

과학기술정보를 위한 메타데이터 레지스트리 관리 시스템

(Metadata Registry Management System for Science and Technology Information)

정 동 원 [†] 신 동 길 ^{**} 정 은 주 ^{**}
(Dongwon Jeong) (Dongkil Shin) (Eunju Jeong)

김 영 갑 ^{***} 이 정 옥 ^{****} 백 두 권 ^{****}
(Young-Gab Kim) (Jeong Oog Lee) (Doo-Kwon Baik)

요 약 지금까지 ISO/IEC 11179를 기반으로 한 많은 메타데이터 레지스트리 관리 시스템들이 개발되었다. 그러나 기존 시스템들의 경우, 국제 표준인 ISO/IEC 11179 명세에 위배되거나 활용을 위한 필수 기능이 미비하다는 단점을 지닌다. 또한 핵심적인 구현 내용이 공개되지 않음으로써 국내 환경에 적합한 메타데이터 레지스트리 관리 시스템 개발에 어려움이 있다.

이 논문의 목적은 기존 시스템들의 문제점을 해결하고 원천 기술 확보를 위한 메타데이터 레지스트리 관리 시스템 개발에 있다. 또한 개발한 시스템을 과학기술정보 분야에 적용하여 실제 메타데이터 레지스트리를 구축 과정을 보인다. 개발된 메타데이터 레지스트리 관리 시스템은 크게 메타데이터 레지스트리 계층, 핵심 컴포넌트 계층, 확장 계층 및 서비스 인터페이스 계층으로 구성된다.

구현된 메타데이터 관리 시스템은 ISO/IEC 11179의 규칙을 준수하고 활용을 위한 필수적인 기능들을 포함하고 있다. 따라서 시스템 개발 프로세스 및 메타데이터 레지스트리 구축을 위한 지침서로서 이용될 수 있다. 또한 컴포넌트를 기반으로 설계 및 구현되었기 때문에 다양한 분야의 메타데이터 레지스트리 관리 시스템 개발을 위한 재사용이 용이하며 시스템 개발 시간과 비용을 감소시킨다.

키워드 : 메타데이터, 데이터 요소, 메타데이터 레지스트리, MDR 관리 시스템, 과학기술정보, ISO/IEC 11179

Abstract Many metadata registry management systems have been developed, which are based on ISO/IEC 11179. However, they do not follow the international standard, ISO/IEC 11179 nor provide some of mandatory functions. It is difficult to a develop metadata registry management system for domestic environment because of their hidden implementations.

The goals of this paper are to solve the problems of the existing systems and to develop a metadata registry management systems for accumulation of primitive technologies. This paper also shows the metadata registry building process for the science and technology information field using the developed system. This system consists of Metadata Registry Layer, Core Component Layer, Extension Layer, and Service Interface Layer.

The developed metadata registry management system follows ISO/IEC 11179 and contains mandatory functions for practical use. Therefore, it can be used as a guideline for building metadata registries and a development process of metadata registry management systems. We can easily reuse

[†] 정 회 원 : 고려대학교 컴퓨터학과 교수
withimp@korea.ac.kr

^{**} 비 회 원 : 고려대학교 컴퓨터학과
dkshin@software.korea.ac.kr
violetto@software.korea.ac.kr

^{***} 학생회원 : 고려대학교 컴퓨터학과
yjkim@software.korea.ac.kr

^{****} 비 회 원 : 건국대학교 컴퓨터공학부 교수
ljo@kku.ac.kr

^{****} 종신회원 : 고려대학교 컴퓨터학과 교수
baik@software.korea.ac.kr

논문접수 : 2004년 3월 16일
심사완료 : 2004년 9월 15일

its components for development of metadata registry management systems in various fields because the system is designed and implemented based on the component-based development methodology. It also decreases time and cost for developing systems.

Key words : Metadata, Data element, Metadata registry, MDR management system, Science and Technology Information, ISO/IEC 11179

1. 서론

메타데이터 레지스트리(MDR, Metadata Registry)는 데이터베이스간 정보 공유 및 교환의 용이성과 상호운용성 향상을 위하여 국제 표준 기구인 ISO/IEC JTC 1에 의해 개발된 ISO/IEC 11179의 핵심 구성요소이다. 메타데이터 레지스트리는 데이터 요소의 집합이라고 정의할 수 있으며, 데이터 요소(Data Element)는 하나의 데이터를 이해하고 활용하기 위해 필요한 메타데이터들로 구성되어 있다[1].

ISO/IEC 11179는 중립적인, 즉 특정 도메인에 종속되지 않는 메타데이터 관리를 위한 메타 모델이다. 이미 외국의 경우, 이를 기반으로 한 여러 메타데이터 레지스트리가 구축되었으며 또한 이를 관리하기 위한 관리 시스템들이 개발되어 왔다. 미국 환경청(Environment Protection Agency)의 환경 정보를 위한 EDR (Environmental Data Registry)[2-4], 호주 건강 복지 기관인 AIHW(Australian Institute of Health & Welfare)의 NHIK(Australian National Health Information Knowledgebase)[5,6], 미국 국방성, 건강국 및 보건증진재단 등이 공동으로 개발한 USHIK(U.S. Health Information Knowledgebase)[7] 등이 그 대표적인 예이다.

그러나 앞서 언급한 메타데이터 레지스트리 관리 시스템들의 경우, ISO/IEC 11179 메타 모델의 필수 조건을 만족시키지 못하거나 일부 기능만을 제공하는 단점을 지닌다. 또한 시스템의 많은 핵심 요소들이 공개되지 않음으로써 국내 메타데이터 레지스트리 구축 및 개발을 위한 시스템 개발에 어려움이 따른다. 국내의 경우, 최근 메타데이터 레지스트리를 위한 연구가 일부 진행되었으나 단순히 메타데이터 레지스트리 정의 방법과 같은 기초적인 단계에 머물러 있다[8,9].

이 논문의 목적은 기존 관리 시스템들의 단점을 해결하고 ISO/IEC 11179의 표준 명세에 따라 메타데이터 레지스트리를 관리 및 운영할 수 있는 시스템 개발에 있다. 특히 과학기술정보 메타데이터를 위해 가장 필수적으로 요구되는 핵심 컴포넌트 개발에 초점을 두었다. 또한 적용 도메인 특성을 고려한 부가 기능을 확장 영역에 추가하였다.

개발된 시스템은 ISO/IEC 11179의 메타 모델을 준수하므로 다른 분야에서의 메타데이터 레지스트리 관리

시스템 개발을 위한 구현 모델로서 이용될 수 있다. 또한 핵심 기능들이 컴포넌트 기반으로 설계 및 구현되었기 때문에 이 논문의 적용 도메인인 과학기술정보 분야는 물론 다른 분야를 위한 메타데이터 레지스트리 구축 및 개발에 재사용이 가능하다. 현재 e-Business 분야에서도 ISO/IEC 11179를 준수하는 공유 및 교환 시스템의 표준을 개발하고 있다[10,11,12]. 따라서 향후 이와 같은 분야에 이용함으로써 문서 공유 및 교환을 위한 시스템 개발을 용이하게 할 수 있다.

