
TMDR 기반의 글로벌 쿼리를 이용한 적응적 쿼리 변환에 관한 연구

황치곤* · 신호영** · 정계동*

*광운대학교 · **경북대학교

A study on the adaptive query conversion using TMDR-based global query

Chi-Gon Hwang* · Hyo-Young Shin** · Kye-Dong Jung*

*KwangWoon University · **KyungBok University

E-mail : duck1052@kw.ac.kr

요 약

본 연구는 네트워크상에 분산된 데이터들의 이질적인 문제를 해결하고 효율적인 데이터 통합을 위한 Topic Maps MetaData Registry(TMDR)기반의 쿼리 변환 기법을 제시한다. TMDR은 분산 데이터를 통합하기 위해 글로벌 스키마를 제공하여 각 로컬 데이터에서 발생하는 정보들 사이의 이질성 문제를 쿼리 변환 기법으로 해결한다. 이 기법은 eXtended Meta Data Registry(XMDR)의 Meta Schema Ontology(MSO)와 Topic Maps에서 연관 관계를 분석하여 로컬의 접근 정보를 관리하는 Meta Location(ML)의 정보를 통해 분산되어 있는 로컬 데이터들을 통합 접근할 수 있도록 한다. 처리 방법은 TMDR을 이용하여 전역처리를 위한 글로벌 쿼리를 생성하고 생성된 글로벌 쿼리는 로컬 스키마로 분해되어 네트워크를 통해 분산되어 있는 시스템들을 접근하도록 함으로써 통합 접근이 가능하게 한다. 우리는 이를 위해 글로벌 쿼리를 로컬에 적응적인 쿼리로 변환하는 방안을 제안한다.

ABSTRACT

This study suggests a query conversion method based on Topic Maps MetaData Registry(TMDR) in order to solve heterogeneity problems distributed in networks and to integrate data efficiently. In order to integrate distributed data, TMDR provides global schema and it solves heterogeneity problem within local data using query conversion method. After analyzing relationship between Meta Schema Ontology(MSO) of eXtended Meta Data Registry(XMDR) and Topic Maps, this method allows integrated access through Meta Location(ML) which manages accessing information of local data. The processing method is to produce a global query for global processing by using TMDR and then to make the produced global query approach to systems distributed through networks so that allows integrated access at the end. For this, we propose a method to convert a global query into a query which is adaptive to local query.

키워드

TMDR, 분산 환경, 쿼리 변환, 이질성 해결, Topic Maps

I. 서론

오늘날 대부분의 정보의 검색은 웹 기반의 서비스이다. 웹을 이용한 정보 검색은 대부분 키워드 중심의 검색이 도입되어 사용 중이다[1].

기업의 합병이나 협업에 의한 새로운 시스템 구축은 데이터 통합과 상호운용성의 문제가 발생한다. 그리고 분 각 로컬 시스템이 관리하고 있는 정보들은 의미적, 구조적 이질성 문제를 가지고 있다. 그러므로 상호운용을 위한 자원의 공유는

많은 문제가 발생할 수 있다[2]. 특히, 실시간으로 정보를 공유해야 하는 경우, 정보 공유 시스템은 의미적 또는 논리적 구조의 충돌을 감지하고 해결할 수 있는 능력이 필요하다. 그러나 대부분의 정보 공유는 기존 로컬 시스템을 수정해야 되는 경우가 많다. 그러므로 각 로컬 시스템의 독립성을 보장하는 데이터 통합은 정보의 의미적 충돌을 관리해야 하는데 쉽지 않다.

