

데이터 가시성 기반의 점진적 메타데이터 레지스트리 구축 방법론

(A Progressive Metadata Building Methodology based on Data Visibility)

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

이 정 옥 ^{***} 서 태 설 ^{****} 백 두 권 ^{****}
(Jeong Oog Lee) (Tae-Sul Seo) (Doo-Kwon Baik)

요 약 메타데이터 레지스트리는 동적인 메타데이터 관리와 다양한 이질적인 메타데이터간 상호운용성 증진을 위한 목적으로 개발되었다. 구축된 메타데이터 레지스트리는 새로운 데이터베이스 설계시 표준 지침을 제공함으로써 근본적이고 데이터 통합 메커니즘을 제공한다. 그러나 대규모 분산 데이터베이스에 대한 점진적인 통합을 위한 상황에서, 기 구축된 데이터베이스에 대한 인덱스를 생성하고 일관성 있는 표준 지침서인 메타데이터 레지스트리를 구축하는데 한계가 있다. 특히, 각 단위 데이터베이스에 대한 이용률 통계 정보 부재 및 단위 기간 내 주어진 비용이 제한된 경우, 우선적으로 통합할 데이터베이스 선택을 통한 점진적인 메타데이터 레지스트리의 체계적인 구축 방법을 제공하지 않는다. 이 논문에서는 이와 같은 환경에서, 점진적으로 메타데이터 레지스트리를 구축할 수 있는 방법론을 제안하고 점진적인 메타데이터 레지스트리 구축을 위해 데이터 가시성과 계층적 메타데이터 레지스트리 구조를 정의한다. 또한 정의한 개념, 구조 및 방법을 실제 도메인에 적용하기 위해 개발된 시스템과 적용한 결과를 기술한다.

키워드 : 데이터 요소, 메타데이터 레지스트리, 메타데이터, 메타데이터 구축 방법론, 계층적 메타데이터 레지스트리, ISO/IEC 11179

Abstract Metadata Registry was developed to dynamically manage metadata and to increase interoperability between various and heterogeneous metadata. The built metadata registry can be used as a standard guideline for creation of new databases and it provides a radical data integration mechanism. However, in the situation that an enormous databases must be integrated progressively, there is a limit to the existing metadata-based approach. In case that each database has no statistical information for its use rate and the restricted cost is given to us for a unit time, existing metadata-based approaches do not provide how to select some databases to be preferentially integrated and to build a metadata registry progressively. In this paper, we propose a methodology that can create progressively metadata registries in the case. The proposed methodology is based on data visibility and hierarchical metadata registry. We also describe the system that have been developed for applying the methodology to a real domain, and then described its results.

Key words : Data element, Metadata registry, Metadata, Metadata building methodology, Hierarchical metadata registry, ISO/IEC 11179

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

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

^{***} 정 회 원 : 건국대학교 컴퓨터·융합과학부 교수
lio@kku.ac.kr

^{****} 비 회 원 : 한국과학기술정보연구원 실장
tsseo@kisti.re.kr

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

논문접수 : 2003년 7월 21일

심사완료 : 2003년 9월 29일

1. 서론

현재까지 분산되어 있는 데이터베이스를 통합하기 위한 많은 연구들이 진행되어 왔으며 대부분의 접근 방법들은 분산되어 있는 단위 데이터베이스들을 일관성 있게 접근할 수 있는 인덱스와 메타데이터를 이용한다. 최근에는 동적인 메타데이터 관리와 다양한 메타데이터들 간의 상호운용성을 증진을 위하여 국제 표준 기구인 ISO/IEC JTC 1/SC 32에 의해 ISO/IEC 11179가 개발되었다[1].

ISO/IEC 11179는 다양한 메타데이터에 대한 등록, 인증 및 표준화 과정을 통해 메타데이터 레지스트리(MDR, Metadata Registry)를 구축하고 메타데이터 레지스트리에 따라 새로운 데이터베이스를 생성할 수 있는 메커니즘을 제공한다. 즉 ISO/IEC11179는 동적이고 일관성 있는 데이터 요소 관리를 위한 메커니즘을 제공한다. 그러나 대규모 데이터베이스를 대상으로 할 경우, 단위 기간에 주어진 제한된 비용과 각 데이터베이스 이용률에 대한 통계 정보가 없는 환경 하에서, 메타데이터 레지스트리를 구축할 수 있는 메커니즘을 제공하지 않는다는 단점을 지닌다. 즉 통계 정보를 지니지 않는 기존의 데이터베이스를 점진적으로 통합하면서 표준 메타데이터 레지스트리를 구축하는 방법론을 제공하지 않는다.

이러한 ISO/IEC11179의 단점을 해결하고 점진적인 메타데이터 레지스트리 구축하기 위해서는 먼저 특정 기준에 따라 대상 데이터베이스 집합을 순차적으로 선택해야 한다. 단위 데이터베이스는 그 중요성에 따라 우선순위가 결정되며, 그 중요성은 응용 분야에 따라 달라질 수 있다. 예를 들어, 과학기술 정보 응용 분야에서, 통합할 데이터베이스의 우선순위는 데이터베이스들의 이용률에 따라 우선적으로 결정된다. 그러나 각 단위 데이터베이스의 이용률에 대한 통계 정보가 없을 경우, 일차적으로 데이터베이스 집합을 선택하기 위한 선택 기준이 없다.

이 논문에서는 이용률에 대한 통계 정보를 지니지 않은 대규모 분산데이터베이스를 대상으로 단위 시간에 주어진 비용이 제한된 환경 하에서, 점진적으로 데이터베이스를 통합하고 메타데이터 레지스트리를 구축하기 위한 방법론을 제안한다. 이 논문에서는 과학기술정보를 제공하는 검색서비스 응용 분야를 대상으로 한다. 각 단위 데이터베이스에 대한 우선순위 결정을 위해서 데이터 가시성(Data Visibility)이란 개념을 정의하였다. 또한 점진적인 메타데이터 레지스트리 구축 및 관리를 위하여 계층적 메타데이터 레지스트리(Hierarchical Metadata Registry) 구조를 이용한다. 마지막으로 제안하는 방법론을 실제 데이터베이스를 대상으로 하여 적용한

결과와 이를 위한 시스템 구조를 기술하고 기존 메타데이터 레지스트리 기반 접근 방법과의 비교평가 결과에 대하여 기술한다.

이 논문의 구성은 다음과 같다. 제2장에서는 기존 통합 방법론 및 메타데이터 레지스트리의 개념과 특성에 대하여 기술한다. 제3장에서는 이 논문에서 정의한 데이터 가시성과 계층적 메타데이터 레지스트리 구조에 대하여 기술하고 제안하는 모델과 이를 위한 시스템 구조는 제4장에서 기술한다. 제5장에서는 구현된 시스템, 적용 사례 및 기존 MDR 기반 접근 방법과의 비교평가 내용을 기술하고 마지막으로, 제6장에서는 전체적인 연구결과에 대하여 정리하고 향후 연구 방향에 대하여 간략하게 언급한다.

2. 관련 연구

2.1 기존 메타데이터 구축 방법론

메타데이터 구축을 통한 데이터 통합 방법들 중에서 가장 일반적인 방법론은 온톨로지 기반의 접근 방법이며 유사한 많은 방법들이 연구되어 왔다. 다양한 지역 스키마들로부터 하나의 전역 스키마를 구축하는 방법 [2-5]에서부터 연방 통합 방법[6], 분산 객체 기술을 이용한 접근 방법[7-9], 중계기 기반의 접근 방법[10,11] 및 사례 기반 접근 방법[12] 등이 있다. 이러한 접근 방법들은 기본적으로 상향식 접근 방법으로서 지역 데이터베이스에 대한 공개 정도에 따라 다양한 접근 방법들로 분류된다[13].

그러나 이들 방법들은 임의의 지역 데이터베이스 중 일부가 변경되거나 새로운 데이터베이스가 추가될 경우, 상호간의 연관성을 일관성 있게 유지하기 위하여 많은 비용을 요구한다. 또한 다양성과 데이터베이스 규모가 증가할수록 유지에 대한 오버헤드가 급격하게 증가하는 단점을 지닌다.