이 논문의 구성은 다음과 같다. 제2장에서는 메타데이터 레지스트리 개념과 기존 관리 시스템의 특징에 대하여 기술한다. 제3장에서는 개발한 시스템의 기능 및 데이터 모델링에 대하여 기술하고 제4장에서는 전체 시스템 구조와 주요 컴포넌트에 대하여 기술한다. 제5장에서는 구현 내용을 기술하고 마지막으로, 제6장에서 결론 및 향후 연구에 대하여 기술한다.

2. 관련연구

이 장에서는 ISO/IEC 11179와 메타데이터 레지스트리의 관계에 대하여 간략하게 언급하고 기존 메타데이터 관리 시스템의 특성 및 문제점에 대하여 기술한다.

2.1 ISO/IEC 11179와 메타데이터 레지스트리

ISO/IEC 11179는 특정 도메인에 종속적인 메타데이터 레지스트리 접근 방법(Domain-dependent Registry, Domain-specific Registry)의 문제점, 즉 메타데이터 포맷의 다양성으로 인해 발생하는 메타데이터간 의미와 형식의 이질성 문제를 해결하기 위해 ISO/IEC 11179에 의해 개발되었다. ISO/IEC 11179의 기본 단위는 데이터 요소로서 하나의 데이터에 대한 표현, 이해 및 활용을 위해 필요한 다양한 메타데이터로 구성되어 있다. 이러한 데이터 요소의 집합이 메타데이터 레지스트리(MDR, Metadata Registry)이다[1].

ISO/IEC 11179는 메타데이터 레지스트리 구축, 개발 및 운영을 위한 메타 모델을 제시하고 있으며, 데이터간 의미 교환 및 공유를 향상시킨다. 또한 응용 분야간의 전체적인 데이터 요청과 상호 교환을 위한 메커니즘 제공, 표준 데이터 요소를 위한 단일 식별자 생성 규칙 제공 등의 장점을 지닌다. 이러한 장점으로 인해 현재 가장 활발하게 연구되고 있는 e-Business 분야에서도

ISO/IEC 11179를 준수하는 문서 관리 표준을 개발하고 있다[10,11,12].

ISO/IEC 11179는 메타 모델이다. 따라서 실질적인 활용을 위해서는 이 표준에서 제시하고 있는 다양한 프로세스 및 규칙들을 수용하여 메타데이터 레지스트리를 구축 및 관리할 수 있는 시스템 개발이 요구된다. 현재 다양한 관리 시스템이 여러 분야에 걸쳐 개발되었으며, 이에 대해서는 다음 절에서 상세하게 기술한다.

2.2 기존 메타데이터 레지스트리 관리 시스템

앞서 언급하였듯이, ISO/IEC 11179는 많은 장점을 지닌다. 현재 다양한 분야에 걸쳐 많은 메타데이터 레지스트리들이 구축되었으며, 또한 이를 관리하기 위한 메타데이터 레지스트리 관리 시스템들이 개발 및 운영되고 있다.

가장 대표적인 예로서, 미국 환경청(Environment Protection Agency)의 환경 정보를 위한 EDR (Environmental Data Registry), 호주 건강 복지 기관인 AIHW (Australian Institute of Health & Welfare)의 NHIK (Australian National Health Information Knowledgebase)[5, 6], 미국 국방성, 건강국 및 보건증진 재단 등이 공동으로 개발한 USHIK(U.S. Health Information Knowledgebase)[7] 등이 있다.

EDR은 환경 데이터에 대한 참조 정보로서 정의, 자원, 그리고 환경 데이터의 사용에 대한 포괄적이고 권위 있는 참조 정보들이다. EDR 시스템은 크게 BioRS (Biology Registry System), SRS(Substance Registry System), CRS(Checmical Registry System), TSN (Taxonomic Serial Number), MetaPro 및 TRS (Terminology Reference System) 등으로 구성되어 있다. 그러나 EDR의 경우, 메타데이터 레지스트리의 확장 즉, 새로운 데이터 요소를 생성하는 프로세스 등이 개발되어 있지 않으며, 데이터 요소의 상태 등이 표준 명세를 따르지 않고 있다. 이는 결국 메타데이터 레지스트리 간 불일치를 야기하게 된다.

그림 1은 EDR의 전체적인 구성 요소를 보여준다.

NHIK는 호주의 건강 복지 기관인 AIHW의 주도로 설계, 구축 및 개발하였으며 건강 관련 메타데이터에 대한 전자 저장소와 질의 도구로 구성되어 있다. NHIK는

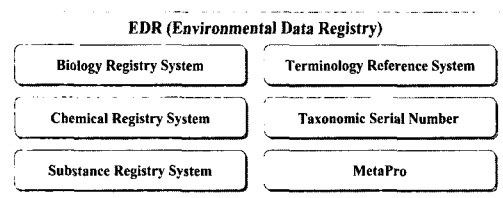


그림 1 EDR의 주요 구성 요소

국가 건강 정보 모델(National Health Information Model), 데이터 사전(National Health Data Dictionary) 및 건강 정보 협정(National Health Information Agreement) 등을 제공한다. 그러나 ISO/IEC 11179에서 제시하고 있는 필수적인 구성 요소를 제공하지 않는 등 여러 가지 규칙을 준수하지 않는다. 이는 메타데이터 레지스트리간 이질성 문제를 초래하게 된다. 부가적으로 데이터 요소를 이해하고 활용하기 위해 필요한 컴포넌트들이 부족하다는 문제점을 지닌다.

USHIK는 호주의 AIHW와 유사한 목적으로 개발되었으며 ISO/IEC 11179를 준수하도록 개발되었다. USHIK는 AIHW의 모델과 밀접하게 연결되어 상호 데이터 요소간의 관계성을 제공한다. 이 외에도 미국 교통부 (Department of Transportation)의 ITS (Intelligent Transportation System)을 예로 들 수 있다[13, 14]. 이 시스템은 여러 개의 데이터 사전으로 구성되어 있으며 ITS 서비스 분야간 비호환성 및 비상호운용성을 해결하기 위해 개발되었다. 그러나 앞서 기술한 시스템들과 유사한 문제점을 지닌다.

3. 기능 모델링 및 데이터 모델

이 장에서는 개발한 시스템의 기능과 이를 위해 요구되는 데이터 모델에 대하여 기술한다.

3.1 기능 모델링

이 절에서는 시스템의 기능적 모델링 결과에 대하여 기술한다. 이를 위해 먼저 기존 시스템들에 대한 분석 결과를 기반으로 정의된 기능들을 기술하고 주요 기능들을 객체지향 모델링 언어인 UML을 이용하여 표현한다[15,16].

3.1.1 표준 지원 측면에서의 시스템 범위

제안하는 시스템 개발의 목적 중 하나는 기존 시스템의 일차적인 문제점인 미비한 표준 준수성을 보완하기 위한 것이다. 이 논문에서 제안하는 시스템에서 지원하는 표준 내용을 요약하면 크게 몇 가지 측면으로 고려할 수 있다.

우선, 기존 시스템들의 경우 데이터 요소가 지나야 하는 필수 속성들을 일부만 지원하고 있다. 그러나 제안하는 시스템은 ISO/IEC 11179에서 명세하고 있는 데이터 요소의 필수 속성들을 모두 지원하고 있다.

