

---

# TMDR 기반의 실시간 데이터 통합 환경 설계

정계동\* · 황치곤\*

Design of The Environment for a Realtime Data Integration based on TMDR

Kye-dong Jung\* · Chi-gon Hwang\*

---

본 논문은 2009년 광운대학교 교내연구비 지원에 의해 연구되었음.

---

## 요 약

본 논문은 레거시를 통합 검색하기 위한 방안으로 XMDR을 확장하는 방안을 제안한다. 이러한 확장은 메타데이터의 관리를 위한 메타 시멘틱 온톨로지, 위치 정보를 위한 메타 로케이션, 그리고 시멘틱 웹을 표현하기 위한 표준 언어인 토픽맵을 결합한다. 본 논문에서는 이것을 TMDR(Topic Map MetaData Registry)이라 한다. 토픽맵은 지식계층으로 인덱스와 같은 역할을 수행한다. 그러나 토픽맵은 데이터의 변화가 빈번한 경우에는 효율이 떨어질 수 있다. 이러한 문제를 해결하기 위해 본 시스템은 메타 데이터 사이의 관계, 실제 데이터 사이의 관계 그리고 메타데이터와 실제 데이터 사이의 관계를 토픽맵으로 표현한다. 표현된 토픽맵은 메타 데이터 간의 관계로 인해 실제 데이터간의 관계 변화를 줄이는 방안을 제시한다.

## ABSTRACT

This study suggests a method for extending XMDR to integrate and search legacy system. This extension blends MSO(Meta Semantic Ontology) for the management of metadata, ML(Meta Location) for the management of location information, and Topic Map which is the standard language used to represent semantic web. This study refers to it as TMDR(Topic Map MetaData Registry). As an intelligent layer, Topic Map functions like an index. However, if the data frequently changes, the efficiency of Topic Map may drop. To solve this problem, the proposed system represents the relation among metadata, the relation among real data, and the relation between metadata and real data as Topic Map. The represented Topic Map proposes a method to reduce the changing relation among real data caused by the relation among metadata.

## 키워드

MDR, Topic Map, TMDR, Ontology, Metadata

## I. 서 론

지속적으로 쏟아지는 거대한 정보의 바다에서 정보의 효율적 검색은 필수적이다. 효율적인 정보 획득은 새로운 정보를 생성하는 것 보다 더 중요하다.

또한, 이러한 정보들은 단일 데이터베이스 시스템을 통해서 제공되는 것이 아니다. 정보요구자가 필요로 하는 정보는 여러 시스템에 분산되어 있을 수 있고 중복되어 있을 수 있다.

이를 해결하기 위해서 기업의 각 조직, 주요 업무 그리고 핵심 어플리케이션에서 발생하는 물리적인 데이터들을 표준 규칙과 메타데이터를 이용하여 중복성이 제거된 데이터, 정확성이 보장된 데이터를 위해 데이터 통합이 필요하다[1]. 효율적인 통합과 레거시의 수정을 최소화하기 위해서는 이질적 환경, 의미적 이질성, 구조적 이질성 문제들을 해결해야 한다.

이를 위해 기존 데이터의 통합에서 발생하는 메타데이터간의 이질적인 문제를 해결하기 위해 XMDR을 사용한다[2]. 본 논문에서 사용되는 XMDR은 MSO(Meta Semantic Ontology), InSO(Instance Semantic Ontology), ML(Meta Location)로 구성된다[3].

이 XMDR의 장점은 몇 가지가 있다. 첫째, 레거시의 수정 없이 그대로 사용할 수 있으므로 재사용성이 높다. 둘째, 정보 교환의 표준을 제공한다. 셋째, 데이터 간의 구조적 · 의미적 이질성을 해소함으로써 단일 시스템과 같은 효과를 본다. 넷째, 레거시를 수정 없이 사용할 수 있어 새로운 시스템을 구축할 필요가 없기 때문에 비용을 절감할 수 있다[3]. 이러한 장점이 있는 반면에 정보간의 관계성을 고려한 정보제공이 부족하다는 단점이 있다. 이를 해결하기 위해 토픽맵 기술을 도입한다. 토픽맵은 온톨로지의 표현 기법 중 하나로 지식 접근을 계층적 접근뿐 만 아니라 관계성을 통한 수평적 접근을 가능하게 한다[4].

