

# XMDR을 이용한 멀티데이터베이스 통합 검색 시스템

허욱<sup>o</sup> 문석재 엄영현 정계동 최영근

광운대학교 컴퓨터 과학과

{her0116, msj1568, class76, gdjung, ygchoi}@kw.ac.kr

## Multi-Database Integration Retrieval System using XMDR

Uk Heo<sup>o</sup>, Seok-Jae Moon, Young-Hyun Eum, Kye-Dong Jung, Young-Keun Choi

Department of Computer Science, KwangWoon University

### 요 약

한 도메인 내의 통합검색시스템에서 메타데이터를 이용하여 분산되어있는 데이터를 공유하고 표준화하여 정보를 통합검색 할 수 있다. 하지만 메타데이터를 특정 도메인에 맞게 설계하여 메타데이터 레거지스 토리로 구축되어진 조직 간에 데이터요소 의미, 구조, 형식상의 이질적인 문제들이 발생하여 통합 검색하기 어렵다. 따라서 본 논문에서는 XMDR(eXtended Metadata Registry)을 이용한 멀티데이터베이스 통합 검색 시스템을 제안한다. 제안된 시스템은 레거지 데이터 시스템들 간의 협업에 필요한 데이터를 공유 및 교환하는데 실시간으로 변화하는 데이터를 일관성 있게 유지할 수 있으며 통합 검색시 단일 인터페이스를 제공하여 각 시스템의 독립성을 유지하면서 데이터의 투명성과 가용성을 향상 시킬 수 있다.

### 1. 서 론

개별적으로 운영되던 데이터베이스들은 인터넷의 발달과 정보화 사회의 흐름에 따라 데이터베이스 시스템의 종류가 다양해지고 다른 데이터베이스 시스템으로부터 데이터를 액세스해야 할 필요성이 생겼다. 공유되는 정보는 효율적으로 처리 및 관리가 되어야 한다. 하지만 도메인 내에서 각 레거지 시스템들 간의 정보가 분산되어 있거나, 동일한 정보가 서로 다르게 표현되는 문제가 발생함에도 불구하고 데이터 통합이 필요하게 되었다. 따라서 개별적으로 운영되던 데이터베이스는 상호연관성을 가지게 됐다. 그러나 개별 데이터베이스를 통합시 플랫폼 이질성, 의미 이질성 등의 문제가 발생했다.[2][3] 이러한 문제점을 해결하기 위해 여러 통합방법이 제안되었다. 특히, 이기종 간의 시스템이나 다양한 애플리케이션에서 나오는 대량의 데이터들은 종류와 형식에 구분 없이 호환가능하고, 정확한 정보를 실시간으로 동기화하여 제공하게 됐다. 예를 들어, 제조업체는 효율적인 재고관리와 정확한 비용 산정, 수요 예측 등을 전개할 수 있게 되었다. 하지만 대량의 정보를 모으고 처리하는 이 기종 응용 시스템들 간의 연계과정에서 정확성이 검증된 정보를 공유하는 데에는 많은 어려움이 있다.

본 논문에서는 XMDR을 이용하여 개별 데이터베이스의 스키마 정보 및 접속 정보를 보다 쉽게 구성하고 통합하여, 개별 데이터베이스의 데이터요소의 의미, 구조, 형

식상 이질성을 해결하는 방안을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 XMDR에 대하여 기술한다. 3장에서는 시스템의 구조에 대하여 기술한다. 4장에서는 시스템 흐름도에 대하여 설명한다. 5장에서는 시스템 적용 예를 기술한다. 마지막으로, 6장에서는 결론 및 향후연구를 기술한다.

### 2. 관련연구

#### 2.1 XMDR(eXtended Metadata Registry)

본 논문에서의 XMDR은 ISO/IEC 11179[1]에서 제안한 정보공유 교환을 위한 표준인 MDR과 정보의 효율적인 이용을 위하여 온톨로지를 포함하여 확장한 개념으로서, 현재 많은 연구가 진행 중에 있다.