이러한 문제점을 해결하기 위하여 제시된 방법론이 표준화된 지침서를 하향식으로 정의하여 제공하는 접근 방법이다. 표준화된 지침서에 따라 데이터베이스를 생성하기 때문에 온톨로지를 기반으로 하는 방법에 비해 유지관리 비용이 매우 절감되는 장점을 지닌다. 이러한 접근 방법으로 다양한 분야에서 각 분야에 적합한 많은 표준 지침서를 정의하여 제공하고 있다. 특히, 과학기술 정보와 관련하여 개발된 표준 지침서의 예로는 MARC [14], RDF[15], Dublin Core[16], ONIX[17] 등이 있다.

위 표준 지침서들의 경우, 특정 도메인에 종속되어 있고 또한 사용자의 요구가 동적으로 반영되지 못하는 단점을 지닌다. 즉 데이터베이스 생성자가 생성하고자 하는 속성을 해당 표준 지침서에서 제공하지 않을 경우, 이를 반영할 수 있는 메커니즘을 제공하지 못한다는 문

제점을 지닌다.

ISO/IEC JTC 1에서는 동적인 메타데이터 관리 메커니즘을 제공하지 못하는 문제점을 해결하기 위하여 메타데이터 레지스트리라는 개념을 정의하고 ISO/IEC 11179[1]를 개발하였다. ISO/IEC 11179에 대해서는 다음절에서 상세하게 기술한다.

2.2 ISO/IEC 11179

ISO(the International Organization for Standardization)와 IEC(the International Electrotechnical Commission)는 메타데이터 포맷을 일관성 있게 관리하기 위하여 ISO/IEC JTC 1 설립하여 ISO/IEC 11179 [1]를 개발하였다. ISO/IEC 11179는 공유 데이터 환경 생성 및 기존 데이터 관리 방법론에 비해 시간과 비용이 절감될 수 있도록 하기 위한 목적으로 개발되었다.

ISO/IEC 11179는 모든 개체, 즉 데이터와 그 개체를 식별할 수 있는 메타데이터를 하나의 데이터 요소(Data Element)로 표현한다. 또한 데이터 요소를 이해하고 공유할 수 있도록 하기 위하여 데이터 요소에 대한 등록과 표준화에 대한 방법을 기술하고 있다. 데이터 요소는 ISO/IEC 11179에서 가장 중요한 개념으로 하나의 데이터에 대한 정의, 식별, 표현 및 허용 값들을 나타내는 최소 단위이다.

그림 1은 ISO/IEC 11179의 기본 개념을 보여준다.

데이터 요소는 객체 클래스(Object Class), 특성(Property) 및 표현(Representation)으로 구성되어 있다. 객체 클래스는 명확한 경계와 의미로 식별될 수 있는 실세계의 개념, 추상적 내용 또는 사물의 집합이다. 특성은 한 객체 클래스의 모든 구성원들이 지니는 공통적인 속성을 의미한다. 표현은 데이터가 표현되는 방식을 의미하며, 값 영역, 데이터 타입 등을 지닌다.

데이터 요소는 데이터의 최소 단위로서 관계형 데이터베이스의 필드, 즉 속성에 해당한다. 그림 1에서, 객체 클래스는 관계형 모델의 개체에 해당하며 객체 지향 모델의 클래스에 해당한다. 그림 2는 ISO/IEC 11179에 이용된 용어와 개념들이 전통적인 데이터 모델에서 사용되는 용어와의 관련성을 명확하게 보여준다.

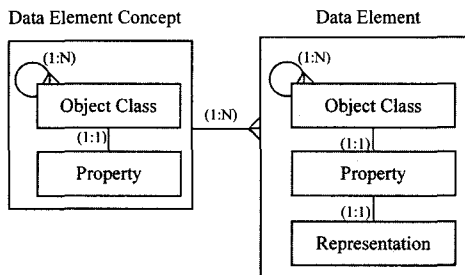


그림 1 데이터 요소의 개념 구조

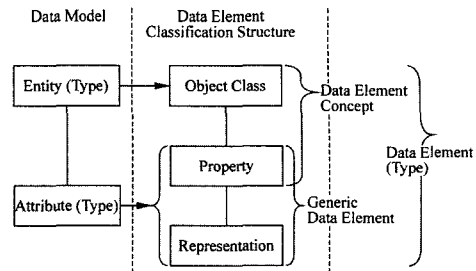


그림 2 데이터 요소와 관계형 모델의 관계

데이터 요소는 ISO/IEC 11179의 최소 단위로서 실질적으로 정의된 데이터 요소의 집합을 메타데이터 레지스트리(MDR, Metadata Registry)라고 한다. ISO/IEC 11179의 개념에서 온 메타데이터 레지스트리는 상호운용, 동적 메타데이터 관리 및 표준화 측면에서 많은 장점을 지닌다. 그러나 제한된 비용과 통계 정보가 없는 데이터베이스들에 대한 메타데이터 구축에 한계가 있다. 또한 메타데이터 레지스트리가 기존의 하향식 방법론에 비해 독립적인 메커니즘을 제공한다. 그러나 실제 응용을 위해서는 상향식 접근 방법의 접목이 요구된다.

따라서 이 논문에서는 단위 시간 내 주어진 비용이 제한적이고 각 단위 데이터베이스들의 이용에 대한 통계 정보가 없는 상황에서 점진적으로 메타데이터를 구축하고 표준화된 지침서인 메타데이터 레지스트리를 구축할 수 있는 방법론을 제안한다. 또한 메타데이터 레지스트리 관리를 위해 구현된 시스템과 구현된 시스템을 이용하여 실제 응용 도메인에 제안된 방법론을 적용한 구현 사례에 대하여 기술한다.

3. 데이터 가시성과 계층적 MDR 구조

이 장에서는 데이터 가시성(Data Visibility)의 개념을 정의하고 계층적 메타데이터 레지스트리(HMDR, Hierarchical Metadata Registry)의 구조 및 데이터 가시성과의 관계성에 대하여 기술한다.

3.1 데이터 가시성 정의

현재까지 메타데이터 레지스트리와 관련하여 연구되고 제안된 대부분의 연구 방법들의 경우, 단순히 기술적인 개발과 가장 정상적인 환경 속에서 이루어졌기 때문에 실제 응용 환경에 적용하는데 여러 가지 문제점을 야기한다. 특히, 제한된 비용과 데이터베이스들의 이용률에 대한 통계 정보가 없는 환경에서 메타데이터 레지스트리를 성공적으로 구축하는 방법에 대한 연구가 거의 이루어지지 않았다.

단위 시간 내에 주어진 비용이 한정되어 있는 상황에서 표준화된 지침서를 생성을 위해서는 일차적으로 가장 우선순위가 높은 데이터베이스들을 선택하고 이들에

대한 분석을 통해 공통 데이터 요소 추출하고 정의해야 한다. 일차 과정이 완료되면 다음 우선순위를 지닌 데이터베이스를 선택하여 위 과정을 반복적으로 수행해야 한다. 그러나 통계정보가 없는 데이터베이스들을 대상으로 할 경우 우선순위를 결정할 수 없게 된다. 이 논문에서는 이와 같은 상황에서, 단위 데이터베이스들의 우선순위를 결정할 수 있는 기준을 결정하는 방법으로서 데이터 가시성을 정의한다.