## II. 관련연구

### 2.1 XMDR

XMDR은 MDR과 온톨로지를 통합한 시스템이다. MDR은 데이터 통합을 위해 XML 기반의 관계형 데이터베이스의 메타데이터를 데이터베이스에 저장하는 기술

이며, 분산된 데이터의 이질성을 해결하기 위한 기술이다[5]. 온톨로지는 데이터간의 관계들로 구성된 의미론적 사전으로 계층적 구조를 가진다[6].

XMDR은 MSO, ML 그리고 InSO로 구성된다. MSO는 MDR과 같이 메타데이터 간의 관계성과 이질적 문제를 해결하기 위해 구축된다. ML은 레거시의 위치정보와 접근정보를 관리하는 부분으로 사용자에게 위치 투명성을 제공하기 위해 구축된다. 마지막으로 InSO는 데이터 간의 관계성과 이질적인 문제를 해결한다[7].

### 2.2 토픽맵(Topic Maps)

토픽맵은 분산 환경 하에서 지식 구조를 정의하고 정의된 구조와 지식 자원을 연계하는데 쓰이는 기술 표준이며, 정보자원의 구성, 추출, 네비게이션에 관련한 새로운 패러다임이라 할 수 있다. 토픽맵의 표준으로는 SGML(Standard Generalized Markup Language)에 기반을 둔 ISO/IEC 13250, 웹 환경에서의 사용을 위해 XML 문체계를 이용한 XTM(Xml Topic Map) 1.0(2001)이 존재한다[8].

토픽맵은 토픽/연관관계 모델에 토픽(topic)/어커런스(occurrence)를 추가함으로써 지식 표현과 정보관리 분야 사이에 연결고리 역할을 담당할 수 있게 되었다. 지식은 정보와 다르다. 어떤 것을 알고 있다는 것과 그것에 대한 정보를 가지고 있다는 것은 다른 것이다. 그런 측면에서 볼 때 지식관리는 ‘생성’, ‘형식화’, ‘전달’의 3가지 지식활동으로 요약할 수 있다. 토픽맵은 3가지 지식 활동 중 형식화를 위한 표준이며 생성과 전달을 위한 도구를 개발하는데도 필수요소이다.

토픽맵은 지식층과 정보층으로 구분되는 이중구조를 가진다. 지식층은 상위 계층으로 토픽과 토픽간의 연관관계로 구성된다. 정보층은 실제 데이터를 나타낸다[9].

## III. TMDR(Topic Map MetaData Registry)

관련연구에서 언급한 XMDR은 MDR의 기술에 온톨로지 기술을 적용한 InSO를 결합한 것이다. InSO는 인스턴스 시멘틱 온톨로지로서 실제 데이터간의 관계를 표현함으로써 의미론적 관계성을 밝히는 것이다. 이는 온톨로지의 일반적인 표현으로 계층구조를 가진다. 이

러한 계층구조에 따른 문제를 해결하기 위한 방법으로 계층구조의 장점인 수직적 지식 표현과 연관관계 (association)에 따른 수평적 지식 표현이 가능한 토픽맵을 결합한 TMDR은 MDR과 토픽맵을 결합한 용어로 데이터 검색, 이주 등의 작업에 효율성을 향상시키기 위한 패러다임으로 제안한다. 본 논문에서 MDR은 XMDR에서 메타데이터의 표준과 레거시의 메타데이터를 매핑하고 관리하기 위한 MSO와 토픽맵의 어커런스 역할을 수행하기 위한 ML부분을 결합하여 MDR이라 한다.

### 3.1 제안하는 토픽맵

본 시스템에서 적용하는 토픽맵은 본래의 의미를 확장하여 지식층의 토픽의 대상을 데이터뿐 만 아니라 데이터를 표현하기 위한 메타데이터까지 확대하여 메타데이터도 토픽의 대상이 되도록 한다. 이것은 데이터 간의 관계, 메타데이터 간의 관계, 데이터와 메타데이터 간의 관계성을 표현할 수 있다. 이러한 토픽들 간의 상호연관성을 분석하여 토픽맵 구조가 생성되고, 생성된 결과물이 토픽맵이다.

데이터의 삽입, 삭제 그리고 수정 등과 같은 변경이 빈번하게 발생하는 B2B 환경에서는 데이터가 실시간으로 빈번하게 발생하고 수정되는 성격을 가진다. 이러한 환경에서는 토픽맵을 운용하기 어렵다.

