 XML 스키마로 업로드 하여 기존 메타데이터 등록 방식보다 쉽게 통합 관리할 수 있다.

참 고 문 헌

[1] “지식정보자원관리 메타데이터 구축 지침”, 한국전산원
 [2] ISO/IEC 11179 Metadata Registry(MDR) part 1~6
 [3] Jung-Eun Lim, O-Hoon Choi, Hong-Seok Na, Doo-Kwon Baik: A Methodology for Semantic Similarity Measurement among Metadata based Information System. SERA 2006: 202-205
 [4] O-Hoon Choi, Jung-Eun Lim, Doo-Kwon Baik: MD R-Based Framework for Sharing Metadata in Ubiquitous Computing Environment. EUC 2005: 858-866
 [5] JuHum Kwon, Chang-Joo Moon, Soo-Hyun Park, Doo-Kwon Baik: Measuring Semantic Similarity Based on Weighting Attributes of Edge Counting. AIS 2004: 470-480
 [6] Dongwon Jeong, Young-Gab Kim, Soo-Hyun Park, Doo-Kwon Baik: Uniformly Handling Metadata Registries. SERA 2004: 81-91
 [7] Dongwon Jeong, Doo-Kwon Baik: Incremental data integration based on hierarchical metadata registry with data visibility. Inf. Sci. 162(3-4): 147-181 (2004)
 [8] JuHum Kwon, Dongwon Jeong, Lee-Sub Lee, Doo-Kwon Baik: Intelligent Semantic Concept Mapping For Semantic Query Rewriting/Optimization In Ontology-Based Information Integration System. International Journal of Software Engineering and Knowledge Engineering 14(5): 519-542 (2004)
 [9] 백두권, “MDR 상호운용을 위한 아키텍처 모델 연구”, 한국데이터베이스 진흥센터
 [10] “의미 검색을 위한 온톨로지의 유사성 검사 및 맵핑”, 과학기술정보원
 [11] US Environmental Protection Agency, <http://www.epa.gov>
 [12] 한국데이터베이스 진흥센터 메타데이터 레지스트리 <http://mdr.dpc.or.kr>
 [13] Dublin Core Metadata Initiative (DCMI), <http://www.dublincore.org>
 [14] Exchange Network. <http://www.exchangenetwork.org>

최 오 훈



2000년 : 고려대학교 컴퓨터학과
 2002년 : 고려대학교 대학원
 (이학석사-전산학)
 2004년 : 고려대학교 대학원
 (박사과정수료-전산학전공)

2002년~2006년 : (주)라임미디어 테크놀로지스
 2006년~2007년 : (주)파슨텍

2005년~현 재 : 건국대학교 컴퓨터응용과학부
 겸임교수

관심분야 : 메타데이터, 데이터 품질, 정보통합, 상황인지, 온톨로지, Home Network

임 정 은



2002년 : 숙명여자대학교 대학원
 (이학석사-전산학)
 2004년 : 고려대학교 대학원
 (박사과정수료-전산학전공)

2002년~현 재 : 고려대학교 컴퓨터정보통신연구소

관심분야 : 시멘틱웹, 메타데이터, 온톨로지, 유비쿼터스 헬스케어(U-Healthcare)

박 성 공



1998년 : 고려대학교 대학원
 (이학석사-전산학)
 2000년 : 고려대학교 대학원
 (박사과정수료-전산학전공)

2007년 : (주)메타라이츠

관심분야 : 메타데이터, 정보통합, 정보보호

나 홍 석



1994년 : 고려대학교 대학원
(이학석사-전산학)
2004년 : 고려대학교 대학원
(이학박사- 전산학전공)

1996년~1999년 : 고려대학교 기초과학연구소 연구원
2000년~2002년 : (주)라임미디어테크놀로지스
대표 및 연구소장
2002년~현 재 : 한국디지털대학교 디지털정보학과
교수

관심분야 : 소프트웨어공학, 데이터아키텍처,
메타데이터, 온톨로지, 디지털콘텐츠,
이러닝(e-learning), 데이터 품질

백 두 권



1977년 : 고려대학교 대학원
(공학석사-산업공학)
1983년 : Wayne State Univ.
(전산학 석사)
1985년 : Wayne State Univ.
(전산학 박사)

1986년~현 재 : 고려대학교 컴퓨터학과 교수
1989년~현 재 : 한국정보과학회(이사/부회장)
1991년~현 재 : 한국시뮬레이션학회
(이사/부회장/감사/회장/고문)
1991년~현 재 : ISO/IEC JTC1/SC32 전문위원회장
2001년~현 재 : (사)도산아카데미(원장)
2004년~2005년 : (사)한국정보처리학회(부회장)

관심분야 : 데이터베이스, 소프트웨어 공학, 시뮬레이
션, 메타데이터, 정보통합, 정보보호