둘째로, 기존 시스템들은 데이터 요소를 개념적으로 그룹화하는 핵심 요소들, 즉 데이터 요소 개념(Data Element Concept), 객체 클래스(Object Class), 개념 도메인(Conceptual Domain) 등을 거의 지원하지 않는다. 그러나 제안하는 시스템에서는 이러한 핵심 요소들을 지원하고 있다.

셋째, 기존 시스템들은 데이터 요소의 상태를 임의로

정의하여 이용하고 있다. 그러나 제한한 시스템의 경우, 표준 명세서에서 정의하고 있는 6가지 상태를 그대로 지원하고 있다.

마지막으로, 기존 시스템에서는 메타데이터 레지스트리의 생성, 이용, 평가 및 관리 등에 관여하는 사용자들 임의로 정의하고 있다. 그러나 제한하는 시스템에서는 ISO/IEC 11179에서 권고하는 사용자 그룹을 모두 지원하고 이에 따라 사용자 권한 그룹을 정의하여 이용하고 있다.

3.1.2 메타데이터 레지스트리 생성 및 운영 프로세스

메타데이터 레지스트리 관리 시스템의 기능은 ISO/IEC 11179의 표준 명세와 밀접한 관련성을 지닌다. 따라서 메타데이터 레지스트리를 생성하고 운영하는 전체 프로세스에 대한 세부적인 분석 작업이 필수적으로 선행되어야 한다. 메타데이터 레지스트리는 데이터 요소들로 구성되어 있다. 하나의 데이터 요소가 생성, 이용 및 폐기되는 전체 프로세스는 크게 데이터 요소 제출, 검증, 등록, 상태 관리 및 버전 관리 등으로 구성된다.

데이터 요소 제출 프로세스는 새로운 데이터 요소 후보를 제안하는 단계이다. 이 단계에서, 데이터 요소의 필수 속성들은 반드시 유효한 값을 지녀야 한다[1]. 유효한 값이 주어졌는지 등에 대한 확인 작업이 데이터 요소 검증 및 후보 등록 프로세스에서 이루어진다. 후보로 등록된 데이터 요소는 'RECORDED', 'QUALIFIED', 'STANDARD', 'PREFERRED' 및 'RETIRED' 상태를 거치게 된다. 이 중 'RECORDED'와 'RETI-

RED' 상태를 제외한 나머지 상태일 경우에만 사용자들에게 공개되어 이용된다. 데이터 요소는 개정될 수 있으며 이러한 경우 버전 관리 프로세스가 요구된다. 더 이상 이용되지 않는 데이터 요소일 경우 폐기되며 상태는 'RETIRED'로 전이하게 된다.

그림 2는 이와 같은 전체 프로세스를 보여준다.

3.1.3 관리 시스템 기능 정의

이 절에서는 기존 메타데이터 시스템들의 분석 결과와 메타데이터 레지스트리의 생성 및 운영을 위한 프로세스를 기반으로 정의한 시스템의 주요 기능들에 대하여 기술한다. 이 논문에서는 정의된 기능들을 크게 사용자 영역과 관리자 영역으로 분류하였다.

사용자 영역은 메타데이터 레지스트리를 이용하거나 새로운 데이터 요소를 생성함으로써 확장하는 기능들의 집합이다. 관리자 영역은 데이터 요소들을 논리적인 단위로 집산화 하는 개념들을 생성 및 관리하거나 사용자에 대한 권한을 부여하는 기능들의 집합을 의미한다. 표 1은 전체적인 기능들을 요약한 것이다.

그림 3은 앞서 정의한 기능들을 UML의 유스케이스 다이어그램을 이용하여 표현한 것이다. 그림 3에서, 액터 즉, 사용자를 다섯 개의 그룹으로 분류하였음을 알 수 있다. 보다 단순하거나 또는 세분화된 사용자 그룹을 정의할 수 있으나 이 논문에서 ISO/IEC 11179에 따라 이와 같이 분류하였다. 그림 3의 유스케이스 다이어그램을 통해 각 사용자의 역할 및 권한을 쉽게 이해할 수 있다. 보다 상세한 사용자 그룹별 권한과 부여 규칙에

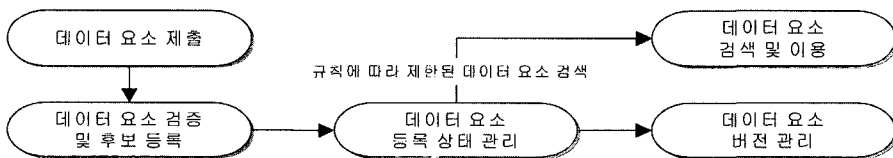


그림 2 데이터 요소 생명주기

표 1 관리 시스템을 위한 주요 기능 정의

영역 분류	기능 정의	기능 설명
사용자 영역	데이터 요소 검색	다양한 옵션으로 데이터 요소를 검색하는 기능
	그룹 (집산화) 요소 검색	논리적 집산화 단위 요소에 대한 검색 기능
	데이터 요소 제출	새로운 데이터 요소를 제안하는 기능으로서 일반적으로 모든 사용자에게 권한이 부여됨
	데이터 요소 상태 변경 제안	데이터 요소의 상태 변경을 제안하는 기능으로서 권한이 제한됨
	스키마 셋 제안	선정 도메인에서의 활용성을 위해 추가된 집산화 개념으로서, 새로운 셋을 제안하는 기능
관리자 영역	투표 기능	제안된 데이터 요소, 요소 상태, 스키마 셋, 개정 등에 대한 투표 기능
	데이터 요소 버전 검색	데이터 요소의 버전을 검색하는 기능
	그룹 요소 관리	집산화 요소인 객체 클래스, 요소 개념, 개념 영역 및 스키마 셋 등을 관리하는 기능
	기관 및 회원 관리	기관 정보 및 회원 관리 기능으로 권한 부여 등의 다양한 세부 기능 포함
	활동 히스토리 관리	특정 위원의 활동 기록을 검색할 수 있는 기능

대해서는 다음 절에서 다룬다.

3.1.4 사용자 분류 및 권한 규칙

이 논문에서는 사용자를 크게 일반 사용자 (Read-Only User), 제안자 (Submitter), 전문가 집단 (Steward), 등록자 (Registrar) 및 집행관리위원 (Executive & Control Committee) 등으로 분류하며, 이는 ISO/IEC 11179를 따른 것이다. 그러나 이 표준의 메타 모델에는 각 사용자 그룹들간의 역할에 대한 명확한 정의가 기술되어 있지 않다. 따라서 이 논문에서는 각 그룹의 역할을 보다 명확히 정의하여 권한 부여 규칙을 생성한다.

권한 부여 규칙은 각 사용자들에게 부여되는 권한을 할당하는 규칙으로 사용자들이 이용하고 조작할 수 있는 서비스의 종류를 의미한다. 권한 규칙은 인공지능 분야에서 전통적으로 사용되어 온 IF~THEN 규칙을 이용하여 정의되어 있다.

다음은 간단한 권한 부여 규칙의 예를 보여준다.