토픽맵은 이중구조를 가지는데 실제 운용데이터는 정보층에 저장되고 관리되어 진다. 이때 정보층의 빈번한 변화에 따라 지식층의 토픽과 토픽간의 관계성도 자주 변경된다. 이는 토픽맵을 유지하기 위한 자원낭비로 성능이 떨어질 수 있다. 이에 따라 데이터만으로 토픽맵을 구축하는 것이 아니라 빈번하게 발생하는 정보에 대한 분석과 관계성을 파악하여 이에 대한 데이터, 메타데이터 그리고 이 둘 간의 관계까지 포함한 상호연관성을 파악하여 토픽맵을 구성한다. 이 토픽맵을 통해 데이터의 유용성이 높은 실시간 시스템에서도 적용할 수 있다.

표1은 기존 토픽맵과 제안하는 토픽맵과의 차이를 요약한 것이다. 기존 토픽맵은 데이터들의 특정 주제로 구성되지만, 제안하는 토픽맵은 메타데이터와 데이터를 토픽으로 하여 구성된다. 이 둘을 비교해 보면, 연관관계 (association)와 역할 타입(Role type)은 양쪽 모두 그대로 사용하지만 토픽의 차이 때문에 실제로는 같다고 볼 수 없다. 토픽은 MDR의 표준 항목인 표준 메타데이터와 데이터로 확장되고, 토픽클래스는 토픽을 계층으로 구성

될 때 상위 계층의 토픽은 하위계층에 대한 토픽클래스가 된다. 어커런스는 ML의 레거시의 접근정보와 권한정보를 통해 표현한다.

표 1. 기존 토픽맵과 비교  
Table 1. Compare with existing Topic Map

토픽맵	제안하는 토픽맵
토픽	MDR의 MSO의 메타데이터 표준 항목, 데이터 주제
토픽클래스	
연관관계	연관관계
역할타입	역할타입
역할	연관관계에 대한 역할
어커런스	정보자원의 위치, MDR의 ML

본 시스템은 구축된 토픽맵에서 연관관계를 선택하고, 연관관계가 선택되면 연관관계와 관련된 역할 타입이 선택되어 지고, 역할 타입에 따른 토픽을 이용할 수 있다. 또한 특정 토픽으로부터 연관관계를 통하여 다른 토픽을 검색할 수 있다. 이러한 과정이 그림 1과 같이 표현되고, 이는 토픽맵의 일부분이다.

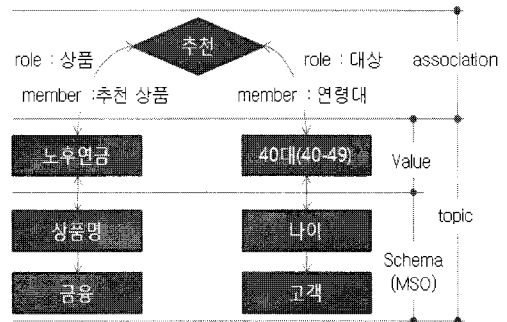


그림 1. 제안하는 토픽맵의 표현  
Fig 1. presentation of suggest the Topic Map

### 3.2 MDR과 토픽맵의 융합

토픽맵은 메타데이터와 데이터 모두를 주제로 하는 토픽과 토픽들 간의 상호연관관계를 표현한 것이다. 이때 사용되는 메타데이터는 레거시의 메타데이터를 통합하기 위한 표준으로 MDR의 MSO에서 관리한다. MSO는 표준과 레거시 사이의 매핑정보도 관리하고, 레거시를 접근하기 위한 정보인 ML도 관리하고 있다. 이

ML은 기존 토픽맵에서 자원의 접근 정보를 표현하는 어커런스로 이용된다. 그러므로 토픽맵과 MDR의 융합이 필요하다. 융합은 관계데이터베이스의 ERD로 그림과 같이 표현된다. 우선 토픽맵은 XML로 기술할 수 있으므로 관계데이터베이스에 저장할 수 있고, MDR 또한 관계데이터베이스로 저장 가능하기 때문이다.

그림 2는 MDR과 토픽맵을 융합하기 위한 ERD를 설계한 것이다. MDR은 메타데이터의 표준과 레거시를 접근하기 위한 위치정보와 권한정보를 표현한 ML이다. 토픽맵의 토픽과 MDR의 융합은 토픽맵의 어커런스를 MDR의 ML과 조인으로 가능하다.