특히, ISO/IEC 11179-3에서는 공유 데이터의 관리를 위한 메타모델, 기본속성이 제시되어 있다. 메타모델은 의미적인 내용과 분산된 환경하의 사용자들이나 정보처리 시스템간의 공유되는 데이터 요소의 구분을 위한 표준과 안내를 제공하고 있다.

본 논문에서 제안한 시스템에 적용된 XMDR은 MSO(Meta-Semantic Ontology), ML(Meta-Location), InSO(Instance-Semantic Ontology)로 구성하였다.[5] 각 요소는 다음과 같이 정의되고 구성된다.

■MSO(Meta-Semantic Ontology): 개별 데이터베이스의 스키마 정보들을 시소러스화하여 표준에 맞추어 매핑한

것으로, 메타데이터의 관계성과 이질성 해결을 목적으로 한다.

■ ML(Meta-Location) : MSO와 연계하여 개별 데이터베이스들의 물리적인 정보인 위치정보, 접근권한 등을 관리한다.

■ InSO(Instance-Semantic Ontology) : 실제 표준 데이터 값과 데이터 값들 간의 연관관계성을 시소러스화하여 정의한 것이다.

위와 같이 정의된 요소들은 ISO/IEC 11179-3에서 제안한 데이터의 속성 명세를 따른다. 데이터의 기본속성은 식별속성, 정의속성, 관계속성, 표현속성으로 다음과 같다.

- 식별속성(Identify Attribute) : 데이터 요소의 식별을 위한 속성
- 정의속성(Define Attribute) : 데이터 요소의 의미를 갖는 속성
- 표현속성(Presentation Attribute) : 데이터 요소의 표현방식에 따른 속성

## 2.2 메타-시맨틱 온톨로지(Meta-Semantics Ontology)

MSO(Meta-Semantics Ontology)는 메타데이터 상호운용에서 발생할 수 있는 동등한 의미를 갖는 메타데이터 간에 일어날 수 있는 문제를 분류하고 그 유형을 정리하였다. 동등의미 메타데이터 간에 발생할 수 있는 구조, 의미, 표현의 불일치 유형으로 나눌 수 있다.

첫째, 구조의 이질성 유형은 다음과 같다.

■ 합성(Composition): 다수의 스키마 요소가 하나의 스키마 요소에 합으로 표현.

(ex. <FirstName U LastName> -> Name)

■ 분해(DeComposition): 하나의 스키마요소가 다수의 스키마 요소로 분리되어 표현.

(ex. Name -> <FirstName U LastName>)

■ 재배치(Rearramngement): 스키마 요소의 순서 정보가 필요로 하며, 그 배치가 다른 표현.

(ex. FirstName- LastName -> LastName-FirstName)

둘째, 의미 이질성 유형은 다음과 같다.

■ 대체(Subsitution): 스키마 요소간의 동등한 의미의 스키마 요소로 표현.

(ex. Name -> author, writer, fullname)

셋째, 표현 이질성 유형은 다음과 같다.

■ 코드형식(Code Format): 코드형식의 다른 표현.

(ex. YYYY-MM-DD <-> MM-DD-YYYY)

■ 단위형식(Unit Format): 단위형식의 다른 표현.

(ex. Pound <-> Kilogram, Mile <-> Kilometer)

■ 데이터형식(Data Format): 데이터 형식의 불일치 표현.  
(ex. VarChar(20) <-> VarChar(50), Integer <-> Float)

위와 같이 분류한 불일치 유형들은 단독으로 존재할 수도 있지만 일반적으로 여러 유형들이 혼합되어 나타난다.

## 2.3 메타-로케이션(Meta-Location)

MLOC(Meta-Location)는 개별 데이터베이스의 위치정보, 접근 권한 정보를 MSO와 릴레이션하여 개별 데이터베이스들로부터 검색이나 트랜잭션을 하기 위한 접속할 때 필요한 정보를 제공하는 역할을 한다.

