

# 메타데이터 레지스트리를 위한 SQL 기반 질의 언어 설계

신동길<sup>○</sup>, 정동원, 이정옥<sup>†</sup>, 백두권  
고려대학교, 건국대학교<sup>†</sup>

{dkshin<sup>○</sup>, withimp, baik}@software.korea.ac.kr, ljo@kku.ac.kr<sup>†</sup>

## A Design of SQL-based Query Language for Metadata Registry

Dongkil Shin<sup>○</sup>, Dongwon Jeong, JeongOog Lee<sup>†</sup>, Dookwon Baik  
Dept. of Computer Science and Engineering, Korea University  
Dept. of Computer Science, Konkuk University<sup>†</sup>

### 요 약

현재 세계 각국에서 많은 메타데이터 레지스트리들이 구축되어 왔다. 그러나 메타데이터 레지스트리가 국제 표준(ISO/IEC 11179)임에도 불구하고 일관성 있는 표준 접근 인터페이스를 제공하지 않음으로써 각각 다른 연산으로 구현되었다. 이는 각각의 메타데이터 레지스트리 구축시 동일한 연산 패턴의 중복된 구현으로 인한 불필요한 비용을 야기한다. 국제 표준인 메타데이터 레지스트리에 대한 접근시 시스템마다 동일한 연산 패턴을 지니게 된다. 이 논문에서는 이와 같이 메타데이터 레지스트리에서 공통적으로 사용되는 연산 패턴을 분석하고 정의한다. 또한 분석된 연산 패턴을 이용하여 표준 SQL을 확장한 SQL/MDR을 정의하고 설계한다. 설계된 SQL/MDR은 메타데이터 레지스트리 관리 시스템 개발에 표준 접근 방법을 제공한다. 또한 이는 개별적인 개발로 인한 추가적인 노력을 감소시키고 메타데이터 레지스트리에 대한 보다 일관성 있는 접근을 가능하게 한다.

### 1. 서 론

최근에 메타데이터 레지스트리를 기반으로 한 많은 시스템들이 구축 및 개발되어 왔다. 메타데이터 레지스트리는 메타데이터의 동적 관리 및 상호운용성 증진을 위해 ISO/IEC JTC 1에 의해 개발된 국제 표준이다[1]. ISO/IEC JTC 1은 ISO와 IEC가 메타데이터 연구를 위해 정보 기술분야 공동 기술 위원회를 구성한 국제 표준 개발 기구이다. 현재까지 국내·외 기관들이 메타데이터 관리를 위해 ISO/IEC 11179를 기반으로 하여 메타데이터 레지스트리를 구축해 왔다[2],[3],[4],[5].

메타데이터 레지스트리(MDR, Metadata Registry)의 구축 및 관리를 위해서는 등록, 데이터 요소 검색 및 등록 상태 변경 등과 같은 다양한 형태의 접근이 필요하다. 이 때 표준에 따라 구축된 메타데이터 레지스트리들은 동일한 형태의 일정한 접근 패턴, 즉 일정한 연산을 지니게 된다. 다시말해, 데이터베이스를 접근함에 있어 일정한 질의 패턴을 지니게 된다. 그러나 현재까지 이러한 동일 연산들에 대한 표준 접근 방법이 연구되지 않았다. 또한 표준 접근 방법을 제공하지 않기 때문에 메타데이터 레지스트리 접근시 반드시 요구되는 제약조건이 고려되지 않는 경우가 발생한다. 이는 예외적인 상황을 발생시키고 표준화된 메타데이터 레지스트리 구축을 어렵게 한다. 추가적으로, 동일한 패턴 연산을 반복적으로 개발함으로써 시간과 비용을 낭비하게 된다.

따라서 반복적인 메타데이터 레지스트리 접근 연산 개발에 따른 시간 및 비용을 줄이고 ISO/IEC 11179 표준명세에 따른 메타데이터 레지스트리 구축 및 일관성 있는 접근 방법에 대한 연구가 필요하다.