데이터 가시성(Data Visibility)이란 데이터의 전문화 정도(Data Level, Degree of Data Specialization)와 사용자의 해당 도메인에 대한 전문 지식 수준(User Level, Degree of User Specialization)을 고려하여 데이터의 이용률(Data Use Rate)을 결정하는 개념이다. 데이터의 전문성은 데이터의 난이도를 의미한다. 예를 들어, 일반적인 동물 이름이나 외형에 대한 이미지 정보 등은 해당 도메인에 대한 지식을 지니고 있지 않아도 모든 사람이 쉽게 이해할 수 있는 정보들이다. 그러나 단백질의 서열과 같은 정보는 해당 도메인에 대한 지식이 있는 전문가들이 이해하고 이용하고 활용할 수 있는 정보들이다. 한국데이터베이스진흥센터의 데이터베이스 이용 패턴 조사[18]에 따르면, 사용자들이 가장 많이 이용하는 분야가 교양, 문화 및 일상생활 분야였으며, 가장 낮은 분야가 전문 분야들로 이루어져 있다. 다시 말해, 일반적으로 쉽게 이해할 수 있는 데이터에 대한 검색 요구가 가장 높음을 알 수 있다. 이는 사용자 레벨, 즉 사용자의 전문성은 해당 도메인에 대한 전문 지식 정도를 나타낸다. 따라서 일반적으로 데이터 레벨과 사용자 레벨에 따라 데이터의 이용률이 달라지게 된다는 것을 알 수 있다.

앞서 언급하였듯이, 데이터 레벨이 낮은 경우, 전문가는 물론 일반 사용자도 이용할 수 있으며 데이터 레벨이 높은 전문적인 지식 정보의 경우 일반적으로 전문가들에 의해 이용된다. 사용자 레벨이 높은 전문가의 경우에는 보다 전문화된 지식 정보까지 이용할 수 있으며 사용자 레벨이 낮은 일반 사용자들은 쉽게 이해할 수 있는 단순 정보에 관심을 지니게 된다.

그림 3은 데이터 레벨, 사용자 레벨 및 데이터 이용률과의 관계성을 표현한 것이다.

그림 3에서, 사용자의 전문성, 즉 레벨이 높아질수록 사용자가 이용할 수 있는 데이터 레벨도 높아지며 그 역 또한 성립한다고 가정한다. 이 때 데이터의 이용이 전문가로 한정되기 때문에 그 이용률은 낮아지게 된다. 반면 데이터 또는 사용자의 전문성이 낮아질수록 사상되는 데이터의 사용자들의 수가 증가하기 때문에 이용률은 증가하게 된다. 이러한 개념은 앞서 기술한 데이터베이스진흥센터의 조사 결과[18]를 바탕으로 일반성을

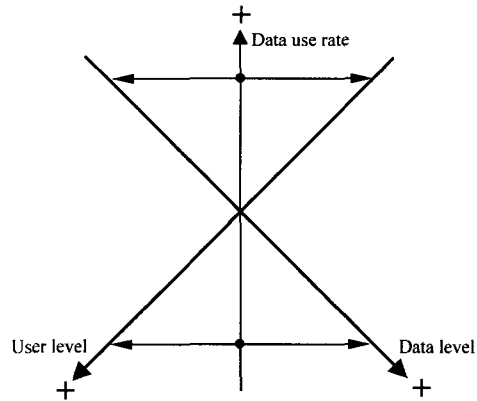


그림 3 데이터, 사용자 레벨 및 이용률간의 관계성

부여한 것이다. 따라서 데이터베이스에 접근하는 집단에 따라 이용률이 달라질 수 있으나, 무작위 선정의 경우에 비해 보다 나은 선택 기준을 제공할 것이다.

이와 같이 사용자 레벨과 데이터 레벨에 따라 데이터에 대한 이용률이 일반적으로 달라지게 된다. 이러한 개념을 이 논문에서는 데이터 가시성이라고 정의하고 값의 영역은 [0 .. 1]로 수렴하도록 정의한다.

정의 1. (데이터 가시성) 데이터 레벨, 사용자 레벨 및 데이터 이용률을 각각 L_D , L_U , U_D 라 정의하고, 산출된 이용률을 데이터 가시성의 값 영역 내로 변환하는 함수를 f_V , 그리고 데이터 레벨과 사용자 레벨을 이용하여 이용률을 산출하는 함수를 f_U 라고 정의한다. 이 때, 데이터 가시성(Data Visibility), V 는 다음과 같다.

$$V = f_V(U_D) = f_V(f_U(L_D, L_U))$$

L_D 와 L_U 는 각각 $(1 \leq L_D \leq 100)$ 와 $(1 \leq L_U \leq 100)$ 사이의 범위를 지닌다. 함수 f_U 는 두 값의 합을 계산하는 함수로서 그 결과 값이 이용률이 된다. 함수 f_V 는 함수 f_U 의 결과 값을 데이터 가시성의 값 범위 내로 변환하는 함수이다. 따라서 이용률은 [0 .. 1] 사이의 범위 내로 수렴하게 된다.

3.2 데이터 가시성과 데이터 계층화

데이터 가시성 개념에 따라서 전체 대상 데이터베이스 집합은 여러 계층으로 나뉘어진다. 사용자의 경우 크게 일반 사용자와 전문가로 분류할 수 있으며, 전문가 그룹은 세분화된 전문 영역으로 인해 보다 세분화된 그룹으로 나뉘어진다. 또한 각 사용자 그룹들은 이용하는 데이터의 레벨이 동일하거나 독립적일 수 있다. 즉, 일반적인 데이터는 일반 사용자 및 전문가들에 의해 이용되며, 데이터 레벨이 높을수록 사용자 그룹과 이들이 이용하는 데이터 집합은 보다 세분화된다.

그림 4는 이러한 관계성을 도식적으로 표현한 것이다. 그림 4에서, 데이터의 계층화는 데이터 레벨과 사용자

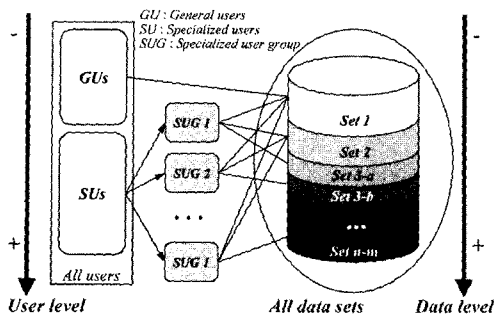


그림 4 데이터 가시성에 따른 데이터 계층화

레벨에 따른 것으로 각각의 레벨이 세분화될수록 데이터의 계층도 증가하게 된다. 데이터 집합, Set 1은 대다수의 사용자에 의해 공통적으로 이용되는 가장 일반적인 데이터를 의미한다. 즉 다수의 데이터 집합 중에서 Set 1의 데이터 가시성이 가장 높으며, 이는 모든 데이터 집합들 중으로 우선순위가 가장 높음을 의미한다. 따라서 메타데이터 레지스트리 구축을 위해 Set 1에 대하여 일차적으로 데이터 요소의 추출 작업이 이루어지게 된다.

데이터 가시성에 의한 전체 데이터의 계층적 분류는 메타데이터 레지스트리의 계층적 구조화와 연관성을 지닌다. 즉 데이터 가시성과 데이터의 계층화 및 이에 따른 데이터의 특성에 의해 데이터 요소들을 계층적으로 분류할 수 있다. 이러한 계층적 분류는 구조적인 복잡성을 증가시키지만 분산 관리, 지역적 이용 메타데이터의 독립적 관리, 특성을 고려한 지역적 메타데이터 운영 등의 많은 장점을 지닌다.

3.3 데이터 가시성에 의한 계층적 메타데이터 레지스트리

데이터 가시성이 높은 데이터 집합의 경우, 대부분의 사용자에게 의해 이용되는 정보임을 의미한다. 이는 데이터에 대한 표준 데이터 요소 정의가 용이하고 표준화의 당위성을 제공한다. 또한 많은 사용자들에 의해 사용되는 정보의 경우, 데이터베이스를 설계할 때 자주 요구되는 스키마 정보임을 의미한다. 즉 데이터베이스 생성과 구축을 위해 일반적으로 요구된다. 그러므로 가시성 높은 데이터에 대한 분석과 이를 통한 데이터 요소의 정의는 일차적으로 요구된다. 또한 대다수의 사용자에게 의해 사용되므로 대부분의 분야에서 공통적으로 이용되는 데이터임을 알 수 있다.