<권한 부여 규칙 1>

IF user.group = 001

THEN user.serviceType = DE.search AND GE.search

위 권한 부여 규칙에서, user.group은 사용자의 권한 그룹 코드를 지니는 필드를 나타내며 '001'은 정의된 권한 코드를 나타낸다. user.serviceType은 사용자의 서비스 타입을 지니는 필드로서 리스트 구조를 이룬다. DE.search와 GE.search는 각각 데이터 요소 검색과 그룹 요소 즉, 집단화 요소에 대한 검색을 의미한다. 즉, 권한 그룹 코드가 '001'인 사용자는 검색 기능만을 이용할 수 있는 권한을 부여 받게 된다.

권한 부여 규칙은 내부적으로 그룹 코드와 기능의 사상관계를 나타내는 GFM(Group-Function Map)으로 구성되어 있다. GFM은 사상테이블과 같은 역할을 수행하며, 메타데이터 관리 시스템에 요구되는 권한 규칙이 방대하지 않기 때문에 직관적이고 단순한 구조를 표현하기에 적절하다.

다음 표 2는 GFM의 구성 예를 보여준다.

표 2에서, 행(기능코드)과 열(그룹코드)이 교차하는

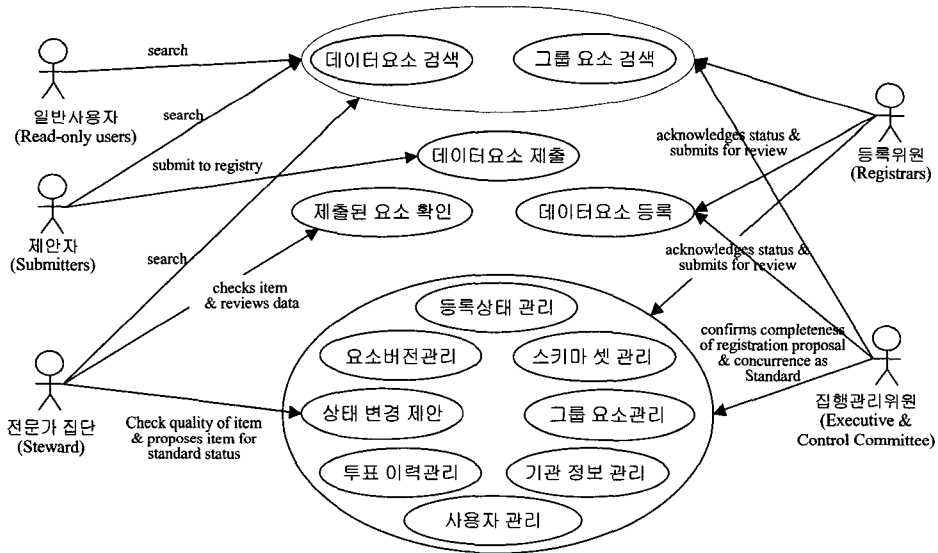


그림 3 메인 유스케이스 다이어그램

표 2 GFM 구성 예

T: TRUE; F: FALSE

기능코드 그룹코드	DE.search	GE.search	DE.submit	DE.register	DE.status	CD.certify	DE.delete
001	T	T	F	F	F	F	F
002	T	T	T	F	F	F	F
003	T	T	F	T	F	F	F
004	T	T	T	F	T	T	F
005	T	T	T	F	T	F	T

DE: Data element; GE: Group element; CD: Candidate(Submitted data element)

셀의 항목이 'T'인 경우, 해당 그룹이 교차되는 기능에 대한 권한을 지님을 의미한다. 예를 들어, '001'과 'DE.search'가 교차하는 셀의 값이 'T'이므로 그룹 001에 포함되는 모든 사용자는 데이터 요소 검색 권한을 지니게 된다. 앞서 기술한 권한 부여 규칙을 이용하여 그룹 003에 속하는 임의의 사용자의 권한 부여 규칙은 다음과 같다.

<권한 부여 규칙 2>

```
IF user.group = 003
THEN user.serviceType = DE.search AND GE.
search AND DE.register
```

3.2 정보 모델링

모델링 결과를 바탕으로 메타데이터 레지스트리 관리 시스템을 위한 데이터 모델을 정의하였다. UML을 이용하여 세부적인 데이터 모델을 기술하였으며 CASE 툴은 현재 널리 이용되고 있는 IBM사의 Rational Rose를 이용하였다[17,18]. 이 논문에서는 정보 모델링을 위한 메타 모델에 대하여 기술하고 구현 레벨의 세부 데이터 모델링 결과에 대해서는 생략한다.

그림 4는 정보 모델링을 위해 필요한 정보들에 대한 메타 모델을 보여준다. 표현된 메타 모델은 개념 모델링 레벨 또는 설계 레벨에서 추출되는 모델로서 실제 구현 단계에서 이용될 수 있는 구체적인 모델의 주요 스키마와 스키마간 관계성을 표현한다.

4. 시스템 구조 및 주요 컴포넌트

이 장에서는 구현된 과학기술정보를 위한 메타데이터 관리 시스템의 구조와 주요 세부 컴포넌트들에 대하여 기술한다.

4.1 시스템 구조

개발한 메타데이터 레지스트리 관리 시스템은 크게 4개의 계층(MDR 계층, 핵심 컴포넌트 계층, 확장 계층 및 서비스 인터페이스 계층)으로 구성된다.

그림 5는 전체적인 시스템 구조를 보여준다.

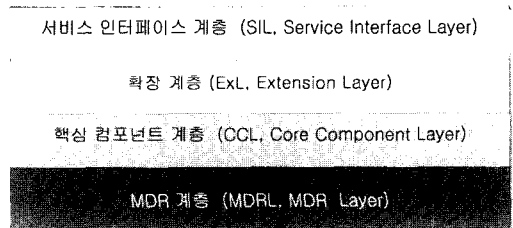


그림 5 시스템 구조: 4계층 구조

MDR 계층(MDRL, Metadata Registry Layer)은 데이터 요소의 집합인 메타데이터 레지스트리, 데이터 요소의 속성으로 관계성을 지니는 기관 및 개인 등의 정보, 객체 클래스, 데이터 요소 개념 등과 같은 집단화 요소인 그룹 요소 등으로 구성된다.

핵심 컴포넌트 계층(CCL, Core Component Layer)은 데이터 요소 및 관련 개념 요소들을 생성하고 확장하기 위해 요구되는 컴포넌트들로 구성된다. 이 계층의 컴포넌트들은 MDRL과 가장 밀접한 연관성을 지니며, ISO/IEC 11179의 표준 명세를 따른다.