이에 우리는 효율적인 데이터 수집과 통합을 위해 데이터들 사이에 발생하는 충돌을 해결할 수 있는 TMDR 기반의 시맨틱 에이전트 시스템을 제안하였다[3]. 분산된 데이터들은 일반적으로 데이터를 저장하기 위한 스키마와 실제데이터로 구성되는데, 이를 한꺼번에 통합할 수는 없다. 그래서 스키마를 결합하여 통합스키마를 만들고 이 통합스키마에 실제 데이터들 간의 이질성을 해결해서 수집해야 한다. 이에 통합된 스키마는 MDR[4]과 같은 방법으로 글로벌 스키마를 생성하여 로컬 스키마와 매핑함으로써 통합된 글로벌 스키마만으로 로컬 데이터를 접근할 수 있도록 한다. 그리고 실제 데이터들 간의 이질성은 Topic Maps[5]에서 토픽과 연관관계(association)를 표현한 것과 같이 메타데이터와 데이터를 표현하는 각 토픽클래스와 토픽으로 등록하고, 메타데이터 사이의 관계, 데이터 사이의 관계 그리고 메타데이터와 실제 데이터 사이의 관계와 같은 연관관계를 제공함으로써 데이터의 의미 및 연관성을 기술할 수 있고 이를 통해 다양한 형태의 의미충돌을 해결할 수가 있다. 시스템 구축은 관련 집단의 구성원들 사이에서 합의하여 도출된 지식을 통해 TMDR을 구축하고, 이를 이중 시스템 간의 조정을 위해 사용한다면 시스템 간의 의미적 이질성을 해결하고 적응적인 쿼리 변환에 이용할 수 있다.

이후 논문의 구성은 2장에서 쿼리 변환을 위한 TMDR에 관한 연구, 3장에서 TMDR을 이용한 적응적인 쿼리 변환 기법 4장에서 시스템 적용 5장에서 결론을 기술한다

II. 쿼리 변환을 위한 TMDR

TMDR은 Topic Maps과 XMDR을 결합한 데이터 통합을 위한 지식 저장소이다. 결합은 다음 방법을 통해 결합한다[3]. XMDR은 로컬 시스템의 접근정보를 관리하는 ML을 두고, Topic Maps은 지식을 접근하기 위한 어커런스 역할을 XMDR에 위임한다. Topic Maps에서 사용하는 스키마 토픽은 XMDR의 글로벌 스키마를 통해서 수행한다

XMDR은 데이터의 상호 운용성을 확보하기 위해 고안된 것으로, 구성은 MSO, INSO(INstance Semantic Ontology), ML이다. 이 중에서 INSO는 데이터 간의 관계성을 표현한다 이 역할은 Topic Maps을 통해서 수행하도록 하여 TMDR을 구성한다. MSO는 스키마의 구조적 의미적 이질성을

해결하기 위해 온톨로지이다. 이것은 글로벌 스키마와 로컬 시스템의 스키마 정보 간의 매핑정보 스키마 간의 이질적 관계에 대한 솔루션 정보 데이터타입 간의 변환정보를 세부적으로 관리한다 ML은 로컬 시스템에 접근하여 검색하기 위한 위치정보, 권한정보, 각 로컬 시스템에 적합한 쿼리로 변환에 사용하게 될 데이터베이스 정보와 테이블 정보 그리고 검색을 위한 우선순위 정보를 관리하고 제공하는 역할을 수행한다 이것은 사용자가 실제 데이터의 위치를 몰라도 접근할 수 있게 하고, 표준 온톨로지에 표현된 데이터의 표준과 매핑을 통해 각 로컬 시스템의 MDR에 접근할 수 있는 방법을 제공하게 된다. 이를 통해 수행되는 통합은 기본적으로 다음을 따른다.

$$I = \langle G, L, M \rangle$$

수식에서 I 는 통합된 결과 이고, 이것은 글로벌 스키마(G), 로컬 스키마(L) 그리고 두 스키마간의 매핑(M)을 통해서 수행한다는 것을 의미한다 이러한 통합된 쿼리 결과를 글로벌 뷰라 하고 글로벌 뷰는 로컬 뷰의 합집합으로 구성된다 이 글로벌 뷰에 기록되는 튜플 t 의 집합은 다음과 같은 관계를 가진다.

$$t(Q^g) \supseteq t(Q^l)$$

Q^g 는 글로벌 뷰를 의미하며 Q^l 는 로컬 데이터베이스를 검색한 결과를 의미한다 로컬 통합 쿼리 결과와 변환된 각 로컬 시스템의 쿼리 결과 간의 관계는 다음과 같다.