이 논문에서는 대부분의 분야에서 공통적으로 이용되고 데이터베이스 생성에 있어서 이용 가능성이 높은 데이터에 대하여 우선적으로 데이터 요소를 정의하여 메타데이터 레지스트리를 구축한다. 이 때, 가장 가시성이 높은 데이터들에 대한 메타 요소들의 집합을 전역 메

타데이터 레지스트리(GMDR, Global Metadata Registry)라고 정의한다. 그리고 각 분야별 또는 사용자별로 독립적으로 사용되는 정보들에 대한 데이터 요소의 집합을 지역 메타데이터 레지스트리(LMDR, Local Metadata Registry)로 정의한다.

전역 메타데이터 레지스트리는 모든 데이터 집합에 대한 공통된 표준 데이터 요소의 집합이며 전역 지침서로 이용된다. 반면, 지역 메타데이터 레지스트리는 각 부분적인 데이터 집합에 대한 공통 데이터 요소 집합이다. 분야별 또는 독립적인 사용자 그룹에 의해서 사용된다는 특성에 따라 LMDR_i(1 ≤ i ≤ n)개의 지역 메타데이터 레지스트리가 존재하게 된다. 그림 5는 데이터 가시성에 따른 메타데이터 레지스트리의 계층화된 구조를 보여준다.

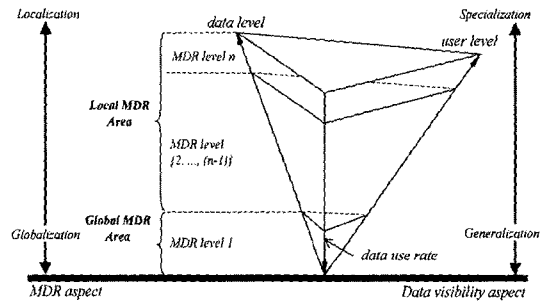


그림 5 계층형 메타데이터 레지스트리 구조

이러한 계층적 메타데이터 레지스트리 구조는 특정 분야의 특성을 고려하여 데이터 요소를 정의함으로써 독립적이고 개별적인 그룹을 위한 표준 데이터 요소 관리 메커니즘을 제공한다.

4. LOG 모델

이 장에서는 앞서 정의한 데이터 가시성과 계층적 메타데이터 레지스트리를 구조를 이용하여 점진적으로 표준 데이터 요소를 정의할 수 있는 방법론과 이를 위해 필요한 시스템 구조 및 일반화 알고리즘에 대하여 기술한다.

4.1 개념적 점진적 메타데이터 레지스트리 구축 모델

이 논문에서는 제안하는 방법론을 LOG(Localized-Global Metadata Registry)라고 정의하며 LOG의 개념 모델은 인터페이스 계층, GMDR(Global MDR) 계층, LMDR(Local MDR) 계층 및 데이터 자원 계층으로 구성된다. 개념 모델의 이러한 분류는 앞서 언급한 데이터 가시성과 계층적 메타데이터 레지스트리 구조에 기반을 두고 있다. 다음 그림은 이러한 개념 모델의 구성 요소를 도식화한 것이다.

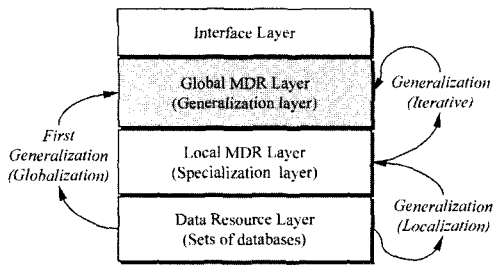


그림 6 LOG 방법론의 개념 모델

그림 6에서, 가장 최상위 계층인 인터페이스 계층 (Interface Layer)은 검색, 질의, 출력 제어 등과 같은 서비스를 지원한다. 인터페이스 계층은 사용자 요청에 대한 적절한 결과를 생성하기 위하여 일차적으로 GMDR 계층을 통해 결과를 요청한다. 만일 원하는 결과를 GMDR 계층에서 제공하지 못할 경우 LMDR 계층을 통해 결과를 생성한다.

두 번째 계층은 가장 공통적인 일반화된 데이터 요소의 집합인 전역 메타데이터 레지스트리를 관리하는 GMDR 계층(Global Metadata Registry Layer)이다. 이 계층에서 관리하는 데이터 요소들은 모든 분야에서 공통적으로 이용되는 데이터에 대한 표준 형식이다. GMDR 계층의 주요 역할은 전역 메타데이터 레지스트리를 관리하고 새로운 일반화된 데이터 요소를 추출하여 추가하는 기능이다. 새로운 데이터 요소는 LMDR의 지역 메타데이터 레지스트리에 정의되어 있는 데이터 요소로부터 추출하게 된다.

LMDR 계층(Local Metadata Registry Layer)은 독립적인 개별 그룹에서만 이용되는 특성화된 표준 데이터 요소들의 집합 지역 메타데이터 레지스트리를 관리하는 계층으로, 단위 데이터베이스로부터 지역적 공통 요소를 추출하고 정의하는 기능을 담당한다. 또한 GMDR의 전역적 공통 데이터 요소를 정의하는 원시 자료원으로 이용된다.

자원 계층(Data Resource Layer)은 분산되어 있는 단위 데이터베이스들의 집합으로서, 실제 데이터를 지니는 계층이다. 자원 계층내의 단위 데이터베이스들 중 일차 대상 집합 선정은 데이터 가시성에 의해 이루어진다.

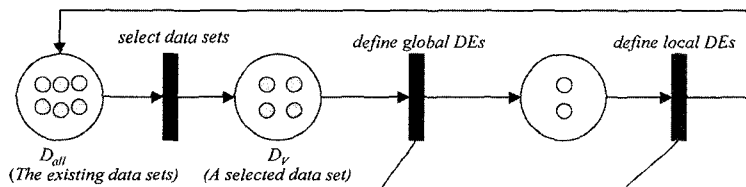
4.2 일반화 알고리즘

지역 메타데이터 레지스트리로부터 전역 메타데이터 레지스트리의 확장을 위해 일반화 알고리즘이 요구된다. 일반화는 데이터 요소의 정의를 통한 데이터 요소 표준화 연산과는 구별된다. 데이터 요소의 정의는 실제 데이터베이스들의 스키마들로부터 공통적인 표준 요소를 정의하는 연산이며 분석 대상은 실제 데이터베이스내의 속성들이 된다. 그러나 일반화는 지역 메타데이터 레지스트리내의 데이터 요소들로부터 하나의 공통된 데이터 요소를 정의하는 것이다. 즉 다양한 분야에서 이미 정의되어 있는 데이터 요소들로부터 하나의 전역 데이터 요소를 추출하는 과정을 의미한다.

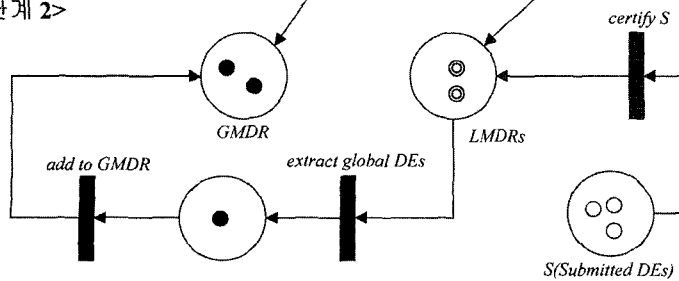
다음 그림 7은 레지스트리 구축 초기부터 데이터 요소에 대한 일반화 과정까지의 전체 프로세스를 페트리 넷 방법을 이용하여 표현한 것이다.

레지스트리 구축 초기에는 전역 데이터 요소 정의를

<단계 1>



<단계 2>



Notations and Symbols

- : Data field of the existing databases
- : New data element submitted
- ⊙ : Local data element in LMDRs
- : Global data element in GMDR
- S : Data element submitted
- DE : Data Element
- D_v : Selected data set by data visibility

그림 7 데이터 요소 정의 및 일반화 프로세스

위해 크게 두 단계의 프로세스가 요구된다. 첫 번째 단계는 실제 스키마들에 대한 비교 분석을 통한 전역 및 지역 메타데이터 레지스트리를 구축하는 단계이다. 두 번째 단계는 기 구축된 지역 데이터 요소들에 대한 일반화 과정을 통해 전역 메타데이터 레지스트리를 확장하는 과정이다.