확장 계층(ExL, Extension Layer)은 해당 도메인을 특성을 고려하고 활용성을 높이기 위해 추가된 컴포넌트들로 구성된다. 이 논문의 선정 도메인인 과학기술정보 분야의 경우, 데이터 요소들을 필요에 따라 논리적으로 집단화하는 개념이 요구된다. 이를 위해 추가된 개념이 스키마 집합이다. 이러한 추가된 개념과 이를 지원하

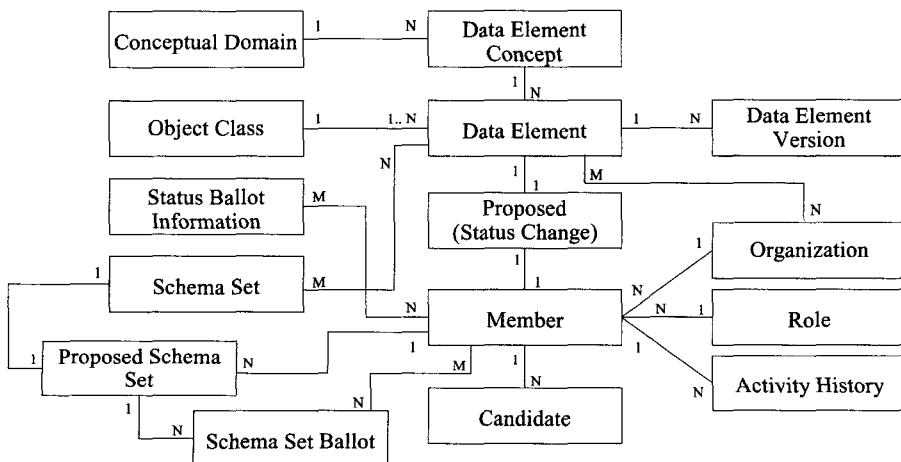


그림 4 주요 정보를 위한 메타 모델

는 컴포넌트들이 ExL에 포함된다.

마지막으로, 서비스 인터페이스 계층(SIL, Service Interface Layer)은 사용자와 시스템간의 통신 채널 역할을 수행하는 계층이다. 사용자의 권한에 따른 화면 제어 등이 이 계층에 속한다.

4.2 주요 시스템 컴포넌트

이 절에서는 시스템의 각 계층을 구성하는 주요 컴포넌트에 대하여 기술한다. 앞서 언급하였듯이, 구현된 시스템은 CCL, MDRL, ExL 및 SIL 등으로 구성되며 각 계층은 여러 개의 세부적인 컴포넌트들로 이루어진다.

그림 6은 SIL을 제외한 각 계층의 세부 컴포넌트들을 보여준다. SIL은 사용자와의 인터페이스를 위한 계층으로 응용, 적용 도메인 및 서비스의 형태에 따라 달라질 수 있기 때문에 이에 대한 상세한 내용은 생략한다.

그림 6에서, MDRL은 크게 MDR과 그룹 요소, 기관 및 사용자에 대한 정보 그리고 권한 할당 규칙 등을 포함한 지식베이스로 구성된다. 가장 중요한 요소는 MDR과 그룹 요소, 즉 집단화 요소로서 메타데이터 레지스트리 관리 시스템의 주 처리 대상이다. 따라서 이들을 관리하기 위한 컴포넌트들이 요구되며 이를 핵심 컴포넌트(Core Component)라고 한다.

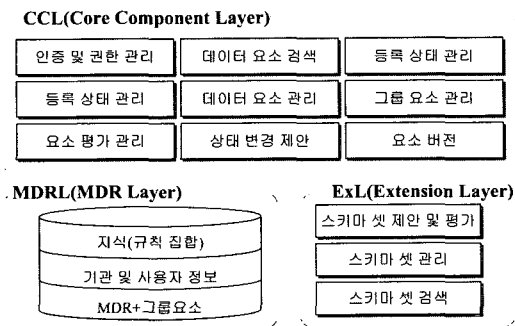


그림 6 시스템을 구성하는 주요 컴포넌트

CCL은 핵심 컴포넌트의 집합으로 구성되며, 메타데이터 레지스트리를 구축하기 위해 필수적으로 요구되는 공통 컴포넌트들이다. 이들 컴포넌트들은 데이터 요소와 그룹 요소에 대한 제안, 검증, 등록, 상태 변경, 이들에 대한 검색 및 개정을 위해 필수적으로 요구된다. 따라서 이 논문의 결과인 핵심 컴포넌트들은 향후 다른 메타데이터 관리 시스템 개발을 용이하게 하고 많은 비용을 절감할 수 있는 효과를 지닌다.

이 논문의 적용 분야인 과학기술정보 분야의 경우, 지식정보 개발의 정확성과 용이성을 위해 스키마 정보를 제공한다. 이 스키마 정보는 데이터베이스 설계를 위한 템플릿으로서 여러 도메인 전문가들이 생성하는 데이터

의 표준 양식이다. 따라서 템플릿들은 과학기술정보를 위해 구축된 데이터 요소들을 논리적으로 집단화한 결과이다. ExL은 이를 지원하기 위한 추가적 확장 컴포넌트들을 지닌다. 즉, 확장된 컴포넌트들은 도메인에 종속된 서비스를 지원하기 위한 시스템 모듈이다.

5. 시스템 구현

이 장에서는 개발된 메타데이터 레지스트리 관리 시스템의 구현 환경과 세부적인 시스템 구조 및 실제 구축된 메타데이터 레지스트리에 대하여 기술한다. 또한 구현된 시스템을 이용하여 이 논문의 대상 적용 분야인 과학기술정보 메타데이터 레지스트리를 생성하는 과정을 보인다.

5.1 개발 환경 및 메타데이터 레지스트리 예

이 절에서는 개발된 시스템의 개발 환경, 세부적인 시스템 구조 및 과학기술정보를 위해 구축된 메타데이터 레지스트리의 예를 보인다.

5.1.1 시스템 개발 환경

제안된 시스템은 윈도우즈 플랫폼을 기반으로 개발되었으며 웹을 기반으로 개발되었다. 개발 언어로는 Java와 웹과의 인터페이스를 위해 Servlet 및 JSP가 이용되었다. 또한 데이터의 관리를 위한 DBMS로는 Oracle이 이용되었다. 개발 환경을 정리하면 다음과 같다.

- 개발 플랫폼: 윈도우즈 시리즈 (2000, XP)
- 사용자 인터페이스: 웹 기반
- 개발 언어: JSP, Servlet, Java
- DBMS: Oracle 9i
- 모델링 언어: UML, E-R 모델
- 모델링 툴: Rational Rose 2002

5.1.2 과학기술정보를 위한 메타데이터 레지스트리 예

이 절에서는 과학기술정보 분야의 실제 데이터와 함께 이를 위해 정의, 구축된 메타데이터 레지스트리의 예를 보인다. 실제 데이터는 한국과학기술정보 연구원에서 과학기술정보 관리를 위해 정의한 데이터이다. 다음 표는 실제 데이터와 이를 위해 정의한 데이터 요소와의 관계를 보여준다.

위의 표에서, "F_DN"과 "AN"은 실제 데이터베이스의 스키마 이름이며 서로 다른 데이터베이스에 정의되어 있다. 의미가 동일하지만 서로 다른 필드 명을 사용하고 있으며 또한 내부적으로 표현 길이 등에도 차이가 있다. 따라서 통합 관리 측면에서 이질성을 지니고 있고 향후 새로운 데이터베이스 설계를 위한 표준 지침을 제공하지 않음으로써 반복적인 이질성 문제를 야기시키게 된다.