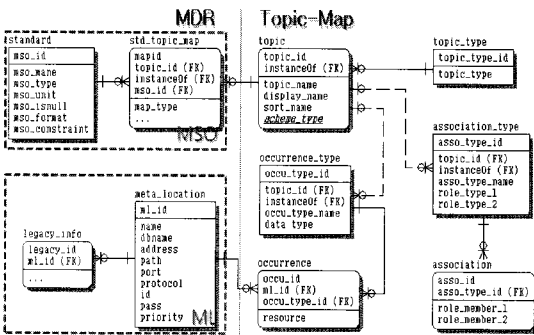


그림 2. XMDR과 토픽맵의 융합을 위한 ERD  
Fig 2. ERD for blend the XMDR with the Topic Map

이는 일반적인 토픽맵을 관계데이터베이스로 표현한 것과 같다. 그러나 몇가지의 차이점을 가진다. 첫째, 메타데이터와 실제 데이터 모듈을 토픽의 주제로 사용할 수 있으므로 이를 구분하는 속성(`schema_type`)이 추가적으로 운용되고 있다. 예를 들면, `schema_type`의 데이터 타입을 논리형으로 하고, 이것이 '1'의 값을 가지면 MDR의 표준을 토픽으로 사용하고, '0'의 값을 가지면 MDR의 표준에 해당하는 값을 토픽으로 한다. 그림 2에서 보면 `standard` 테이블과 `topic` 테이블은 `std_topic_map` 테이블을 통해서 두 테이블을 매핑시킨다.

둘째, 어커런스는 토픽맵에서 토픽을 표현하는 자원영역을 나타내는데, 본 시스템은 ML을 이용한다. `occurrence` 릴레이션은 `occurrence_type` 릴레이션과 `meta_location` 릴레이션을 연동하여 토픽맵의 어커런스 역할을 수행한다.

그림2에 표현된 토픽맵의 각 부분은 토픽, 토픽타입,

관계타입, 관계, 어커런스 타입과 어커런스로 구성된다. 이 부분을 간략히 살펴보면 다음과 같다.

- 토픽(topic) : 메타데이터의 주제('상품', '분류', '금융', '이름', '직업' 등)와 메타데이터들의 인스턴스인 실제 데이터의 주제('40대', '주택부금', '청약저축', '남자' 등)
- 토픽타입(topic\_type) : 토픽을 표현하기 위한 상위 토픽으로 고객의 성별인 '남자'의 토픽타입은 '성별'으로 표현되는 것과 같다.
- 관계(association) : 연관관계와 연관관계에 대한 역할 멤버를 표현한다. 예를 들면, 혼인관계에서 멤버는 남편과 아내의 이름이 된다.
- 관계타입(association\_type) : 연관관계를 가지는 연관관계의 유형과 연관관계를 가지는 멤버의 역할타입(role type)을 표현한다. 관계에서 혼인관계에 대한 역할타입은 남편과 아내가 된다.
- 어커런스(occurrence) : 어커런스는 토픽을 실제 데이터와 엮어놓는 역할을 한다. 본 시스템에서는 토픽을 ML과 묶어 놓음으로 해결할 수 있다.
- 어커런스타입(occurrence\_type) : 어커런스 타입은 어커런스의 유형으로 토픽이 가지는 관계의 유형이다.

#### IV. 시스템 설계

제안하는 TMDR을 이용한 시스템은 4단계의 기억계층으로 구성된다. 각 계층의 구성은 그림 3과 같이 비즈니스(business) 계층, 데이터베이스 서비스(database service) 계층, TMDR 서비스(TMDR service) 계층, 데이터 자원(data resource) 계층으로 구성된다.

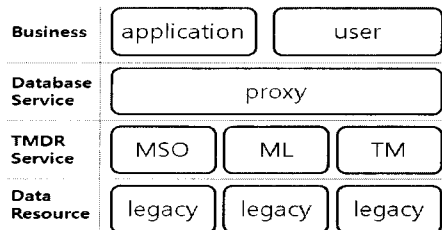


그림 3. 제안하는 시스템의 기억계층  
Fig 3. Storage hierarchy of suggest system

#### 4.1 시스템의 구성과 운용

비즈니스 계층은 어플리케이션 프로그램, 사용자 등으로 서비스를 제공받는 사용자 계층이다. 데이터베이스 서비스 계층에서 생성된 정보를 제공받아 이용하는 계층이다.

데이터베이스 서비스 계층은 프록시로 구성된다. 비즈니스 계층에서 요구하는 정보를 제공하기 위한 계층으로 서비스 데이터 저장 계층이다. 프록시는 비즈니스 계층에서 요구하는 실시간 정보를 제공한다. 사용자의 요구를 TMDR 서비스를 이용하여 데이터 자원 계층에 있는 레거시의 수집된 정보를 일시 보관하여 사용자 또는 시스템에게 제공한다.