첫 번째 단계는 기존의 실제 스키마들에 대한 비교 분석 작업을 통한 전역 및 지역 메타데이터 레지스트리를 구축하는 과정이다. 이 과정에서 분석할 데이터베이스의 우선 순위는 이 논문에서 제안하는 데이터 가시성에 의해 결정된다. 따라서 일차적으로 데이터 가시성이 가장 높은 데이터 집합에 대한 분석 작업을 통해 전역 메타데이터 레지스트리를 구축하고, 우선 순위에 따라서 반복적인 과정을 통해 순차적으로 지역 메타데이터 레지스트리들이 구축된다.

<단계 1>이 완료된 후, 지역 메타데이터 레지스트리들로부터 새로운 전역 데이터 요소를 추출하고 정의하는 일반화 과정이 수행된다. 두 번째 단계에서, 새로운 데이터 요소가 제출되고 이에 대한 인증 및 표준화 과정을 통해 개별적이고 독립적인 그룹별 지역 레지스트리의 확장이 이루어진다. 확장된 지역 메타데이터 레지스트리들로부터 새로운 전역 공통 데이터 요소가 추출되어 전역 메타데이터 레지스트리에 추가된다. 이 과정을 데이터 요소에 대한 일반화라고 정의한다.

4.3 시스템 구조

이 논문에서 제안하는 방법론을 위해 개발된 시스템은 크게 네 개의 계층으로 구성된다. 개발된 시스템은 사용자와의 상호작용을 위한 사용자 인터페이스 계층,

전역 메타데이터 레지스트리 관리를 위한 전역 메타데이터 레지스트리(Global MDR) 계층, 지역 메타데이터 레지스트리 관리를 위한 지역 메타데이터 레지스트리(Local MDR) 계층 및 실제 분산 데이터베이스들의 집합인 데이터 자원 계층으로 이루어진다.

그림 8은 시스템의 전체적인 구조와 핵심 컴포넌트들을 보여준다. 사용자 인터페이스 계층은 서비스를 이용하는 서비스 사용자(Service User), 데이터 요소 제출자(Submitter), 등록자(Registrar), 관리위원(Control Committee) 및 관리자/Administrator) 등을 위한 컴포넌트들로 구성된다. 서비스 사용자 컴포넌트는 정보를 검색하고 이용하는 사용자를 위한 기능을 수행한다. 다른 컴포넌트들은 데이터 요소를 등록하고, 등록된 요소 제출, 등록, 평가 및 시스템 관리 역할을 수행한다.

GMDR 계층은 크게 네 가지의 컴포넌트로 구성된다. 전역 MDR(Global MDR)은 이미 앞서 언급했던 전역 메타데이터 레지스트리를 나타내며 전체 응용 도메인에 대해 공통적인 전역 데이터 요소 집합이다. 전역 메타저장소(Global Meta Repository)는 데이터 요소에 사상되는 실제 스키마 정보와 분산 데이터 베이스 위치 정보, 액세스 정보 등을 지닌다. 이 정보는 사용자의 질의에 대한 분산 질의 처리를 위해 이용된다. 전역 지식베이스는 데이터 요소와 실제 스키마 필드간의 사상 규칙, 사용자 권한 규칙, 데이터 일반화를 위한 필요한 추출 및 정의 규칙 등을 관리한다. 전역 MDR 에이전트는 인터페이스와 통신, 일반화를 위한 프로세스 관리, LMDR 계층과의 연계 등 GMDR 계층의 모든 컴포넌트를 관리하는 코어 컴포넌트이다.

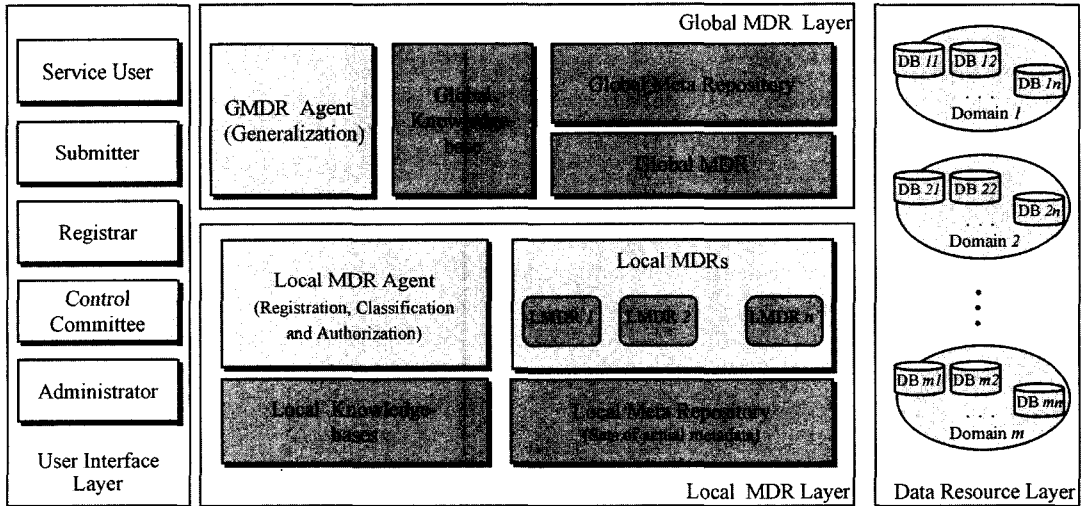


그림 8 LOG 방법론을 위한 시스템 구조 및 핵심 컴포넌트

LMDR 계층의 컴포넌트들은 GMDR 계층과 거의 동일하다. 즉 지역 MDR(Local MDR), 지역 메타 저장소(Local Meta Repository) 및 지역 지식베이스 등은 GMDR 계층의 전역 MDR, 전역 메타 저장소 및 전역 지식베이스 각각의 기능과 유사하다. 그러나 LMDR 계층의 컴포넌트들은 개별적이고 독립적인 그룹 또는 특정 분야의 사용자들에 의해서만 사용되는 데이터 요소와 관련된 정보들을 관리한다. 지역 MDR 에이전트(Local MDR Agent)은 전역 MDR 계층의 전역 MDR 에이전트 및 인터페이스 계층과 통신을 담당하며 지역 MDR 계층의 모든 컴포넌트들을 관리한다. 특히 지역 MDR 에이전트는 제출자가 제출한 새로운 데이터 요소에 대한 인증, 분류 및 표준화와 같은 ISO/IEC 11179에서 정의한 기본적인 기능을 수행하게 된다. 따라서 LMDR 계층은 물리적으로 분산되어 운영될 수 있다. 그러나 이 논문에서는 이러한 분산 모델에 대해서는 향후 연구 주제로 남겨둔다.

사용자가 검색을 위한 질의를 입력할 경우, 인터페이스 계층의 서비스 사용자 컴포넌트와 GMDR 계층의 전역 MDR 에이전트와의 통신을 통해 질의 결과를 생성하게 된다. 만일 질의에 대한 적절한 결과가 생성되지 않을 경우, 서비스 사용자 컴포넌트는 LMDR 계층의 지역 MDR 에이전트에게 질의 결과를 요청하게 된다. GMDR 계층 및 LMDR 계층의 메타데이터 레지스트리는 분산되어 있는 실제 스키마 필드들에 대한 인덱스 역할을 수행하기 때문에 사용자의 검색 요청에 대해 2단계의 검색 연산을 수행하게 된다.

5. 시스템 구현 및 적용 사례

이 장에서는 LOG 모델과 제안된 시스템 구조를 기반으로 실제 구현된 내용과 과학기술정보 데이터베이스를 대상으로 한 적용 사례에 대하여 기술한다.

5.1 기능 설계 및 범위

LOG 모델을 위해 구현되어야 하는 기능들은 기본적인 메타데이터 요소 관리 기능들과 더불어 일반화를 위한 모듈 및 계층적 메타데이터 레지스트리 관리를 위한 모듈이 추가적으로 요구된다. 그림 9는 LOG 모델을 위한 시스템 구현을 위한 기능적인 특성을 UML 메소드를 이용하여 설계한 것이다.