## 2.4 인스턴스-시맨틱 온톨로지(Instance-Semantics Ontology)

InSO(Instance-Semantics Ontology)는 협업 환경에서 데이터 값이 상호운용할 때 실제 값(value)들 사이에 연관성을 구조화하여 구체적으로 관계성을 정의한 것이다. 단순한 검색이 아니라 검색 키워드에 대한 의미성, 유사성을 고려하여 검색에 대한 결과를 확장하려는 데 목적이 있다. 값들 간의 연관관계는 동등관계, 포함관계, 부분관계, 참조관계의 4가지로 분류할 수 있으며 다음과 같다.

■ 동등관계: 데이터를 표현하는 방식은 다르지만 의미상 동등한 의미를 가지는 관계.

(ex. Computer = Com = 컴퓨터)

■ 포함관계: 데이터를 의미하는 것에 포함하는 관계.

(ex. 본체 -> Memory, CPU)

■ 부분관계: 데이터간의 의미상위, 하위개념관계.

(ex. 전자제품-컴퓨터-서버)

■ 참조관계: 데이터간 의미상 참조가 되는 관계.

(ex. PC -> Moniter, Keyboard)

예를 들어, 'Computer'에 대해 요청한다면 'COM', '컴퓨터' 등과 같은 의미는 갖지만 다른 표현으로 된 데이터도 같이 검색되어 Computer에 대한 정보를 더욱 확대하여 검색할 수 있도록 한다.

## 2.5 XMDR 스키마 설계

XMDR 설계는 2.2, 2.3, 2.4에서 기술한 방법으로 이질적인 문제를 해결하고 데이터 상호운용의 일관성을 위해 표준 스키마 항목을 결정한다.

결정된 표준스키마 항목과 개별 데이터베이스 스키마 정보를 표준 스키마 항목에 매핑시키고 이질적인 문제를 해결하기 위해 요소 사이의 변환, 데이터 타입간의 변환정보를 생성한다. 그림은 메타-시맨틱 온톨로지,

메타로케이션, 인스턴스-시멘틱 온톨로지를 릴레이션을 표현한 스키마 모델이다.

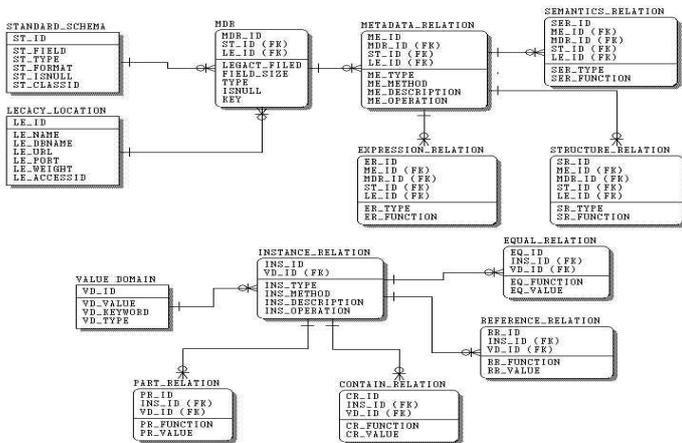


그림 1. XMDR의 스키마 관계도

### 3. 시스템 구조

#### 3.1 시스템 개요

본 논문에서 제안한 XMDR을 이용한 멀티데이터베이스 통합 검색 시스템의 구조는 (그림 2)과 같다.

핵심요소는 크게 XMDR(eXtended Metadata Registry), VM(Value Mapper), MMA(Metadata Mapping Agent)가 있다.



그림 2. 시스템 구조

#### 3.2. Domain Layer

Domain Layer는 사용자 인터페이스인 UIA와 사용자 질의를 생성하는 QMA로 구성한다.

#### 3.2.1 User Interface Agent (UIA)

사용자가 사용할 수 있는 표준 인터페이스를 제공하고 XMDR을 데이터 요소를 추출하여 질의를 요청하고 질의 결과를 보여주는 역할을 한다.

#### 3.2.2 Query Managing Agent (QMA)

UIA(User Interface Agent)에 요청한 질의를 이용하여 사용자 질의를 생성, 삭제 등의 관리를 하며, 사용자 질의를 XMDR Layer에 전달한다.