이 논문에서는 이를 위해 먼저 동일한 패턴을 지니는 접근 연산들을 분석한다. 또한 동일한 연산들과 제약조건을 수용하고 일관성 있는 접근을 위한 메타데이터 레지스트리 질의 언어를 정의 및 설계한다. 질의 언어는 국제 표준 질의어인 SQL을 기반으로 하였으며, 이 논문에서는 설계된 SQL 기반의 메타데이터 레지스트리 질의 언어를 SQL/MDR이라고 정의한다.

### 2. 관련연구

ISO/IEC 11179는 데이터를 쉽게 이해하고 공유할 수 있도록 하기 위해 국제 표준 기구인 ISO/IEC JTC 1에 의해 개발되었다. 이는 데이터베이스들간의 공유 및 교환을 극대화하고 표준방식으로 일관성 있게 데이터베이스를 표현하고 구축하기 위한 국제 표준이다. ISO/IEC 11179는 데이터 요소라는 중요한 개념을 지니고 있으며, 이는 정의, 식별, 표현 및 허용 가능한 값들을 속성 집합으로서 데이터를 명세하기 위한 최소단위이다. 이러한 데이터 요소들의 집합이 메타데이터 레지스트리이다.

현재 국내에서 구축된 메타데이터 레지스트리는 한국전자통신연구원에서 컴포넌트의 원활한 유통을 위해 구축한 컴포넌트 메타데이터 레지스트리[3], 한국과학기술정보 연구원에서 개발한 서지정보 메타데이터 레지스트리[2]가 있다. 해외에서 구축된 메타데이터 레지스트리는 미국 환경보호청의 EDR(Environmental Data Registry)[4], 호주의 NHK(Australian National Health Information Knowledgebase)[5], 미국의 ITS(Intelligent Transportation System)[6] 등이 있다. EDR은 환경 데이터에 대한 참조 정보로서 정의, 자원, 그리고 환경의 데이터의 사용에 대한 포괄적이고 권위 있는 참조 정보를 가지고 있다. NHK는 건강과 관련된 메타데이터에 대한 전자 저장소(Electronic Repository)이며 질의 도구이다. ITS는 미 교통부에서 50개 주와 지역의 교통국이 ITS 시스템 내부에서 혹은 중앙정부와 ITS 외부 시스템이 가지고 있는 데이터를 의미에 따라 교환하고 상호작용 할 수 있도록 하는 메타데이터 레지스트리로 구축된 시스템이다.

이들은 모두 국제 표준인 ISO/IEC 11179를 기반으로 구축되었음에도 불구하고 동일한 접근 연산을 위해 서로 상이한 접근 방법을 개발하여 이용하고 있다. 즉, 일관성과 표준화가 결여된 메타데이터 레지스트리 접근 방법을 사용한다. 이는 불필요한 노력과 비용 낭비를 초래하게 되며, 또한 서로 다른 접근 방법을 사용함으로써 국제 표준인 메타데이터 레지스트리의 일관성 결여를 야기한다.

따라서 메타데이터 레지스트리에 대한 표준 인터페이스를 제공하기 위한 일관성 있는 접근 방법에 대한 연구가 필요하다. 이미 데이터베이스 분야의 경우, 공간 데이터, 시간 데이터 및 멀티미디어 데이터 등에 대한 일관성 있는 접근을 위한 표준

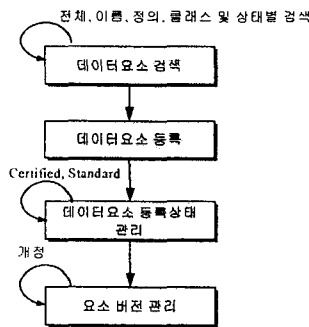
인터페이스 개발이 진행되어 왔다. 공간 데이터의 표준 접근을 위한 공간 질의어(Spatial Query Language) 연구[7], 시간 데이터를 위한 시간 질의어(Temporal Query Language) 연구[8], 멀티미디어 데이터를 위한 SQL/MM[9] 등이 그 예이다.