그림 9는 LOG 모델을 위한 시스템의 주요 유스케이스 다이어그램을 보여준다. 그림에서, 대상 데이터베이스의 선택, 데이터 요소의 추출 및 전역 데이터 요소를 위한 일반화 등을 위해 "Select target databases", "Extract data elements" 및 "Generalization" 유스케이스들이 관여하게 된다. "Data element search", "Registration", "Classification" 및 "Evaluation" 유스케이스들은 각각 제공되는 데이터 요소의 검색, 요소 등록, 요소의 분야 분류 및 평가 등을 위한 프로세스들이며, MDR을 위한 가장 기본적인 기능들이다.

이 논문에서는 제안한 LOG 모델을 한국과학기술정보 연구원에서 관리하는 과학기술정보 데이터베이스에 적용하여 계층적 메타데이터 레지스트리를 구축하였다. 한국과학기술정보연구원은 방대한 데이터베이스를 관리하면서 일반 사용자들에게 다양한 서비스를 제공하고 있다. 제한된 자원과 통계 정보를 지니지 않은 소스 데이터베이스들을 대상으로 하여 데이터 가시성을 적용하였으며, 이를 기반으로 일차적으로 대상 데이터베이스를 선택하

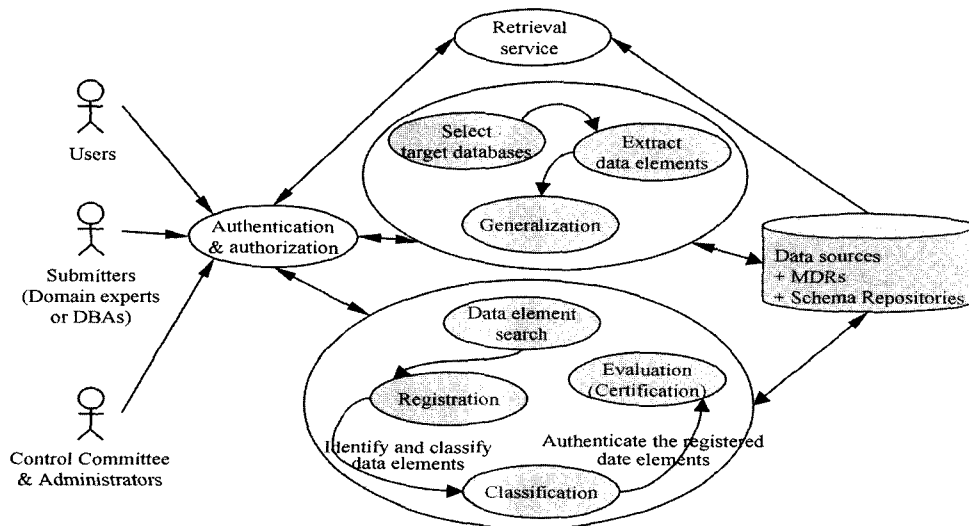


그림 9 LOG 모델을 위한 기능 설계

표 1 선택 대상 데이터베이스 목록

데이터베이스	설명	표기
DIGS	정기 간행물 기사 색인 정보를 관리하는 데이터베이스	D _{DIGS}
DIMD	학위 논문 정보 관리하는 데이터베이스	D _{DIMD}
BIST	과학기술 관련 학술잡지, 회의자료 및 보고서 자료 등을 관리하는 데이터베이스	D _{BIST}
INFO	과학기술 분야의 주요 잡지 정보를 관리하는 데이터베이스	D _{INFO}
THESIS	과학기술 분야의 석박사 학위 논문의 요약 정보와 원문을 관리하는 데이터베이스	D _{THESIS}
KRIST	과학기술 관련 연구개발보고서를 원문과 함께 관리하는 데이터베이스	D _{KRIST}
SOCIETY	학회지 및 학술대회 논문집 등을 원문과 함께 관리하는 데이터베이스	D _{SOCIETY}

였다. 선택된 데이터베이스 목록은 다음 표와 같다.

5.2 메타데이터 관리 시스템 구현

앞서 기술한 선택된 대상 데이터들에 분석을 통해 데이터 요소들을 추출하여 정의하였다. 정의된 데이터 요소들은 새로운 데이터베이스를 생성할 때 표준 지침서로 이용된다. 일차적으로, 각 데이터베이스들로부터 다수의 LMDR을 구축하고 구축된 n개의 LMDR에 일반화 알고리즘을 적용하여 GMDR을 구축하게 된다. 이러한 과정을 위해서는 데이터 요소 등록 및 평가 등의 모듈에 대한 구현이 우선적으로 요구된다. 이들 모듈을 통해 LMDR이 구축되며 구축된 다수의 LMDR로부터 일반화 과정을 통해 전역 데이터 요소 즉, GMDR이 구축된다. 이와 같은 과정을 반복적으로 수행함으로써 점진적으로 GMDR이 확장되며 계층적 메타데이터 레지스트리 구조를 지니게 된다.

그림 10은 데이터 요소 제출 및 평가를 통한 데이터

요소 구축 과정을 보여준다.

그림 10-a에서, 데이터 요소를 제출하기 위해 필요한 기본적인 속성들을 나타낸다. 이러한 속성들은 ISO/IEC 11179를 기반으로 하고 있으며, 식별, 정의, 관계, 표현 및 관리 등 크게 다섯 개의 범주로 분류된다. 식별 속성(Identifying Properties)들은 데이터 요소를 식별하기 위해 필요한 값들을 표현하며 데이터 요소의 식별자나 이름 등을 나타낸다. 정의 속성(Definitional Properties)은 데이터 요소의 의미를 기술하는 속성들이며, 관계 속성(Relational Properties)들은 클래스 또는 다른 속성과의 관계를 기술하는 속성들이다. 표현 속성(Representational Properties)은 데이터 요소의 데이터 타입, 길이 등 보다 물리적인 요소와 관련된 정보들을 기술하고 관리 속성(Administrative Properties)은 데이터 요소를 제출하거나 관리하는 기관 또는 담당자 등의 정보를 기술한다.

(a) 데이터 요소 제출

Property Category	Name	Value	Mandatory
Identifying properties	Name	Announcement_Date	Mandatory
	Identifier	Auto-generated generation	Mandatory
	Version	Auto-generated generation	Conditional
Definitional properties	Registration authority	KRIST	Conditional
	Synonymous name	Publishing data	Optional
Relational properties	Definition	denotes the data that papers or texts are published	Mandatory
	Classification scheme	Literatures	Optional
	Keyword	Publishment	Optional
Representational properties	Object Class	Publishing class	Optional
	Representation category	Character	Mandatory
	Form of representation	Auto-internal generation	Mandatory
	Value format	2003-01-15, January 15, 2003	Mandatory
Administrative properties	Maximum length	20	Mandatory
	Minimum length	10	Mandatory
	Responsible organization	KRIST	Optional
	Registration status	Auto-generated generation	Conditional
	Submitting organization	Meta Creator Inc	Optional
	Create by user name	Dong-won Jeong	Mandatory
	Create date	Auto-generated generation	Mandatory
	Data steward organization name	Dis-Service-Blog	Mandatory

(b) 데이터 요소 평가

제출된 요소에 대한 평가 인터페이스 화면. 평가 위원들에 의한 평가가 이루어진 후, 해당 LMDR을 위한 데이터 요소로 등록된다.

Guideline

1. Select a corresponding area of yours
2. Select a data element submitted
3. If you agree to this proposal, click Approval
4. If not, click Reject and give the reason

Corresponding area: Select your field

New data elements submitted

Area	Submitted	Recorded
논문 저자가 속한 대학명	2	Recorded
문헌의 출판지	2	Recorded

2 개의 요소가 검색되었습니다.