### 3.3 XMDR Layer

XMDR에서 MSO, ML, ISO의 정보를 이용하여 데이터 요소 사상을 관리담당하며, XMDR Manager과 VMA로 구성된 Value Mapper를 구성한다.

#### 3.3.1 Value Mapping Agent (VMA)

Value Mapping Agent는 Metadata Mapping Agent에 보내어질 글로벌 쿼리를 XMDR Manager를 참조하여 생성한다.

#### 3.3.2 XMDR Manager

개별 데이터베이스간의 접속정보 및 데이터요소 의미 이질성의 해결을 담당한다.

### 3.4 Database Layer

실제로 연결되는 데이터베이스들을 통합 관리하는 역할을 담당한다.

#### 3.4.1 Metadata Mapping Agent (MMA)

개별 데이터베이스의 스키마 정보의 생성 및 변경 정보를 Value Mapping Agent에 제공하는 역할을 한다. 또한 VMA에서 전달된 글로벌 쿼리를 분석하여 로컬쿼리로 만든 다음 데이터베이스에 요청하게 된다.

#### 3.4.2 Database

실제 연동되는 개별 데이터베이스를 말한다.

### 4. 시스템 흐름도

멀티 데이터베이스에서 통합 검색하기 위해 데이터 공유 및 교환할 때 수시로 데이터는 변화한다. 이런 변화하는 데이터를 일관성 있게 유지하기 위한 방법으로 SQL/XMDR 메시지 교환을 이용하여 통합 검색을 한다.[5] SQL/XMDR 메시지 교환 및 사상 메커니즘을 본 논문의 시스템에 적용하였다. 이 시스템의 흐름도를 크게 클라이

언트와 서버로 나누어 설명한다.

#### 4.1 클라이언트 흐름도

클라이언트에서 데이터의 흐름도는 (그림3)과 같다.

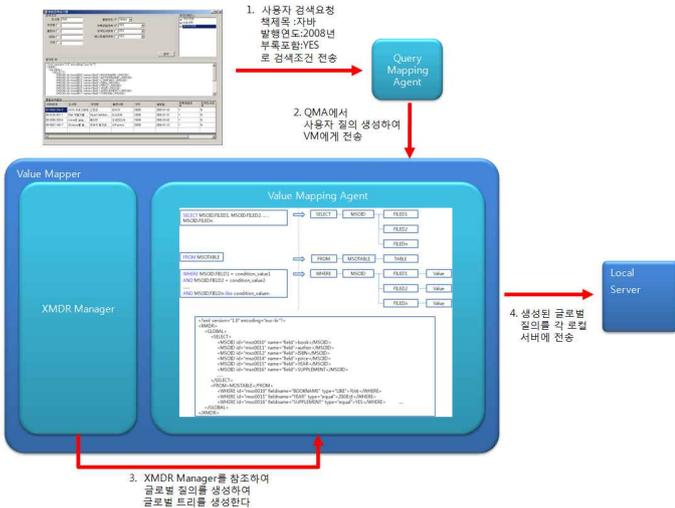


그림 3. 클라이언트의 데이터 흐름도

먼저 사용자가 조건을 선택하여 검색을 하면 그에 대한 검색 조건이 QMA(Query Mapping Agent)에 전송되면 QMA는 사용자 질의를 생성하여 XMDR Layer에 있는 VM(Value Mapper)에게 전송된다. VM은 사용자 질의를 받아 XMDR Manager를 참조하여 VMA(Value Mapping Agent)를 이용하여 글로벌 질의를 만든다. 글로벌 질의는 트리 구조로 변환하여 글로벌 트리를 생성한다. 글로벌 트리가 저장된 XML문서를 데이터베이스가 있는 각 서버에 전달하게 된다.

사용자가 책명이 “자바” 이면서 부록을 포함하고, 발행이 “2008년” 에 된 책을 찾고 있다고 한다. 이 조건이 QMA(Query Mapping Agent)에 전달되면 그에 맞는 사용자 질의가 생성된다. 생성된 사용자 질의는 XMDR Layer에 있는 XMDR Manager를 참조하여 VMA(Value Mapping Agent)를 이용하여 다음과 같은 글로벌 질의를 만든다.