이와 같은 문제점을 해결하고 일관성 있는 정보 구축 및 관리를 표준적인 필드 즉, 데이터 요소를 정의하였

표 3 기존 시스템들과 제안 시스템의 비교

실제 데이터		해당 데이터 요소	설명
F_DN	AN	Bibliography number	서지 정보에 부여되는 식별코드
F_PY	PY	Publishing year	출판된 연도를 나타내는 필드
F_LA	LA	Language	논문, 연구 보고서 등의 기술을 위해 사용된 언어
F_TI	TI	Title	논문, 연구 보고서 등의 제목을 나타냄
F_AU	PNH	Author name	저자의 이름
F_KW	KW	Keyword	논문, 연구 보고서 등의 키워드
F_CC		Subject classification	주제가 별 분류를 위한 값
F_CY		Publishing country	출판된 국가 이름(국가 코드가 아닌 전체 이름)

다. 정의된 데이터 요소는 데이터베이스 설계를 위한 표준 지침으로 서 이용된다. 만일 기 정의된 데이터 요소 집합 즉, 메타데이터 레지스트리에서 원하는 필드를 찾을 수 없을 경우, 제안하는 관리 시스템을 통해 동적으로 요청하여 생성하여 이용할 수 있다.

5.2 사용자 및 권한 생성

개발된 메타데이터 관리 시스템을 이용하여 과학기술 정보 메타데이터 레지스트리를 구축하기 위해서는 관련된 여러 종류의 사용자들을 생성해야 한다. 이를 위해서는 사용자 그룹 생성, 사용자 생성, 사용자 그룹 지정 및 권한 규칙 생성 등의 프로세스가 요구된다. 생성된 사용자들은 각각의 권한에 따라 서비스의 이용 권한과 함께 일정의 역할을 수행하게 된다.

그림 7은 위의 전체 실행 과정을 보여준다.

그림 7에서 <7-A>는 권한 코드 즉, 권한 그룹을 생성한 결과 화면이다. 생성된 권한 그룹을 특정 사용자에게 부여한다. <7-B>는 사용자 'user2'에게 권한 그룹 'registrar'를 부여하는 화면을 보여준다. 따라서 사용자 'user2'는 데이터 요소 검색, 데이터 요소 제출, 스키마 셋 제안 등의 권한을 지닌다. 반면, 데이터 요소 관리나

사용자 관리 등의 기능은 사용할 수 없다. <7-C>는 'user2'에게 권한을 부여한 결과를 보여준다.

5.3 과학기술정보 메타데이터 레지스트리 구축

과학기술정보 메타데이터 레지스트리를 구축하기 위해서는 데이터 요소의 제출, 제출된 데이터 요소에 대한 검증 및 등록 과정이 순차적으로 수행된다. 등록된 데이터 요소는 관련 전문가와 위원들에 의해 상태 변경이 제안되어 보다 정제된 상태로 전이하게 된다. 이러한 과정을 거쳐 데이터 요소들이 생성 및 추가됨으로써 메타데이터 레지스트리가 구축된다.

그림 8은 데이터 요소의 제출 과정에서부터 등록 상태가 변경되는 과정을 전체적으로 보여준다.

그림 8에서, <8-A>는 제출자 (Submitter)가 데이터 요소를 제안하기 위해 입력해야 하는 필수 속성과 입력 예를 보여준다. 주어진 정보는 데이터 요소 후보로 제출되며 제출된 후보는 확인 작업을 거쳐 데이터 요소로 등록된다. 초기에 등록된 데이터 요소의 상태는 'Recorded'이며 일반적으로 이 상태의 데이터 요소는 서비스되지 않는다. <8-B>는 검증 작업을 거친 후보가 데이터 요소로 등록된 상태를 보여준다.

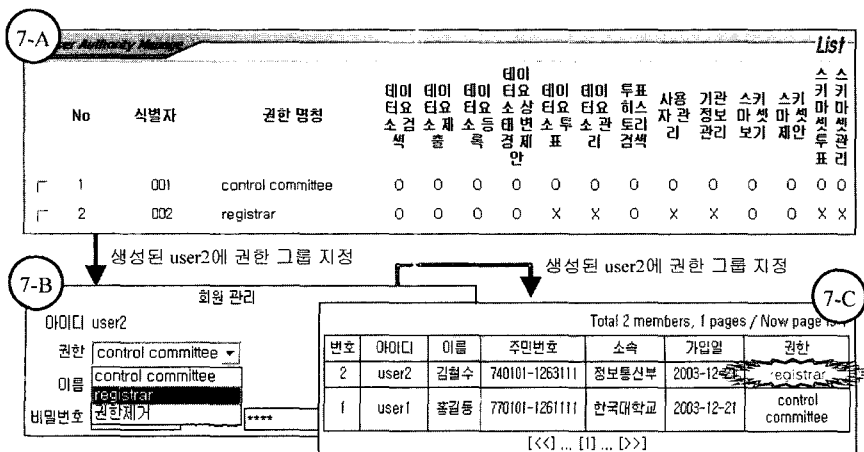


그림 7 사용자 및 권한 생성 과정

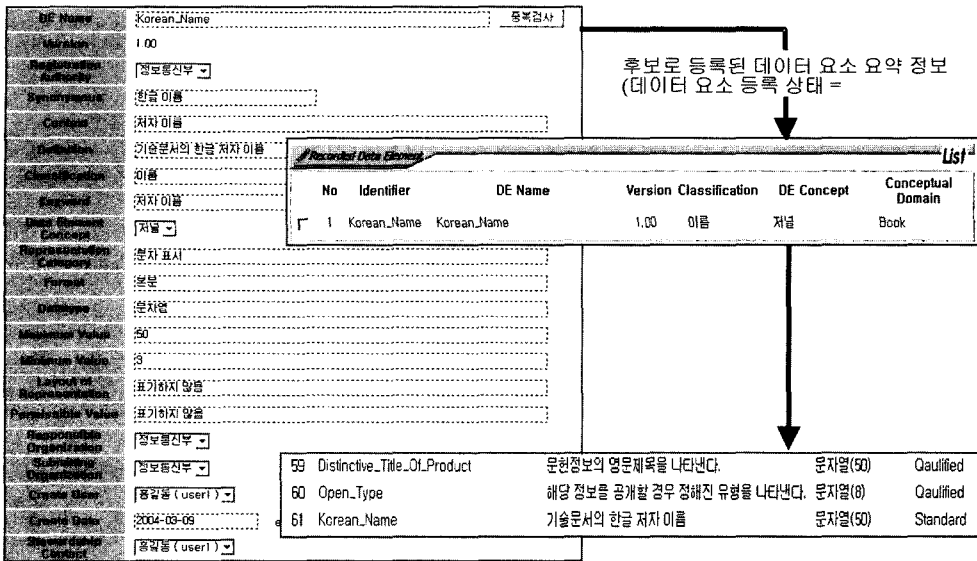


그림 8 데이터 요소 생성 및 등록 상태 전이 과정

등록된 데이터 요소는 위원들의 평가를 통해 서비스 가능한 상태로 전이된다. 이를 위해서는 먼저 임의의 권한을 지닌 전문가가 특정 데이터 요소의 상태 변경 제안을 위해 타당성과 함께 제안을 한다. 제안된 데이터 요소는 해당 분야의 전문 위원들의 평가를 통해 상태가 전이된다. <8-C>는 초기에 등록했던 데이터 요소 (Korean_Name)의 상태가 'Recorded'에서 'Standard'로 변경되었음을 보여준다.