그림 10 데이터 요소 제출 및 평가 프로세스

그림 10-b는 제출되거나 등록된 데이터 요소를 평가하는 인터페이스를 보여준다. 이 과정을 통해 데이터 요소는 등록, 인증 또는 표준화의 단계를 거치게 된다. 일단 등록되어 인증된 데이터 요소는 각 해당 LMDR에 추가되며, 이와 같은 프로세스를 반복적으로 수행하면서 다수개의 LMDR이 구축된다. LMDR들에 대한 구축이 완료되면, 각 LMDR 내에 있는 데이터 요소들에 대한 일반화 과정을 통해 전역 데이터 요소가 정의된다. 즉 점진적으로 GMDR이 구축되며 계층적인 메타데이터 레지스트리 구조를 형성하게 된다.

계층적인 MDR 구축이 완료된 후, 새로운 데이터베이스를 생성하고자 할 경우, 구축된 MDR 내의 데이터 요소를 이용하여 데이터베이스 스키마를 정의하게 된다. 이 때, 원하는 데이터 요소를 찾기 위하여 일차적으로 GMDR을 검색한다. 만일 원하는 데이터 요소를 찾지 못할 경우, 해당 분야의 메타데이터 레지스트리, 즉 LMDR을 검색하게 된다. 제 2단계 검색을 통해 원하는 데이터 요소를 검색하지 못할 경우에는 그림 10-a에서와 같이 직접 데이터 요소를 제출하게 된다.

그림 11은 데이터 요소를 검색하기 위한 프로세스를

구현된 인터페이스를 통해 보여준다.

5.3 비교 평가

ISO/IEC 11179에 기반을 두고 있는 메타데이터 레지스트리를 기반으로 한 접근방법은 동적으로 메타데이터를 관리할 수 있다는 장점을 지닌다. 그러나 구축되어 있는 기존의 데이터베이스들을 통합할 수 있는 메커니즘을 제공하지 않는다. 특히 제한된 비용과 통계 정보를 지니지 않는 데이터베이스들을 대상으로 점진적인 메타데이터 구축 방법을 제공하지 않는다. 또한 지역적으로 분산되어 있는 단위 데이터베이스들간의 관계성을 관리할 수 있는 방법을 제공하지 않는다. 따라서 표현력이 부족하며 다양한 응용 서비스 개발에 부적합한 구조를 지니고 있다.

반면, 이 논문에서 제안하는 LOG 방법론의 경우, 기존 MDR 기반의 접근 방법이 지니는 표준성을 지니고 있으며 MDR이 지니는 상호운용성, 동적 메타데이터 관리 및 표준화 메커니즘 제공 등의 장점을 지닌다. 추가적으로, LOG 방법은 기존 데이터베이스를 통합할 수 있는 메커니즘을 제공하며 제한된 비용과 통계 정보를 지니지 않는 데이터베이스를 대상으로 점진적인 메타데

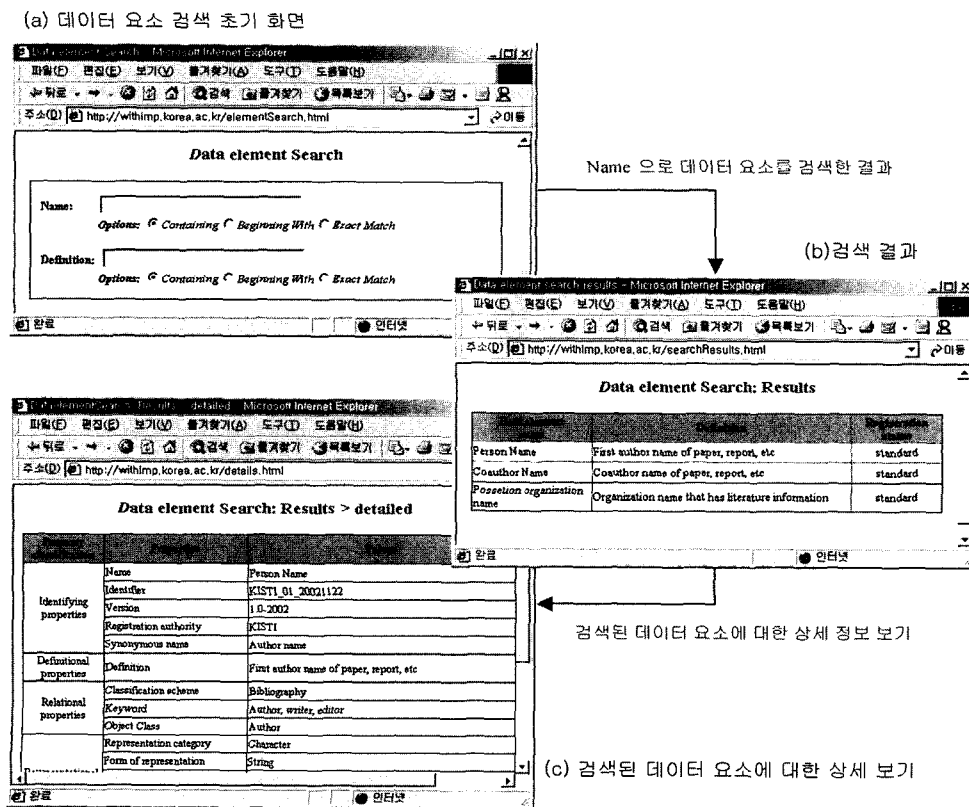


그림 11 데이터 요소 검색 인터페이스

표 2 MDR 기반 모델과 LOG 모델에 대한 비교

평가 항목	기존 MDR 기반 모델	LOG 모델
표준 여부	ISO/IEC 11179	ISO/IEC 11179
상호운용성	high	high
구조적 복잡도	$O(c)$	$O(n(R_{DB}) * n(H_{MDR}))$
동적 데이터 요소 관리 메커니즘	support	support
표현력	weak	strong
기존 DB 통합 메커니즘	N/A	support
분야별 특성 반영 정도	N/A	support

이타를 구축할 수 있는 메커니즘을 제공한다. 이러한 LOG 방법에서 제공하는 메커니즘은 방대한 데이터베이스를 대상으로 하여 메타데이터 구축할 수 있는 보다 현실적인 방법이다. 또한 계층적인 메타데이터 구조, 즉 계층형 MDR 구조를 제공함으로써 다양한 분야를 위한 통합된 메타데이터 정의가 가능하며, 각 분야별 특성을 다양하게 반영할 수 있다. 표 2는 기존 MDR 기반의 접근방법과 LOG 방법을 비교 평가한 내용이다.

위 표에서, LOG 모델이 구조의 복잡도 측면에서 기존의 MDR 기반 모델보다 비교적 높음을 알 수 있다. 이 논문에서 제안하는 LOG 모델은 계층적인 구조를 지니고 있으며 구조들간의 관계성을 유지해야 한다. 이에 비해, 기존 MDR 기반의 모델은 단순 메타데이터 레지스트리 구조를 지니고 있기 때문에 구조의 복잡도가 낮다. 그러나 LOG 모델은 기 구축된 데이터베이스를 통합할 수 있는 메커니즘의 제공과 강한 표현력을 통한 다양한 응용성 등 많은 장점들을 얻을 수 있다.

동적 데이터 요소 관리 메커니즘은 ISO/IEC 11179의 기본 기능이므로 두 모델 모두 이를 지원한다. 그러나 기존 방법론은 단순히 ISO/IEC 11179에서 제공하는 동적 메타데이터 관리를 기능들만을 제공한다. 즉 기 구축되어 있는 실제 데이터베이스 내에 있는 필드들과 데이터 요소와 사상 관계 정보를 지니지 않는다. 그러나 제안하는 모델은 메타데이터 요소들과 실제 데이터 필드들간의 관계 규칙을 관리하고 유지하기 때문에 표현력이 강하다. 이는 기 구축된 데이터베이스와의 사상 관계 정보 생성을 통해 온톨로지 기반 방법론과 같은 상향식 통합 메커니즘도 제공함을 의미한다.