■ 생성된 글로벌 질의 :

```
SELECT mso0010.book, mso0011.author, mso0012.ISBN,
mso0014.price, mso0015.YEAR, mso0016.supplement
FROM MSOTABLE
WHERE mso0010.book LIKE '%자바%'
and mso0015.YEAR = '2008'
and mso0016.supplement='YES'
```

위와 같이 생성된 질의는 트리로 표현이 가능하다. 위에서 만들어진 글로벌 질의는 글로벌 트리를 생성하여

XML문서로 저장한 다음 데이터베이스가 있는 각 Local Server로 XML문서를 전달한다.

#### 4.2 서버 흐름도

각 서버에서의 데이터 흐름도는 (그림4)와 같다.

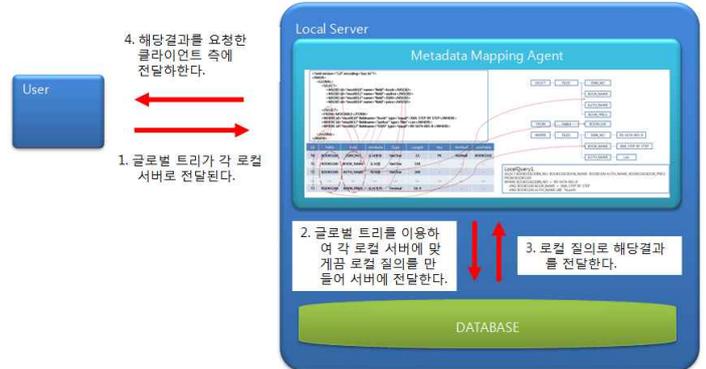


그림 4. 서버의 데이터 흐름도

(그림4)에서 나온 것처럼 글로벌트리가 만들어지면 데이터베이스가 있는 각 서버로 글로벌 트리가 저장된 XML문서가 전달되게 된다. 그럼 MMA(Metadata Mapping Agent)는 (그림4)에서와 같이 글로벌 트리를 각 로컬 스키마에 맞게끔 분석하여 아래와 같이 각 로컬 질의를 만들게 된다.

■ 생성된 로컬 질의 :

```
SELECT BOOKS100.BOOK_NAME,
BOOKS100.AUTH_NAME,
BOOKS100.ISBN_NO,
BOOKS100.BOOK_PRICE
BOOKS100.BOOK_YEAR
BOOKS100.BOOK_SUPPLE
FROM BOOKS100
WHERE BOOKS100.BOOK_NAME LIKE '%자바%'
AND BOOKS100.BOOK_YEAR='2008'
AND BOOKS100.BOOK_SUPPLE='YES'
```

위와 같이 만들어진 로컬 질의를 MMA는 데이터베이스에 전달하여 생성된 로컬 질의에 맞는 결과물을 내놓게 된다. 각 데이터베이스에서 나온 결과물은 MMA로 보내어지고 다시 MMA는 (그림4)에 있는 VMA에 그 결과물을 보내어 XMDR Manager를 참조하여 결과물을 통합하게 된다. 통합된 결과물은 QMA를 거쳐 (그림5) 검색결과화면에 그 통합된 결과를 보여주게 된다.

### 5. 시스템 적용 및 비교 분석

#### 5.1 시스템 적용

본 논문의 시스템 적용사례의 시스템 환경은 다음과

같다.

- 운영체제 : Windows Server 2003
- 데이터베이스 : MS-SQL Server 2005, Oracle8i
- 개발도구 : Visual Studio .Net

본 논문에서 제안한 XMDR을 이용한 통합 검색시스템은 여러 데이터베이스를 XMDR을 이용하여 정보를 통합하여 검색할 수 있는 서비스를 제공하는 도서 통합검색 시스템을 구현하였다. (그림5)은 본 논문에서 제안한 시스템을 이용하여 사용자가 통합 검색하는 화면이다. 사용자는 검색을 원하는 서버를 선택하여 검색할 수 있으며, 사용자가 원하는 조건을 선택하여 검색할 수 있게 하였다. (그림5)에서 사용자가 입력한 조회조건 파라미터 값을 이용하여 XMDR을 참조하여 글로벌 질의를 생성하여, 생성된 글로벌 질의 트리 구성이 질의문 뷰어에 보여진다.