이 논문은 국제 표준인 메타데이터 레지스트리에 대한 일관성 있는 접근을 위한 인터페이스 정의 및 설계를 목적으로 한다. 먼저 이러한 목적을 위해 메타데이터 레지스트리 접근시 필요한 연산 패턴을 조사하고 분석한다. 분석된 연산 패턴을 이용하여 표준적인 메타데이터 레지스트리 질의 언어를 정의 및 설계한다. 이를 위해 국제 표준 질의어인 SQL을 기반으로 하여 메타데이터 레지스트리 질의 언어인 SQL/MDR을 설계한다.

SQL/MDR은 메타데이터 레지스트리에 대한 표준 인터페이스를 제공함으로써 반복적인 메타데이터 레지스트리 접근 연산 개발에 따른 시간 및 비용을 줄이고 ISO/IEC 11179 표준명세에 따른 메타데이터 레지스트리 구축 및 일관성 있는 접근 방법을 제공한다.

### 3. 메타데이터 레지스트리 접근 패턴

#### 3.1 메타데이터 레지스트리 기본 연산



ISO/IEC 11179에 명시된 메타데이터 레지스트리의 관리를 위한 기본적인 기능은 데이터요소 검색, 데이터요소 등록, 데이터요소 등록상태 관리 및 데이터요소 버전 관리 등이 있다. 이 기능을 통해 메타데이터 레지스트리를 확장하고 관리한다.

데이터요소 검색은 메타데이터 레지스트리에서 사용자가 등록된 데이터요소에 대한 명세를 찾아 볼 수 있도록 다양한 옵션의 검색기능을 제공한다. 사용자는 데이터요소 검색을 통해 자신이 원하는 데이터요소를 실제 데이터베이스 구축시 사용할 수 있으며 데이터요소에 대한 메타정보를 정확히 확인할 수 있다.

데이터요소 등록은 사용자가 임의의 메타데이터 레지스트리 내에 필요로 하는 데이터요소가 없을 경우, registrar를 통해 새로운 데이터요소를 등록하는 기능이다. 메타데이터 레지스트리는 데이터요소 등록을 통해 지속적으로 확장되며 등록된 데이터요소는 Control Committee의 인증과정을 거쳐 표준 데이터요소로 채택된다.

데이터요소 등록 상태관리는 임의의 데이터요소 상태를 변경하는 연산으로서, 일반적으로 데이터요소 'recorded', 'certified' 및 'standard' 등의 상태를 지닌다.

데이터요소 버전 관리는 등록상태의 변경이나 데이터요소 개정에 따른 버전을 관리하는 기능이다.

이런 기능들은 모든 메타데이터 레지스트리가 기본적으로 준수해야 한다. 즉 모든 메타데이터 레지스트리 관리 및 접근을 위한 연산들의 일정한 패턴을 갖는다. 이 논문에서는 이러한 유사한 연산패턴을 기반으로 일관성 있는 접근방법을 설계하고 제안한다.

#### 3.2 연산 패턴 분석

데이터요소 검색에 사용되는 연산패턴은 메타데이터 레지스트리에 접근해 SELECT 연산시 옵션으로 검색하는 패턴이 있다. 이는 메타데이터 레지스트리의 가장 기본적인 기능으로 모든 메타데이터 레지스트리는 이와 같은 데이터요소 검색패턴을 갖는다.

데이터요소 등록시 데이터요소 이름은 반드시 다른 이름을 갖는 데이터요소 이름을 입력해야 한다. 그렇기 때문에 메타데이터 레지스트리는 중복여부를 체크하는 패턴연산을 갖는다. 메타데이터 레지스트리 개발시 미리 정의된 SQL/MDR을 이용하면 시간을 절약할 수 있으며 개발시 중복여부를 자동으로 체크할 수 있다.

데이터요소 상태 관리는 데이터요소가 Control Committee의 투표를 통해 등록에서 인증 및 표준으로 변경시 해당 요소의 상태를 수정한다. 데이터요소의 상태는 Recorded, Certified, Standard로 변경되며 변경되는 상태에 따라 상태정보와 관련된 메타정보가 함께 자동으로 수정된다. 이 연산패턴을 이용하여 데이터요소의 상태 단계별로 수정되는 항목을 자동으로 처리할 수 있다. 이처럼 데이터요소 투표후에 데이터요소의 특정 메타정보 내용을 수정하는 패턴연산이 발생한다.