$$q = q^{L_1 D_1} \cup q^{L_2 D_2} \cup q^{L_3 D_3} \cup \dots \cup q^{L_n D_n} = \bigcup_{i=1}^n q^{L_i D_i}$$

q 는 모든 시스템의 쿼리를 수집한 결과이고 $q^{L_i D_i}$ 는 로컬 쿼리로 i 번째 로컬 시스템의 로컬 데이터베이스를 접근한 쿼리 결과이다

TMDR을 구성하는 6가지의 릴레이션 집합으로 표현된다. 릴레이션은 테이블 하나를 말하는 것이 아니라 하나 또는 그 이상의 테이블로 구성될 수 있기 때문에 릴레이션 군이라 볼 수 있다 이는 데이터 상호운용을 수행하기 위한 기반 영역(Domain : D)이 된다. 이 영역 D 는 다음과 같은 릴레이션으로 이루어진다

$$D = (G, L, M, T, A, m)$$

릴레이션 G, L, M 은 XMDR에 포함된 릴레이션이다. 이들 릴레이션은 글로벌 스키마(G), 로컬 스키마(L), 그리고 두 스키마간의 매핑(M)을 표현한다. 릴레이션 T, A 는 Topic Maps에 포함된 릴레이션이다. m 은 XMDR과 Topic Maps을 매핑을 위한 릴레이션이다[6].

III. TMDR을 이용한 적응적 쿼리 변환

글로벌 쿼리는 사용자가 실제의 스키마와 관계 없이 원하는 정보와 조건을 제시하고, 로컬 시스템에 대한 세부 내용을 몰라도 필요한 정보를 찾을 수 있게 한다. 로컬 시스템에 접근하여 쿼리를 수행하는 것은 TMDR을 통해 생성된 로컬 쿼리로 수행한다. 글로벌 쿼리는 글로벌 스키마를 바탕으로 사용자의 요구를 적용하여 생성한다 이것은 TMDR에 따라 해당 로컬 시스템과 의미적 이질성을 적용하고, 각 로컬 시스템에 적합하게 적용된 로컬 쿼리로 변환하여 로컬 시스템을 접근한다.

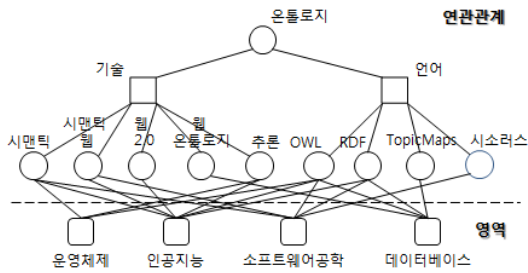


그림 1 TMDR과 키워드를 통한 쿼리 적응 사례

적응적 쿼리 변환을 위한 지식은 TMDR에서 사용자 키워드를 통하여 추출함으로써 쿼리 조건을 확장할 수 있도록 한다. TMDR에 생성된 내용은 주기적인 갱신하고, 사용자 키워드를 기반으로 관계성을 추출하여 적응적 쿼리 확장에 이용할 수 있다. 그림 1은 사용자 키워드 'Topic Maps'을 통하여 추출된 적응 사례이다. 그림 1에서 원은 키워드들로 추출된 빈도에 따라 가중치를 부여하여 정확도를 향상할 수 있고, 적용된 내용에 따라 분야별, 관계별로 쿼리를 확장할 수 있다. 이에 따른 글로벌 쿼리는 FROM절을 제외하고 필드와 조건만 나열하는 형식을 이용한다. 필드는 글로벌 스키마를 이용하고, 조건은 글로벌 스키마와 그림 1과 같은 Topic Maps를 이용한다. 기본형식은 다음과 같다.

```
SELECT <gbl_item_list> WHERE <cond>;
```

위 형식에서 <gbl_item_list>는 글로벌 스키마로 인터페이스에서 사용자가 선택한다 <cond>는 사용자의 요구 조건이다. 이러한 규칙을 통하여 글로벌 쿼리는 생성된다. 글로벌 쿼리를 로컬 쿼리로 변환하는 과정은 다음 단계를 거쳐 수행한다.