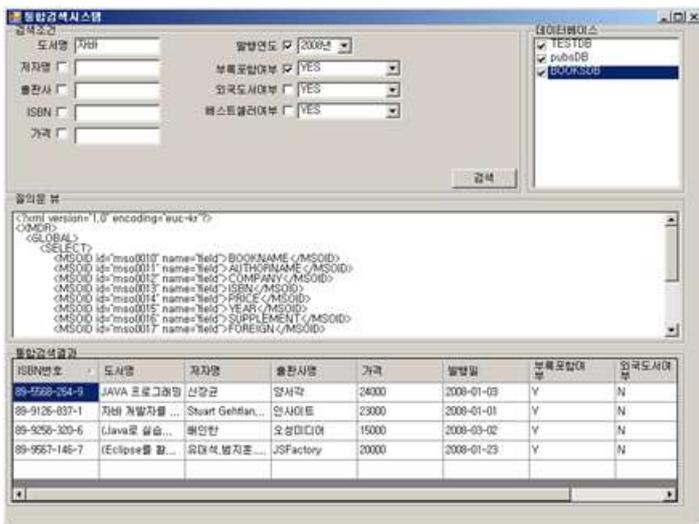


그림 5. 통합검색화면 및 검색결과화면

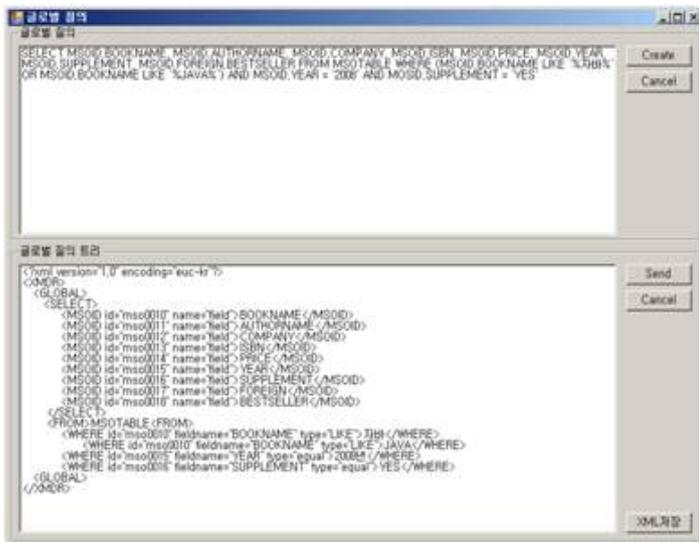


그림 6. 글로벌 질의 및 트리 생성 화면

(그림6)에서 구성된 글로벌 질의 트리를 XMDR메세지로 생성하여 각 Local Server로 전송한 다음에 (그림7)에서와 같이 메시지를 파싱하여 Local Server의 데이터베이스 스키마 항목을 1:1로 매핑을 하게 된다. 매핑을 하면서 Local Server의 데이터베이스 스키마에 맞는 질의가 생성되고, 만들어진 질의를 통해 데이터베이스를 검색하게 된다.

(그림7)는 하나의 Local Server의 로컬 질의로 변환하는 처리 과정을 보여주는 화면이다. (그림7)의 과정이 완료되면 (그림5)의 통합검색결과에 그 검색된 결과가 나타나게 된다.