TMDR 서비스 계층은 데이터베이스 서비스 계층을 지원하기 위한 계층이다. 이 계층은 기존의 레거시의 정보를 통합관리하고 연관관계를 부여하여 연관처리를 하기위한 본 시스템의 핵심계층이다. 데이터 자원 계층의 레거시를 통합하기 위한 표준 메타데이터와 메타데이터의 매핑정보를 저장 관리하는 MSO, 각 레거시의 접근을 위한 위치 정보와 접근을 위한 권한 및 보안 정보를 저장 관리하는 ML, 메타데이터와 데이터 그리고 이 둘 사이의 의미적 이질성 해결하기 위한 토픽맵을 생성, 저장 관리하는 토픽맵으로 구성된다.

데이터 자원 계층은 기존에 구축되어 있는 시스템들로, 운용 데이터들이 생성되고 저장되어 있는 공간으로 대부분 관계 데이터베이스로 사용한다. 시스템은 이렇게 기억 계층 구조를 가지고 있다. 이 계층을 이용한 시스템의 운용과정은 그림 4와 같다.

데이터의 흐름은 전체적인 구조에서 메타데이터를 관리하기 위한 데이터의 흐름과 운용 데이터를 검색하고 이용하기 위한 실제 데이터의 흐름 이렇게 두 가지 흐름을 가진다. 첫째, 데이터 검색의 흐름은 그림에서 직선 화살표로 **data search flow**라고 표현된 것이다. 실제 데이터를 검색하고 이용하기 위해 레거시에 접근하기 위한 흐름들이다. 이 흐름에서 중요한 부분은 인터페이스를 통한 요구사항을 MSO와 토픽맵을 이용하여 요구사항에서 검색항목과 검색조건을 분리하는 것이다. 검색항목은 MSO를 이용하여 메타데이터들의 표준을 추출하고, 검색조건은 토픽맵에서 연관관계를 추출하여 요구사항을 충족하기 위한 토픽맵의 연관관계를 추출하도록 한다. 이 추출된 정보를 바탕으로 DQP(Distributed Query Process)에서 수행하기 위한 질의를 생성한다. 둘

째, 레거시간의 인터페이스를 위한 메타데이터 관리가 필요하다. 이러한 메타데이터의 관리에는 MDR에서 수행한다. 그림 4에서 보는 것과 같이 MDR은 MSO와 ML로 구성되는데 이는 관련연구에서 언급한 XMDR의 MSO와 ML이다. 이러한 운용과정에서 핵심 역할은 TMDR 계층이다.

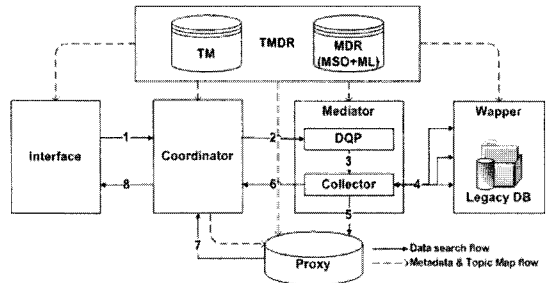


그림 4. 시스템 운용 과정  
Fig 4. operation process of system

#### 4.2 TMDR의 스키마 매핑

분산된 레거시를 통합하기 위해서는 메타데이터 관리가 중요하다. 이를 위한 메타데이터 모델은 레거시의 메타데이터를 통합하여 하나의 표준을 제공하고, 레거시의 위치투명성을 제공하기 위한 ML로 구성된다. 사용하는 데이터의 소유자를 밝히기 위한 사용자 정보를 포함한다. 이 과정은 TMDR에서 사용자 정보에 따라 요구 항목과 요구조건 항목에 대한 MDR과 레거시의 로컬 메타데이터의 매핑과정이다. 이 과정을 표현한 것은 그림 5와 같다.

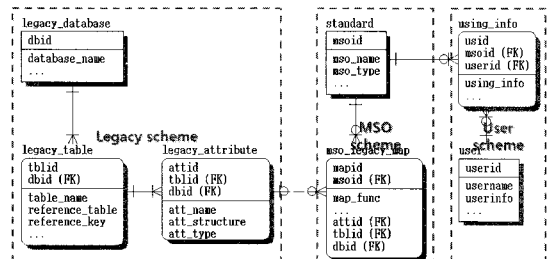


그림 5. 메타데이터 스키마 매핑  
Fig 5. Schema mapping of metadata

이 과정은 우선 사용자의 로그인에 대한 사용자 접근 정보를 확인하고 사용자가 표준항목을 접근할 수 있도록 한다. 사용자의 접근에 따른 표준 인터페이스 제공을