적용한 질의 사례는 키워드를 "TopicMaps"으로 하는 연구결과를 검색한다. 검색 대상은 제목, 저자, 학회 순으로 한다.

첫째, 사용자 검색에 대한 글로벌 질의를 생성한다. 글로벌 스키마는 로컬 스키마에 대한 대표

항목으로 실제 테이블은 없기 때문에 FROM절은 없다. 위의 사례를 적용한 질의는 다음과 같다

```
SELECT title, author, society
WHERE keyword='TopicMaps'
```

둘째, 질의 분해 과정으로 글로벌 질의에서 표준항목과 인스턴스를 추출한다. 표준항목은 title, author, society, keyword, lab이며, 인스턴스는 'TopicMaps'이다.

표 1 표준항목과 이에 대응되는 로컬 항목

	표준항목	local1
제목	title	ptitle
저자	author	pauthors
학회	society	comm
키워드	keyword	key
연구실	lab	laboratory

셋째, TMDR을 이용하여 표준 항목과 로컬 항목간의 매핑 및 인스턴스의 관계성을 추출한다. local1과 매핑된 정보는 표 1에서 제시한 것과 같고, 인스턴스인 'TopicMaps'는 그림 1에서 온톨로지의 언어 'OWL', 'RDF' 그리고 '시소러스'와 관계를 가지고 있다. 추가적으로 적용분야도 포함할 수 있다. 여기서는 온톨로지의 언어 관계와 '데이터베이스' 영역으로 확장한다.

넷째, 질의 변환 과정이다. 이때는 표준에 해당하는 로컬정보를 이용하므로 FROM절이 포함된다. 생성된 질의는 다음과 같다. 이렇게 변환된 질의는 각 로컬시스템으로 전송되어 질의를 수행한다.

```
SELECT p.ptitle, a.pauthors, c.comm
FROM paper p, paper2author pa, authors a,
community c
WHERE p.key='TopicMaps' OR p.key='OWL'
OR p.key='RDF' OR p.key='thesaurus'
AND a.part='database';
```

IV. 적용

이 장은 본 논문을 적용을 위한 테스트 구현과 실험용 데이터를 통하여 수행된 결과를 분석한다. 그림 2과 3은 사용자의 요구 사항 입력에 따른 쿼리 생성과 Topic Maps를 통하여 연관관계를 쿼리에 포함하는 과정을 보이고 있다.

그림 2는 사용자에게 제공되는 글로벌 스키마를 생성된 인터페이스이다. 이 인터페이스는 세 영역으로 분리하여 각 기능들을 수행한다. 먼저, Search item 영역이다. 이 영역은 글로벌 스키마를 제공하여 요구하는 항목을 선택한다. 이를 통해서 사용자는 로컬 시스템에 대해서 모르더라도

검색 할 수 있다. 둘째, search condition 영역이다. 이 영역은 조건을 선택하는 영역으로 조건의 항목과 값을 선택한다 이 두 영역은 글로벌 쿼리를 생성하기 위한 데이터가 된다 글로벌 쿼리는 그림 2의 오른쪽의 텍스트를 통해 확인할 수 있다. 셋째, select search site의 선택 영역으로 접근할 로컬 시스템의 선택 영역이다 이 영역은 통합 검색에 포함된 사이트들 목록이다 사용자가 검색하고자 하는 사이트들을 선택하도록 제공한다 이것은 ML을 통해 접근한다.