분야별 특성 반영 측면에서, 기존의 MDR 기반 방법론은 단순한 구조를 지니고 있기 때문에 분야별 특성을 고려하여 분산 관리할 수 있는 구조를 제공하지 않는다. 그러나 LOG 모델의 경우, 계층적인 구조를 지니고 있기 때문에 전역 데이터 요소와 지역 데이터 요소를 분리하여 관리할 수 있다. 즉 계층적 구조를 이용하여 공통된 데이터 요소와 서로 독립적인 데이터 요소들을 분리하여 지역적으로 분산 관리할 수 있는 있다.

따라서 제안된 LOG 모델이 구조적 복잡도를 제외한 많은 비교항목에서 나온 구조를 제공한다. 비록 구조의 복잡도가 기존 MDR 기반 방법론에 비해 높지만 보다 많은 장점을 얻을 수 있다. 또한 이러한 추가적인 장점은 기존의 MDR 기반 모델과 비교하여 보다 실용적인 접근 방법이며, 현재 이슈화되고 있는 메타데이터 레지스트리들간의 상호운용성을 지원할 수 있는 구조를 제공한다.

6. 결론

이 논문에서는 점진적, 계층적으로 메타데이터 레지스트리를 구축할 수 있는 방법을 제안하였으며 제안한 방법론을 LOG(Localized-Global metadata registry)라고 정의하였다. 이 방법은 단위 시간당 제한된 비용과 통계 정보를 지니지 않는 방대한 데이터베이스를 대상으로 하여 메타데이터 레지스트리를 구축할 수 있는 메커니즘을 제공한다. 이를 위한 데이터 레벨과 사용자 레벨 및 데이터 이용들간의 관계성을 나타내는 데이터 가시성이라는 개념을 정의하고 LOG 방법론을 제안하였다. 이 방법론은 기존 MDR 기반의 접근 방법들의 문제점인 표현력의 문제, 기존 데이터베이스를 통합을 위한 메커니즘의 부재 등의 문제를 해결할 수 있는 보다 현실적인 방법론이다.

이러한 LOG 방법론의 실제화를 위하여 네 개의 계층(사용자, GMDR, LDMR 및 자원 계층)으로 구성된 시스템을 설계하고 구현하였다. 또한 제안한 방법과 이를 지원하는 시스템을 한국과학기술정보연구원의 실제 데이터베이스를 대상으로 하여 메타데이터 레지스트리를 구축하였다. 마지막으로, MDR 기반의 기존 접근 방법과 이 논문에서 제안하는 LOG 방법론을 비교 평가하였다.

결론적으로, LOG 방법론은 현실적으로 존재하는 단위 시간당 비용과 통계 정보를 지니지 않는 데이터베이스 우선순위 결정 문제를 극복하고 메타데이터 레지스트리를 구축할 수 있는 실용적인 메커니즘을 제공한다. 또한 기존 MDR 접근 방법에서 제공하지 않는 이미 구

축되어 있는 데이터베이스들을 수용하여 통합할 수 있는 구조 및 방법을 제공한다.

현재 정의되고 구축된 LOG 방법론 및 시스템은 보다 자동화된 일반화 알고리즘이나 기 구축된 데이터베이스의 자동 사상 등을 제공하지 않는다. 따라서 향후 이에 대한 연구와 더불어 제안된 방법론에 대한 실험을 통한 검증 작업이 요구된다. 이러한 문제가 해결될 경우, 제안하는 방법의 실제 적용에 있어 보다 많은 비용을 절감할 수 있다.

참 고 문 헌

[1] ISO/IEC JTC1/SC32, "ISO/IEC 11179: Specification and standardization of data elements - Part 1~6," ISO/IEC JTC 1, 2003.

[2] S. Ram, "Special issue on heterogeneous distributed database systems," IEEE Computer, Vol. 24, No.3, 1991.

[3] A. P. Sheth, "Semantic Issues in Multidatabase Systems," SIGMOD Record, Vol. 20, No.4, pp. 5-9, December 1991.

[4] C. A. Knoblock et al., "Modeling Web Sources for Information Integration," In Proceedings of 11th Nat'l Conference on Artificial Intelligence, 1998.

[5] W. Kim, I.-J. Choi, S. Gala and M. Scheevel, "On resolving Schematic Heterogeneity in Multidatabase Systems," Distributed and Parallel Databases, An International Journal, Vol. 1, No. 3, pp. 251-279, 1993.

[6] A. Sheth and J. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous and Autonomous Databases," ACM Computing Surveys, Vol. 22, No.3, pp. 183-236, 1990.

[7] M. Tamer Ozsu, Umeshwar Dayal and Patrick Valduriez, "Distributed Object Management," Morgan Kaufmann Publishers, San Mateo, California, 1994.

[8] R. G. G. Cattell, D. Barry, M. Berler, J. Eastman, D. Jordan, C. Russell, O. Schadow, T. Stanienda and F. Velez, "The Object Database Standard ODMG 3.0," Morgan Kaufmann, 2000.

[9] F. Manola, S. Heiler, D. Georgakopoulos, M. Hornick and M. Brodie, "Distributed Object Management," International Journal of Intelligent & Cooperative Information Systems, Vol. 1, No. 1, 1992.

[10] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. D. Ullman, V. Vassalos and J. Widom, "The TSIMMIS Approach to Mediation: Data Models and Languages," Journal of Intelligent Information Systems, Vol. 8, No. 2, pp. 117-132, 1997.

[11] G. Wiederhold, "Mediators in the architecture of future information systems," IEEE Computer, Vol. 25, No. 3, pp. 38-49, 1992.

[12] A. El Abbadi, M. L. Brodie, S. Chakravarthy, U.

Dayal, N. Kamel, G. Schlageter and K.-Y. Whang, "A CaseBased Approach to Information Integration," Proceedings of 26th International Conference on Very Large Data Bases(VLDB 2000), Cairo, Egypt, 2000.

[13] M. Uschold and M. Gruninger, "Ontologies: principles, method and applications," The Knowledge Engineering Review, Vol. 11, No. 2, pp. 93-136, 1996.

[14] B. Furrie, "Understanding MARC Bibliographic: Machine-Readable Cataloging," Follett Software Co., 1998.

[15] E. Miller, "An introduction to the Resource Description Framework," D-Lib Magazine, 1998.

[16] S. Sugimoto, T. Baker and S. Weibel, "Dublin Core: Process and Principles," International Conference on Asian Digital Libraries (ICADL 2002), pp. 25-35, 2002.

[17] M. H. Needleman, "ONIX (Online Information Exchange)," Serials Review, Vol. 27, No. 3, pp. 102-104, 2001.

[18] 한국데이터베이스진흥센터, 데이터베이스 백서, 한국데이터베이스진흥센터, 서울, 1997.



정 동 원
 1997년 군산대학교 전산과(학사). 1998년 한국전자통신연구원(위촉 연구원). 1999년 충북대학교 전산과(석사). 1999년 ICU 부설 한국정보통신교육원 전임강사 2000년 (주)지구넷 선임연구원. 2000년~현재 고려대학교 대학원 컴퓨터학과 박사과정. 2000년~현재 고려대학교 정보통신연구소 연구원 2001년~현재 (주)라임미디어 테크놀로지 연구원. 2002년~현재 TTA TC 08 SG 02 특별위원. 관심분야는 분산 컴퓨팅, 이동 에이전트, 데이터베이스, 메타데이터 레지스트리 기반 정보 통합, 정보 표준화 등



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



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, 정보검색, 지능형에이전트, 웹기반정보시스템 등



서 태 설

1980년~1983년 연세대학교 기계공학과(학사). 1984년~1985년 KAIST 생산공학과(석사). 1993년~현재 KAIST 서울분원 자동화 및 설계공학과(박사과정)
 1988년~2000년 산업기술정보원. 2000년~현재 한국과학기술정보연구원(실장)
 2001년~현재 응용데이터 연구반 위원(TTA). 2001년 현재 한국 ISO/IEC JTC1/SC32 전문위원(기술표준원). 2002년~현재 데이터기술위원회(TC08) 의장(TTA). 관심분야는 자동화 설계, 과학기술정보 표준화 등

백 두 권

정보과학회논문지 : 컴퓨팅의 실제
 제 9 권 제 2 호 참조