여러 연산 패턴중 이 논문에서는 검색패턴 연산을 소개하며 검색관련 연산은 다음과 같다.

검색 관련 연산	
등록된 데이터요소 검색연산	
인증된 데이터요소 검색연산	
표준화된 데이터요소 검색연산	
제안된 데이터요소 검색연산	

### 4. SQL/MDR

#### 4.1 검색을 위한 연산자

메타데이터 레지스트리 관리 및 접근을 위한 연산들은 일정한 패턴을 지니며 이미 3장에서 분석된 연산패턴들에 대하여 기술하였다. 이런 연산패턴들을 기반으로 하여 일관성 있는 메타데이터 레지스트리 접근방법인 SQL/MDR을 설계하기 위해서는 각 연산 패턴들에 대한 메타데이터 레지스트리 연산자를 정의해야 한다. 다음은 데이터요소 검색시 사용되는 데이터요소에 대한 메타데이터 레지스트리 연산자들을 정의한 것이다.

구분	설명
STATUS RECORDED	상태가 'Recorded'인 요소검색
STATUS CERTIFIED	상태가 'Certified'인 요소검색
SATUS STANDARD	상태가 'Standard'인 요소검색
SATUS SUBMIT	상태가 'Submit'인 요소검색

#### 4.2 SQL/MDR 구문

앞서 분석한 메타데이터 레지스트리에서 사용되는 연산패턴 중 검색유형에 대해 표준 SQL을 확장하여 BNF(Backus Naur Form)로 정의하였다. 검색패턴은 메타데이터 레지스트리에서 데이터요소를 검색하는 구문으로 표준 SQL문의 WHERE 절에 <mdr search predicate>를 추가 정의하였다. 다음은 검색패턴의 BNF이다.

```

<query specification> ::=
SELECT [DISTINCT | ALL] <select list> <table expression>
  
```

```

<select list> ::=
ASTERISK | <MDR selection><select sublist> [ { COMMA <select
sublist> }... ]

<MDR selection> ::= <MDR search predicate>

<table expression> ::=
<from clause> <where clause> <group by clause> <having
clause>

<where clause> ::= WHERE <search condition>

<search condition> ::=
<boolean term> | <search condition> OR <boolean term>

<boolean term> ::=
<boolean factor> | <boolean term> AND <boolean factor>

<boolean factor> ::=
[NOT] <boolean test>

<boolean test> ::=
<boolean primary> [IS [ NOT ] <truth value>]

<boolean primary> ::=
<predicate> | <left paren> <search condition> <right paren>

<predicate> ::=
<comparison predicate> | <MDR search predicate>

<MDR search predicate> ::=
<field name> <MDR status condition expression>

<field name> ::= <value expression>

<MDR status condition expression> ::= STATUS_RECORDED |
STATUS_CERTIFIED | STATUS_STANDARD | STATUS_SUBMIT
    
```

위의 BNF는 WHERE절의 <predicate>에 <MDR search predicate>를 추가 하였다. <MDR search predicate>는 <field name>과 <MDR status condition expression>으로 구성되며 <field name>은 검색할 스키마의 컬럼 이름을 나타낸다. <MDR status condition expression>은 STATUS\_RECORDED, STATUS\_CERTIFIED, STATUS\_STANDARD 중 하나를 선택할 수 있도록 함으로써 메타데이터 레지스트리에서 정의한 데이터 요소의 세가지 상태 즉, recorded, certified, standard를 검색할 수 있게 한다. 이와 같이 선택항목을 SQL을 확장하여 미리 정의함으로써 SQL문 질의시 더욱 명시적으로 표현할 수 있으며 일관성 있는 메타데이터 레지스트리를 개발할 수 있다.