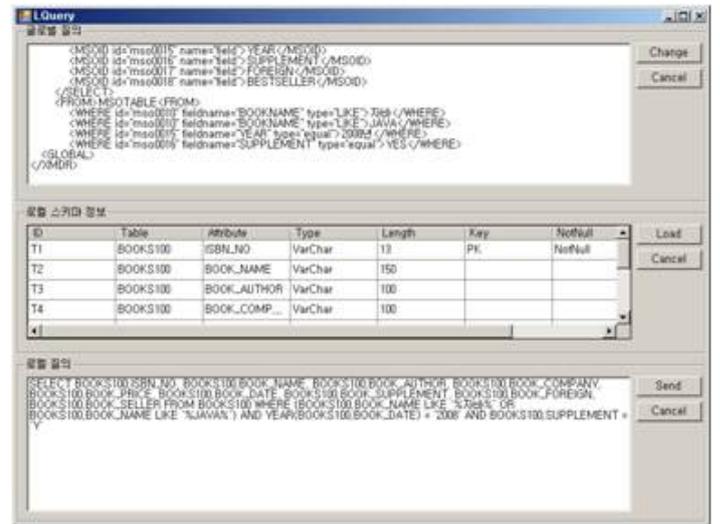


그림 7. 글로벌 질의에서 로컬 질의로 변환 화면

## 5.2. 비교 분석

본 논문에서 제안한 시스템은 멀티데이터베이스 환경에서 이질성 문제 해결과 데이터 통합검색을 지원하기 위한 시스템이다. 따라서, 표1에서와 같이 세 가지 프레임 워크를 대상으로 다음 7가지 항목으로 구분하여 비교하였다. 그 항목으로는 XMDR지원여부, 데이터 교환 자동화 여부, 범용 명세 스키마 지원여부, 적용범위, 저장소 구조, 메시지 교환, 데이터 연관성, 실시간 데이터 동기화에 대한 부분으로 비교가 된다.

표1. 데이터 교환을 위한 프레임워크간의 비교

	BizTalk	X-MAP	DBMS-Grid	본 시스템
XMDR 지원	지원하지 않음	지원하지 않음	지원하지 않음	지원함
데이터 교환 자동화	지원함	일부 지원함	지원함	지원함
데이터 동기화	일부 지원함	지원하지 않음	일부 지원함	일부 지원함
범용명세 스키마	지원하지 않음	지원함	지원함	지원함
저장소 구조	중양 집중	분산	분산	분산과 집중
메시지 교환 기법	사용 안함	사용	사용	사용
데이터 연관성	일부 지원함	일부 지원함	일부 지원함	사용
시스템 독립성	지원하지 않음	지원함	지원함	지원함

## 6. 결론 및 향후 연구

본 논문은 XMDR을 이용하여 멀티데이터베이스를 통합 검색하는 시스템을 제안하였다. 제안된 시스템은 XMDR을 이용하여 데이터요소의 의미 이질성과 개별 데이터베이스의 스키마 정보, 접속정보를 보다 효율적으로 구성하여 문제를 해결하였다. 따라서 본 시스템은 개별 데이터베이스간의 데이터 통합검색에 적합하며, 이질적인 데이터베이스 간의 데이터 통합검색에도 적합하다. 향후 연구과제로는 Grid개념을 도입하여 통합검색 뿐만 아니라 데이터 교환 및 수정이 가능한 시스템으로 확장할 필요가 있으며, 서비스 제공의 유연성으로 웹 서비스로의 확대 할 수 있도록 할 것이다.

## 참고문헌

- [1] Kevin D. Keck and John L. McCarthy, "XMDR: Proposed Prototype Architecture Version 1.01", <http://www.XMDR.org/>, February 3, 2005
- [2] W.Kim and J.Seo, "Classifying scematic and data heterogeneity in multi-database systems", IEEE Computer, Vol.24, No.12, pp.12-18, 1991
- [3] David Wang, "Automated Semantic Correlation Between Multiple Schema Information Exchange", M.I.T, MM, May, 2000
- [4] Ray Gates, "Introduction to MeD-Tutorial on ISO/IEC11179", Metadata Open Forum 2004, 2004
- [5] 문석재, 정계동, 최영근, "분산 데이터 상호운영을 위한 SQL/XMDR 메시지기반의 Wrapper를 이용한 데이터 허브 시스템", 한국해양정보통신학회, Vol.11, No.11, p.2047-2058, 2007
- [6] 박우창, "전역 스키마 생성도구를 이용한 데이터베이스 통합 및 질의 시스템", 한국인터넷정보학회, 인터넷정보학회논문지 제8권 제3호, pp.65-74, 2007. 6