위한 부분이다. 표준항목(standard)으로 사용자의 요구 항목이 결정되어 지면 이에 따라 표준과 레거시 항목의 매핑정보(mso\_legacy\_map)를 이용하여 각 레거시의 데이터베이스에 저장된 데이터의 이질성 문제를 해결하기 위한 기능을 수행할 수 있다. 그림 5의 과정은 사용자의 접근에서 MSO를 통한 표준 메타데이터를 이용하고 이를 통하여 레거시까지 접근하는 전체의 과정이다. 이는 표준 질의의 생성과 생성된 표준 질의를 통한 레거시를 위한 개별적인 질의로 파싱을 수행할 수 있도록 한다.

4.3 TMDR의 토픽맵을 이용한 질의 생성

TMDR에서 질의의 생성은 일반적인 질의와 다르다. 여기서 질의의 생성은 두 부분으로 분할된다. 생성방법은 먼저 사용자 요구에 대한 항목을 MSO에서 표준항목을 추출하는 부분이다. 이는 4.2에서 기술하였다. 두 번째는 토픽맵에서 토픽과 연관관계를 조건으로 이용하기 위한 부분이다. 이렇게 두 부분의 결합으로 질의가 생성된다.

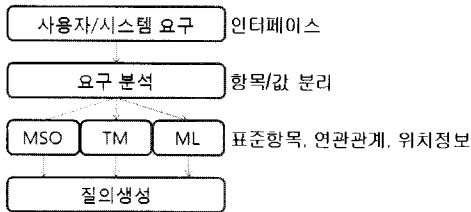


그림 6. 질의 생성 단계  
Fig 6. Phase of Query Generation

그림 6은 질의 생성 과정이다. 질의 생성 과정은 사용자의 요구나 시스템의 요구를 인터페이스를 통해 받아들이면 이 요구를 분석하여 검색이나 조건의 항목이 되는 항목과 조건이 되는 부분으로 분리한다. 이렇게 분리된 것은 항목에 해당하는 것은 MDR의 MSO를 통해서 표준 항목을 추출하고, 조건의 항목도 마찬가지로 MSO에서 추출한다. 검색의 조건에 해당하는 부분은 토픽간의 관계성을 표현한 토픽맵을 통해 조건과 연관관계를 가지는 항목과 조건을 추출한다. 이에 추출된 항목의 데이터베이스 위치, 토픽맵의 어커런스 위치는 MSO의 ML에서 수행하므로 ML을 통해서 추출하여 표준 질의를 추출하여 질의를 생성한다. 수행 과정은 그림 7과 같으며, 이는 그림 6의 과정을 도식화한 것이다.

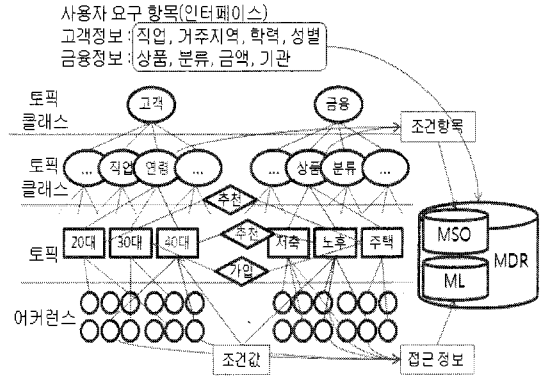


그림 7. 질의 사례  
Fig 7. Example of Query

그림 7은 질의 사례로 사용자의 인터페이스 조작에 따라 요구항목이 결정되고, 제시된 연관관계를 통해서 조건을 선택한 경우를 바탕으로 가정하여 표현했다. 우선 사용자의 요구항목은 고객정보와 이 고객정보들과 관련된 금융정보를 선택했다. 각 요구항목은 MDR의 MSO를 통하여 표준항목을 추출한다. 그리고 사용자의 조건은 토픽맵으로 제공되므로 사용자가 선택한 조건은 토픽이 되고, 이 토픽에 따른 금융 토픽과의 연관관계를 추출한다. 즉, 그림과 같이 고객에 대한 연령대가 40대인 조건을 선택하고 금융정보를 같이 보겠다고 했다면 '연령=40대'라는 조건과 추천 연관관계에 따라 금융상품으로 노후가 선택되는 것을 볼 수 있다. 이에 따라 노후 토픽의 상위토픽인 상품 토픽을 통하여 '상품=노후'조건이 추가된다.