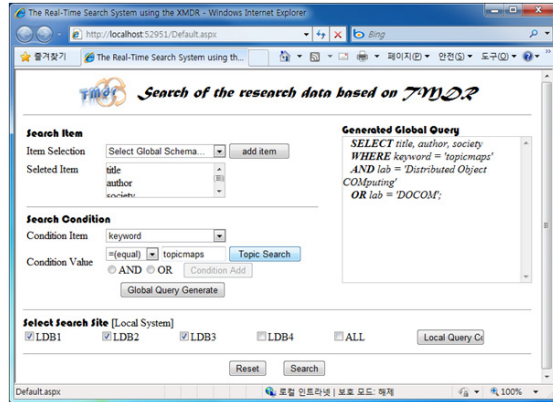


그림 2 사용자 인터페이스와 글로벌 쿼리

그림 3은 XMDR의 Topic Maps 영역으로 연관 관계를 검색하는 것이다. 이 그림은 사용자가 입력한 항목과 값에 따른 토픽의 계층 구조를 제공하고, 토픽 계층구조에서 선택한 토픽의 연관 관계를 제공함으로써 연관 관계를 쿼리에 추가한다

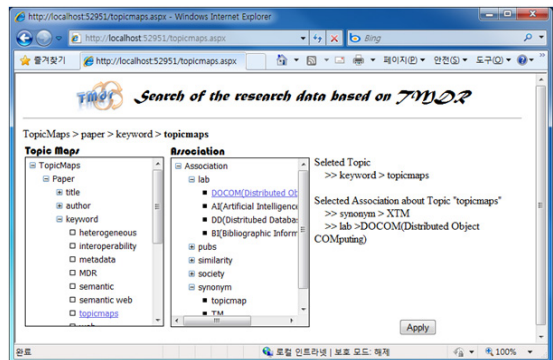


그림 3 연관관계 추가를 위한 Topic Maps

이러한 방법으로 TMDR은 각 로컬에 존재하는 데이터베이스를 수정 없이 접근할 수 있다

V. 결론

우리는 분산된 데이터베이스 통합하기 위해 발생하는 문제점으로 스키마와 데이터들이 가지고 있는 구조적, 의미적 이질성 문제를 제시하였다 이를 해결하기 위해 Topic Maps과 XMDR을 유

기적으로 결합한 지식 저장소로 TMDR을 이용하여 글로벌 쿼리를 효율적으로 로컬에 적용할 수 있는 쿼리로 변환시키는 방안을 제안하였다

이에 스키마의 이질성은 MSO를 적용하고, 인스턴스의 이질성은 Topic Maps의 association을 이용하여 해결하였다. 그리고 시스템의 접근은 ML이 담당하도록 하였다. 이러한 방식에 따라 질의 변환을 수행함으로써 사용자는 로컬에 대한 접근방법이나 스키마에 대한 지식 없이도 네트워크상에 존재하는 다양한 정보들을 접근 할 수 있으며, 다른 지식 검색 분야에 확장하여 적용할 수 있도록 일반화해야 한다.

참고문헌

- [1] Salton, G., and McGill, M.J., "Introduction to Modern Information Retrieval," McGraw-Hill, New York, 1983.
- [2] Alon Y. Halevy, Naveen Ashish, Dina Bitton, Michael Carey, Denise Draper, Jeff Pollock, Arnon Rosenthal, Vishal Sikka, "Enterprise information integration: successes, challenges and controversies", Proceedings of the 2005 ACM SIGMOD international conference on Management of data, 2005. 06.
- [3] 정계동, 황치곤, 최영근, "멀티 온톨로지 기반의 키워드 연관성을 이용한 전문가 검색 시스템," 한국정보통신학회논문지, v. 16, no. 1, pp. 183-190, 2012. 01.
- [4] Ray Gates, "Introduction to MDR-Tutorial on ISO/IEC 11179", Metadata Open Forum 2004, Xian, 2004. 5.
- [5] <http://www.isoTopicMaps.org/> ISO/IEC 13250 Topic Maps (Second Edition), 2002. 05.
- [6] Chigon Hwang, Seokjae Moon, Hyoyoung Shin, Gyedong Jung, Youngkeun Choi, "TMDR-Based Query Conversion for Data Integration in a Distributed Network," Communications in Computer and Information Science, Volume 195, pp.36-45, 2011.