5.4 데이터 요소 및 그룹 요소 검색

과학기술정보를 위한 메타데이터 레지스트리가 구축 되면 이는 새로운 데이터베이스 설계시 표준 지침서로 이용된다. 지침서, 즉 메타데이터 레지스트리의 이용을 용이하도록 하기 위해서는 설계하고자 하는 데이터베이스 필드의 유효한 값 도메인과 의미가 동일한 데이터 요소를 검색할 수 있는 서비스가 필수적으로 요구된다.

이미 언급하였듯이, ISO/IEC 11179에는 데이터 요소들을 논리적인 개념으로 집단화할 수 있는 그룹 요소들을 지낸다. 이러한 그룹 요소는 데이터 요소에 대한 디렉토리 검색 서비스를 제공한다.

그림 9는 데이터 요소를 검색을 위한 시스템 초기 화면과 주어진 키워드를 이용하여 검색한 결과를 보여준다. 또한 검색된 데이터 요소 중 특정 데이터 요소의 속성들에 대한 상세 정보를 보여준다.

그림 9에서, <9-A>는 데이터 요소 검색 초기 화면을 보여주며, 검색 범위, 키워드 및 검색 옵션으로 구성되어 있다. 검색 범위는 특정 주제를 대상으로 검색하는

연산과 유사하며, 데이터 요소의 속성들과 그룹 요소들이 검색 범위로 사용된다. <9-B>는 데이터 요소 이름이 주어진 키워드로 시작되는 데이터 요소에 대한 검색 결과를 보여준다. 마지막으로, <9-C>는 검색 결과 중에서 선택된 데이터 요소인 'Record_Reference_Number'에 대한 상세 속성 정보를 보여준다.

지금까지 개발된 관리 시스템을 이용하여 레지스트리 구축 및 검색과 같은 주요 과정에 대하여 기술하였다. 이 외에도 개발된 메타데이터 레지스트리 관리 시스템은 과학기술정보 분야에 종속적인 서비스인 스키마 셋 관리 기능, 기관 및 사용자 관리, 버전 관리 및 이력 관리 등 다양한 기능을 제공한다.

5.5 비교 평가

이 절에서는 기존 시스템들과 이 논문에서 제안하는 시스템간의 비교 평가 내용에 대하여 기술한다. 다음 표는 비교평가 내용을 요약한 것이다.

비교 대상 시스템들은 이미 언급한 미국 환경청의 EDR, 호주의 NHIK, 미국 국방성의 USHIK, 그리고 미국 교통부의 ITS 등이다. 위의 표에서 제안된 시스템과 기존 시스템들은 모두 메타데이터 관리를 위해 ISO/IEC11179를 기반으로 개발되었다. 그러나 세부적인 표준 준수성이 미흡하고 기능성이 부족하다는 단점을 지닌다.

기존 시스템들의 경우, ISO/IEC 11179에서 명세하고 있는 핵심 요소들, 즉 데이터 요소 개념(Data Element Concept), 객체 클래스(Object Class), 개념 도메인

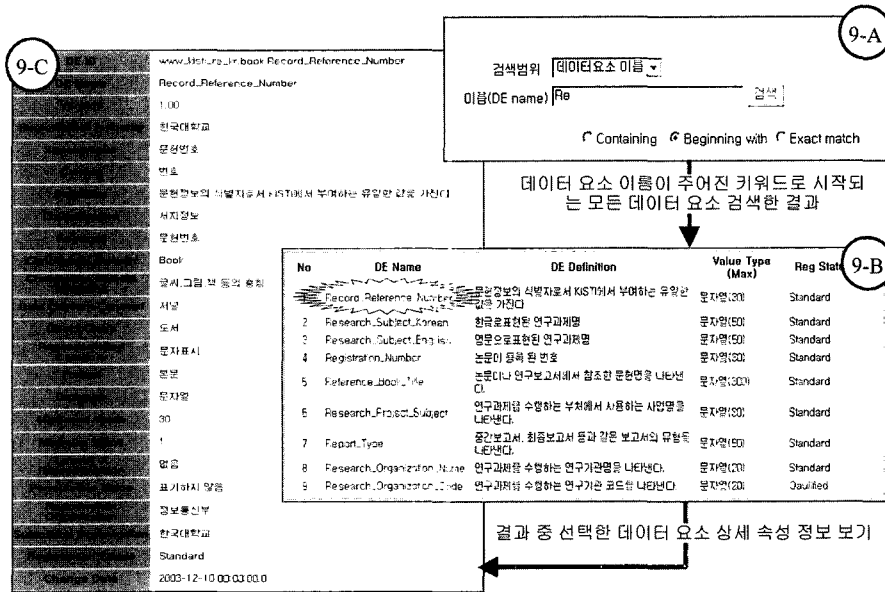


그림 9 데이터 요소 검색 기능

표 4 기존 시스템들과 제안 시스템의 비교

비교 항목	기존 시스템	제안 시스템
표준 준수	ISO/IEC 11179	ISO/IEC 11179
핵심 요소들의 정의 기능	N/A	지원
핵심 요소들에 대한 검색 기능	N/A	지원
명명 규칙의 준수성	일부 시스템만 준수	준수
데이터 요소의 필수 속성 지원	일부 필수 속성만을 지원	권고하는 모든 속성 지원
데이터 요소의 검색 기능	일부 시스템만 지원	지원
사용자 유형 준수 여부	일부 시스템만 준수	준수
핵심 요소와의 연계 검색 기능	N/A	지원
재사용성	N/A	컴포넌트화

(Conceptual Domain) 등에 대한 정의 및 검색 기능을 제공하고 있지 않다. 이는 표준성 저하와 직결되며 다른 메타데이터 레지스트리와의 상호운용시 많은 비용을 초래할 수 있다. 그러나 제안한 시스템에서는 이러한 핵심 요소를 정의하여 내부 구조에 수용하였으며, 이를 검색하는 기능도 제공하고 있다.

메타데이터 레지스트리의 구축에 있어 가장 중요한 요소는 데이터 요소이다. ISO/IEC 11179에서는 데이터 요소들이 반드시 지나야 하는 속성들을 정의하고 있다. 그러나 EDR 시스템을 제외한 대부분의 시스템들은 많은 필수 속성들을 제공하지 않는다. 이 또한 표준의 준수성을 저하시키고 결국 데이터에 대한 명확한 이해를 어렵게 하여 이용성을 저하시키게 된다. 제안한 시스템에서는 필수 속성들을 모두 수용함으로써 표준을 준수하고 있다. 이는 데이터 요소에 대한 충분한 설명과 함께 이용에 대한 용이성을 제공한다. 또한 향후 다른 메

타데이터 레지스트리와의 상호운용시 또 다른 불일치 문제를 일으키지 않는다는 장점을 지닌다.