데이터웨어하우스를 구축하고자 할 때, 이질적 데이터를 통합하기 위한 메타데이터의 관리, 데이터 사이의 관계성 고려, 데이터를 추출, 변환, 저장하는 과정에서 추출과 변환을 위한 기법이 필요하다. 이 때 본 시스템에서 제안하는 방식을 적용할 수 있다.

V. 비교평가

본 논문을 평가하기 위해서 유사한 시스템과 비교평가 한다. 비교 대상은 XMDR, 토픽맵을 본 논문과 비교한다. 토픽맵은 시멘틱 웹을 구현하기 위한 지식 표현방법이고, XMDR은 기존의 분산 시스템을 통합하기

위한 시스템이다. XMDR의 지식 표현 방법으로 InSO (Instance Semantic Ontology)를 사용한다. 이러한 시스템과 TMDR은 다음 항목에서 비교 평가한다.

표 2. 기존 시스템과의 비교  
Table 2. Compare with the existing system

		토픽맵	XMDR	TMDR
메타 데이터	통합	부분	지원	지원
	관계성	부분	지원	지원
	이질성	부분	지원	지원
데이터	관계성	연관관계	매핑	연관관계
	이질성	지원	지원	지원
온톨로지 표현		토픽 연관성	메타데이터 연관성, 데이터 연관성	메타데이터와 데이터 상호 연관성
지식구조		수직, 수평	수평	수직, 수평
지식관리 대상		토픽	데이터	메타데이터 + 데이터

비교 평가 항목은 표 2와 같이 메타데이터, 데이터, 온톨로지 표현 그리고 지식관리 대상에 대해서 수행한다. 메타데이터는 통합 방안 고려 여부, 관계성 그리고 이질성 해결 방안 고려 여부로 세분화 된다. 데이터는 데이터 간의 관계성과 이질성 해결 방안으로 세분화 하여 비교한다.

기존 데이터베이스가 구축된 형태가 다르기 때문에 이에 대한 스키마 정보도 다르다. 이런 기존 시스템을 통합하기 위한 스키마 정보의 통합이 메타데이터의 통합이다. 토픽맵은 데이터를 토픽의 단위로 인식하여 관계성을 고려할 수 있으므로 부분 지원한다. XMDR과 TMDR은 메타데이터를 위한 MSO를 제공하므로 지원한다. 관계성과 이질성 또한 동일하다.

데이터 관계성은 토픽맵은 토픽간의 연관관계를 통하여 지원하고, XMDR은 스키마와 데이터간의 연관관계 매핑을 통하여 지원한다. 본 시스템에서는 메타데이터와 데이터의 연관관계를 지원한다. 데이터 이질성은 비교 대상이 되는 모든 시스템이 지원한다.

온톨로지의 표현은 온톨로지를 표현하는 방법에 대한 비교 평가 항목이다. 토픽맵은 데이터의 주제를 정한 토픽과 토픽간의 연관관계를 통하여 지원한다. XMDR은 메타데이터간의 연관관계를 MSO를 통해 표현하고,

데이터간의 연관관계 InSO를 통해 표현한다. 본 시스템에서는 데이터뿐 만 아니라 메타데이터까지의 연관관계를 표현함으로써 시스템 운용의 효율성을 높인다.

지식구조는 수직과 수평구조로 비교한다. 수직구조는 지식을 표현하기 위한 구조를 계층적으로 표현한다. 이런 구조는 동일한 부모를 가지는 지식을 한번의 접근으로 이용할 수 있다는 장점을 가진다. 반면에, 상위계층에서 하위계층으로 지식을 검색하는 것으로 동등한 수준일 경우 같은 부모를 가지는 지식이 아니면 상위계층으로 이동하여 필요로 하는 하위계층으로 이동해야 하는 단점이 있다. 수평구조는 수직구조와 반대의 특성을 가진다. 따라서 이러한 경우는 수직구조와 수평구조를 모두 지원하는 것이 효율적이다. 마지막 비교항목인 지식관리의 대상은 지식을 관리하기 위한 대상을 무엇으로 하는지에 대한 비교평가 항목이다. 토픽맵은 지식층에서 토픽과 토픽간의 관계를 표현한다. XMDR은 InSO에서 실제 데이터간의 관계성과 이질성을 해결한다. TMDR은 토픽맵에서 제공하는 토픽맵을 이용하되 토픽의 단위를 메타데이터, 데이터 그리고 이들 간의 관계를 표현하여 토픽맵의 지식층을 확대하였다.