#### 4.3 검색을 위한 SQL/MDR

다음은 메타데이터 레지스트리에서 가장 기본적인 기능인 데이터 요소 검색의 예이다.

<질의 1> 등록상태가 RECORDED인 요소를 검색하시오.

```

[SQL]
SELECT *
FROM table_name
WHERE status='recorded'
    
```

```

[SQL / MDR]
SELECT STATUS_RECORDED
FROM table_name
    
```

질의 1은 Control Committee가 데이터요소 상태가 Recorded인 요소 투표를 통해 데이터요소를 인증할 때와 데이터요소를 등록한 Submitter가 방금 등록한 데이터요소를 검색할 때 자주 사용되는 질의문이다. SELECT문에 직접 STATUS\_RECORDED을 이용해 질의 함으로써 질의문을 단축시키고 메타데이터 레지스트리에서 자주 사용되는 연산패턴을 표준화 하여 검색하였다. 또한 WHERE절 이후에 명시적으로 STATUS\_RECORDED를 이용해 질의할 수 있다.

<질의 2> 등록상태가 RECORDED이고 요소이름이 KOREA인 요소를 검색하시오.

```

[SQL]
SELECT *
FROM table_name
WHERE status='recorded'
AND de_name='korea'
    
```

```

[SQL / MDR]
SELECT *
FROM table_name
WHERE status STATUS_RECORDED de_name 'korea'
    
```

질의 2는 데이터요소의 등록상태가 recorded이면서 데이터요소 이름이 korea인 요소를 검색하는 질의문이다. Control Committee와 Submitter가 등록 상태가 recorded인 데이터요소 이름으로 직접검색 할 때 생성되는 질의 문이다. <질의 1>과 <질의 2>의 예와 같이 메타데이터 레지스트리에서 자주 사용되는 연산패턴을 SQL문으로 확장함으로써 표준화된 메타데이터 레지스트리를 개발 할 수 있으며 표준 프로세스를 따르도록 유도한다.

#### 5. 결론

이 논문은 국내/외에서 개발된 메타데이터 레지스트리의 연산패턴을 분석하고 분석된 연산패턴중 일부를 SQL문으로 확장 정의하였다. 확장된 SQL/MDR은 메타데이터 레지스트리의 가장 기본이 되는 연산으로 구성되어 있어 이를 이용하여 메타데이터 레지스트리의 개발 시간과 비용을 줄이고 ISO/IEC 11179의 명세를 준수하여 일관성 있는 시스템을 개발 할 수 있다. 또한 국내외의 관심에 비해 그 사례가 없는 메타데이터 레지스트리에 적용되는 표준 SQL을 확장하였다.

향후 메타데이터 레지스트리의 연산패턴을 좀더 세밀히 분석하고 정의하여 반드시 필요한 연산을 갖는 SQL/MDR의 확장이 요구된다. 확장된 SQL/MDR을 실제 메타데이터 레지스트리에 적용시켜 개발 비용단축과 표준 명세를 따라 개발하는데 어떤 이득이 있는지에 대한 검증이 필요하다.

#### 참고문헌

1. ISO/IEC JTC1 SC32, ISO/IEC 11179: Specification and standardization of data elements, Part 1~6, 2003
2. 한국과학기술 정보연구원, 과학기술 지식정보 공유 및 표준화 관리 모델 개발연구, KISTI 연구보고서, 2002.
3. 한국전자통신연구원, 컴포넌트등록 및 검색시스템 연구, ETRI 연구보고서, 2000.
4. <http://www.epa.gov/edr/>
5. <http://www.aihw.gov.au/>
6. <http://www.dot.gov>
7. Egenhofer, M., Spatial SQL: A query and presentation language, IEEE Transactions on Knowledge and Data Engineering, 1994.
8. Pissinou, N., Snodgrass, R., et al., Towards an infrastructure for temporal databases: Report of an invitational. arpa/nsf workshop. In Sigmod record, 1994
9. ISO/IEC JTC1 SC32, ISO/IEC 13249: Information technology - Database languages - SQL Multimedia and Application Packages, Part 1~5, 2003