기존 시스템들 중 일부 시스템의 경우, 데이터 요소에 대한 검색 기능조차 제공하지 않고 있다. 이는 메타데이터 레지스트리를 이용과 활용에 큰 저해 요인이 된다. EDR의 경우에는 다양한 검색 기능을 제공한다. 그러나 핵심 요소와의 연계를 통한 데이터 요소 검색 기능을 제공하지 않음으로써, 특성에 따른 분류별 검색이 불가능하다. 그러나 제안된 시스템에서는 핵심 요소에 대한 검색 기능은 물론 핵심 요소와의 연계성을 고려한 데이터 요소 검색을 제공한다. 이는 데이터 요소의 구조와 개념을 쉽게 이해하고 활용할 수 있도록 해 준다.

마지막으로, 제안된 시스템 컴포넌트를 기반으로 개발되었다. 즉, 메타데이터 레지스트리 관리를 위해 필요한 필수 기능들을 컴포넌트로 개발하였다. 이는 다른 분야에서 해당 분야의 데이터를 위한 메타데이터 레지스트

리 관리 시스템 개발을 재활용함으로써 동일한 기능 개발에 드는 비용을 절감할 수 있다. 따라서 개발 기간 또한 단축할 수 있는 장점을 지닌다.

6. 결론 및 향후 연구

이 논문에서 메타데이터 레지스트리 관리 시스템을 설계하고 개발한 결과를 기술하였다. 또한 개발된 관리 시스템을 이용하여 과학기술정보를 위한 메타데이터 레지스트리를 구축한 내용을 보였다.

개발된 시스템은 메타데이터 레지스트리 계층(MDRL, Metadata Registry Layer), 핵심 컴포넌트 계층(CCL, Core Component Layer), 확장 계층(ExL, Extension Layer), 서비스 인터페이스 계층(SIL, Service Interface Layer) 등으로 구성되어 있다. MDRL과 CCL은 ISO/IEC 11179에서 명세하고 있는 필수 공통 요소들로 구성되어 있기 때문에 다양한 분야에서의 메타데이터 레지스트리 관리 시스템 개발시 재사용이 가능하다. SIL은 이 논문의 적용 도메인인 과학기술정보 분야에서 요구되는 추가적인 기능들로 구성되어 있다. 따라서 SIL은 서로 다른 컴포넌트들로 구성될 수 있는 확장 영역이다.

기존에 개발된 메타데이터 레지스트리 시스템들의 경우, 서비스 측면에서 기능이 미비하거나 ISO/IEC 11179의 표준 명세를 정확하게 준수하지 않는 문제점들을 지니고 있다. 이 논문에서는 이러한 표준을 준수하면서 필수적으로 요구되는 다양한 기능을 설계하고 구현하였다. 개발한 시스템은 컴포넌트 기반으로 개발되었다. 이는 다른 분야를 위한 메타데이터 관리 시스템 개발시 많은 시간과 비용을 절감할 수 있다.

향후에는 보다 다양한 분야에 개발된 시스템의 컴포넌트들을 적용함으로써 추가적인 핵심 컴포넌트를 개발하고 정의된 컴포넌트들을 정제하는 작업이 요구된다. 또한 기 개발된 메타데이터 레지스트리 관리 시스템들과의 상호 연동을 위한 연구가 요구된다. 이를 위해서는 다중언어 지원, 데이터 요소들간의 사상, 메타데이터 레지스트리 관리 시스템간의 통신 프로토콜 등이 연구되어야 한다.

참 고 문 헌

- [1] ISO/IEC JTC 1/SC 32, "ISO/IEC 11179: Specification and standardization of data elements, Part 1~6," 2003.
- [2] Harman, J., "Overview of EPA's System of Registries," SoR Users Conference, National Center for Supercomputing Applications, Virginia, U.S., March 20, 2003.
- [3] Environmental Protection Agency (EPA), "Data Standards Publications and Guidances," 2003.
- [4] Eversole, S., "Overview of EPA's System of Registries: Environmental Information Management System (EIMS)," SoR Users Conference, National Center for Supercomputing Applications, Virginia, U.S., March 20, 2003.
- [5] Australian Institute of Health and Welfare (AIHW), "Australian National Health Information Knowledgebase," <http://www.aihw.gov.au/>.
- [6] Australian National Health Data Committee, "National Health Data Dictionary," 2003.
- [7] Health Informatics Standards Board(HISB) Standard Development Organizations (SDO), "United States Health Information Knowledgebase," <http://www.usih.org/>.
- [8] ETRI, Research on the Registration and Search System of Component, Research Report, 2000.
- [9] KISTI, A Study on the Development of Standardization and Management Model for Science and Technology Information, Research Report, 2002.
- [10] Organization for the Advancement of Structured Information Standards (OASIS), "ebXML Technical Architecture Specification v1.04," Specification, February 2001.
- [11] Organization for the Advancement of Structured Information Standards (OASIS), "Registry Information Model v2.0," Specification, November 2001.
- [12] ebXML Business Process Project Team, "Proposed Revisions to ebXML Technical Architecture Specification v1.0.4," White Paper, May 2001.
- [13] U.S. Department of Transportation, "U.S. Transportation System," <http://www.dot.gov/>.
- [14] ITS Architecture Development Team, "ITS Logical Architecture- Volume I, Volume II: Process Specifications, Volume III: Data Dictionary," 2003.
- [15] Booch, G., Rumbaugh, J., and Jacobson I., The Unified Modeling Language User Guide, Addison-Wesley, 1999.
- [16] Object Management Group (OMG), "OMG Unified Modeling Language Specification," Version 1.5, March 2003.
- [17] IBM, Rational Rose, <http://www-306.ibm.com/software/rational/>.
- [18] Quatrani, T., Visual Modeling with Rational Rose 2002 and UML, 3rd Ed., Addison Wesley, 2003.



정 동 원

1997년 군산대학교 전산과(학사). 1998년 한국전자통신연구원(위촉연구원). 1999년 충북대학교 전산과(석사). 1999년~2000년 ICU 부설 한국정보통신교육원(전임강사). 2001년~현재 (주)라임미디어 테크놀로지(연구원). 2002년~현재 TTA 표준화위원회(특별위원). 2004년 고려대학교 컴퓨터학과(박사) 2004년~현재 고려대학교 컴퓨터학과(연구교수). 관심분야는 분산 컴퓨팅, 이동 에이전트, 데이터베이스, 메타데이터 레지스트리, 유비쿼터스



신 동 길

2000년 광운대학교 컴퓨터공학과(학사) 2002년~현재 고려대학교 대학원 컴퓨터학과 석사과정. 관심분야는 이동 에이전트, 데이터베이스, 메타데이터 레지스트리



정 은 주

2001년 한성대학교 컴퓨터공학 학사 2003년~현재 고려대학교 대학원 컴퓨터학과 석사과정. 관심분야는 이동 에이전트, 정보 보호, 메타데이터 레지스트리

김 영 갑

정보과학회논문지 : 컴퓨팅의 실제
제 10 권 제 3 호 참조



이 정 욱

1992년 고려대학교 컴퓨터학과 학사 1994년 고려대학교 대학원 컴퓨터학과 석사. 2001년 고려대학교 대학원 컴퓨터학과 박사. 2002년~현재 건국대학교 자연과학대학 컴퓨터·응용과학부 교수. 관심분야는 데이터베이스, 정보통합, 정보 검색, 지능형에이전트, 웹기반정보시스템 등

백 두 권

정보과학회논문지 : 컴퓨팅의 실제
제 10 권 제 3 호 참조