#### IV. 결론

본 논문에서는 데이터의 통합검색을 효율적으로 수행하기 방안으로 TMDR을 이용하는 통합 환경을 제안했다. TMDR은 간단하게 표현해서 MDR과 토픽맵의 결합이라 볼 수 있다. 그러나 적용하는 본 시스템에서의 토픽맵은 기존의 방식으로 작성된 것이 아니라 MDR에서 구축된 메타데이터의 표준과 실제 데이터를 토픽으로 구성하고, 이에 필요한 연관관계를 표현하는 방안을 제안하였다. 본 시스템에서 제안하는 이 구조는 데이터 웨어하우스와 같이 대용량의 데이터를 특정 지식 구조에 따라 수집하고자 할 때, 데이터 추출하여 지식 시스템을 구축하기 위한 형틀을 제공할 수 있다. 이는 데이터 웨어하우스뿐 만 아니라 데이터를 수집하고 이를 이용하기 위한 모든 기술이나 업무에서 적용할 수 있는 방안이다.

본 시스템은 XMDR이 가진 장점은 모두 그대로 유지하면서 추가적인 장점을 가진다. 먼저, 데이터를 접근하기 위한 지식구조로 토픽맵을 도입함으로써 데이터간의 연관관계를 표현할 수 있다. 둘째, 지식표현이 계층구

조로 표현되는 일반적인 구조에서 가질 수 있는 단점을 보완하여 수평적 접근을 가능하게 했다. 셋째, 토픽의 범위를 메타데이터로 확장함으로써 레거시들의 효율적 통합과 연관성 있는 데이터 접근을 가능하도록 했다.

이후 연구는 TMDR의 토픽맵을 구축하기 위한 각 부분에 대한 구체화가 필요하다. 특히, 데이터웨어하우스를 위한 TMDR 적용과 분산 질의 처리 부분이다. 일반적인 분산 질의 처리는 정형화된 질의가 일반적인 경우인데 반해 본 논문에서는 정형화된 질의가 아니므로 효율적인 분산 질의 처리를 연구해야 할 필요성이 있다.

### 참고문헌

- [1] 나정욱, "oracle data hub 실시간 기업을 위한 데이터 통합", Oracle Korea Magazine 2005 Summer, 2005.
- [2] Kevin D. Keck and John L. McCarthy, "XMDR: Proposed Prototype Architecture Version 1.01", <http://www.XMDR.org/>, February 3, 2005.
- [3] 황치곤, 정계동, 최영근, "XMDR 기반의 통합 검색을 위한 데이터 그리드 Wrapper 설계", 한국해양정보통신학회, Vol.12, No.5, p.921-929, 2008.
- [4] Members of the Topic Map.org Authoring Group, "XML Topic Maps(XTM) 1.0", [www.Topic Maps.org/xtm/](http://www.Topic Maps.org/xtm/), 2001.8.
- [5] Ray Gates, "Introduction to MDR-Tutorial on ISO/IEC 11179", Metadata Open Forum 2004, Xian, May 17, 2004.
- [6] T.R.Grubber, "Towards principles for the design of ontologies used for knowledge sharing", International Journal of Human-Computer Studies, Vol.43, No.5, 1995, pp.1-2.
- [7] 황치곤, 정계동, 강석중, 최영근, "분산환경에서 효율적인 데이터 검색을 위한 XMDR 기반의 토픽맵 시스템 설계", 한국해양정보통신학회, Vol.13, No.3, p.586-593, 2009.
- [8] 오삼균, "토픽맵s 응용 표준 및 활용 가이드라인 개발", 한국전자거래진흥원, 2003. 10.
- [9] H.Holger Rath, "The Topic Map handbook", empolis, 2003. 2. 6.

### 저자소개



정 계 동(Kye-dong Jung)

1985년 광운대학교 전자계산학 (이학사)

1992년 광운대학교 산업정보학 (이학석사)

2000년 광운대학교 컴퓨터과학(이학박사)

1993년 ~ 2004년 광운대학교 정보과학원 교수

2005년 ~ 현재 광운대학교 교양학부 교수

※관심분야: XML 분산시스템, 분산 컴퓨팅기술, 이동에이전트



황 치 곤(Chi-gon Hwang)

1995년 창원대학교 경영학과(학사)

2004년 광운대학교 정보통신학과 (공학석사)

2008년 ~ 현재 광운대학교 컴퓨터 과학 박사과정

2006년 ~ 현재 (주)인찬 연구원

※관심분야: 웹서비스, XMDR, 그리드컴퓨팅, 이동에이전트, 상호운용