# TMDR 기반의 효율적인 키워드 검색 방안에 관한 연구

정계동*, 황치곤*, 신효영**, 최영근*
*광운대학교 컴퓨터과학과
**경복대학 인터넷정보과
e-mail:duck1052@kw.ac.kr

# A Study on Efficient keyword search Scheme based on TMDR

Kye-dong Jung*, Chi-gon Hwang*, Hyo-young Shin**, Young-keun Choi*
*Dept of Computer Science, Kwang-Woon University
**Dept of Internet Information, Kyung-bok College

요 약

검색 키워드의 모호성으로 인해 검색된 정보의 질이 낮아지는 문제가 있다. 이러한 문제를 해결하기 위해 키워드의 정확성을 보강할 방안으로 온톨로지를 구성하여 지원할 수 있다. 온톨로지를 표현하기 위한 방법 중 다양한 방법들이 있으며, 그 중 수직, 수평 검색이 모두 가능한 토픽맵을 이용하는 방법이 효율적이다. 그리고 기존 웹뿐만 아니라 인명정보를 관리하는 데이터베이스 시스템들의 정보를 통합하기 위한 방안도 필요하다. 그러므로 본 시스템에서는 온톨로지 표현 기법인 토픽맵과 기존 데이터베이스 시스템 통합을 위한 패러다임인 MDR을 융합하는 방안을 제안한다.

## 1. 서론

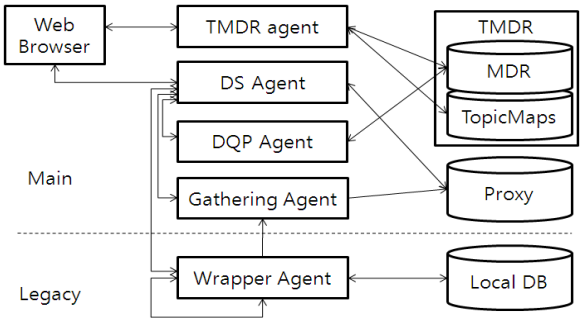
현재 수많은 정보들이 생성되고 소멸되는 과정을 계속적으로 반복하고 있다. 그리고 이러한 정보는 사용하는 사람들에 따라 큰 가치를 가지게 된다. 이때 정확한 정보뿐만 아니라 관련정보도 필요하다. 이런 관련 정보를 필요한 시기에 제공하는 것이 정보 검색의 목적이다. 그리고 이러한 정보들의 검색은 일반적으로 키워드 기반으로 수행되고 있다[1]. 이런 키워드 기반의 검색은 키워드 출현 빈도수가 높은 검색하기 때문에 관련 정보를 검색해 주지만, 검색된 정보가 실제 사용자에게 유용하지 않은 정보들도 다수 검색된다. 예를 들어 우리나라 사람들 중 흔한 성과 이름으로 조합한 "김철수"를 구글에서 검색하면 약 276,000건의 결과가 나온다. 그럼 이 결과는 모두 동일인이 아니라 "김철수"라는 이름을 가진 동명이인에 대한 모든 사람의 결과이다.

우리는 이러한 애매모호한 결과를 해결하기 위해 인명 검색을 위한 지식과 각 인명의 부가적인 정보를 연결하는 방법으로 토픽맵을 이용한다. 토픽맵은 분산 환경 하에서 지식 구조를 정의하고 정의된 구조와 지식 자원을 연계하는데 쓰이는 기술 표준이다[2]. 그리고 기존 시스템들의 데이터베이스를 통합하기 위한 방안으로 MDR은 데이터 통합을 위해 XML 기반의 관계 데이터베이스의 메타데이터를 데이터베이스에 저장하는 기술이며, 분산된 데이터의 이질성을 해결하기 위한 기술이다[3].

지식을 나타내기 위한 온톨로지 표현 기법인 토픽맵과 분산된 데이터베이스 통합을 위한 MDR을 결합한 TMDR이 있다[4]. 이것을 이용해 기존 시스템에 저장된 인명관련 정보를 효율적으로 검색할 수 있는 방안을 제안한다.

## 2. TMDR을 이용한 검색 시스템

제안하는 시스템은 (그림 1)과 같은 구조를 가지고 있고 5개 에이전트의 상호 작용을 통하여 시스템이 운용된다. 시스템에서 사용하는 저장소는 상호 작용을 위한 TMDR과 검색 결과의 저장을 위한 Proxy DB를 가진다.

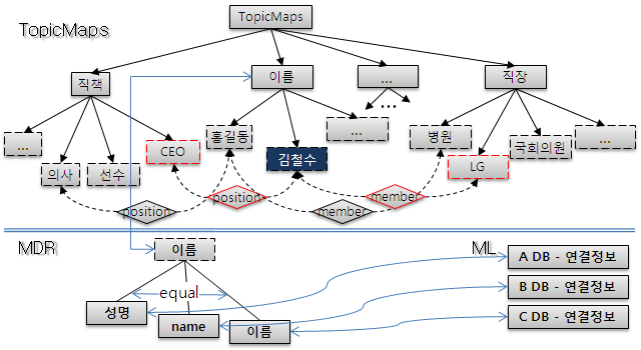



(그림 1) 제안 시스템

각 에이전트에 역할은 각각 다음과 같다. TMDR 에이전트는 TMDR에 접근하여 토픽맵의 지식과 MDR의 표준 항목의 정보를 웹브라우저에 제공한다. DS(Data Service) 에이전트는 웹브라우저나 응용프로그램에 데이터서비스를 제공하는 에이전트로 다른 에이전트들과 작업수행의 지시와 검색 항목을 통해 Proxy에 검색결과가 축적될 테이블을 생성한다. DQP(Distributed Query Process) 에이전트는 사용자가 선택한 키워드와 연관 지식을 DS 에이전트에서 넘겨받아 실제 질의를 수행하기 위한 질의를 작성한다. 각 레거시에 접근하기 위한 접근정보를 MDR의 ML(Meta Location)을 참조한다. Gathering 에이전트는 DS 에이전트의 요구에 의해 Wrapper에이전트에서 수행

한 검색결과를 받아 Proxy에 저장하는 역할과 Proxy에 축적된 정보의 내역을 DS에이전트에 전달하여 사용자 인터페이스에 제공할 수 있도록 한다. Wrapper 에이전트는 DS 에이전트에서 전송한 검색요구사항을 받아 로컬 데이터베이스에 맞게 질의를 변환하고 변환된 질의문은 수행하고 수행된 결과를 Gathering 에이전트에게 전송한다.

TMDR의 구조는 토픽맵의 지식표현 부분과 MDR을 결합한다. 토픽맵은 정보 개념 또는 정보 개체가 되는 토픽(topic), 토픽과 토픽 사이의 연관관계(association) 그리고 토픽에 대한 자원 또는 그 링크인 어커런스(occurrence)로 구성된다. 이 토픽맵을 이용하여 키워드 모호성을 해결한다. 본 시스템에 적용하기 위한 토픽맵은 어커런스 역할은 MDR의 ML을 통해서 수행할 수 있으므로 토픽과 연관관계를 통한 토픽의 연관관계를 정의하고, 연관관계를 이용한 수평적 접근이 가능하도록 하는 토픽맵이다. 제안되는 토픽맵의 사례는 (그림 2)와 같이 표현된다.

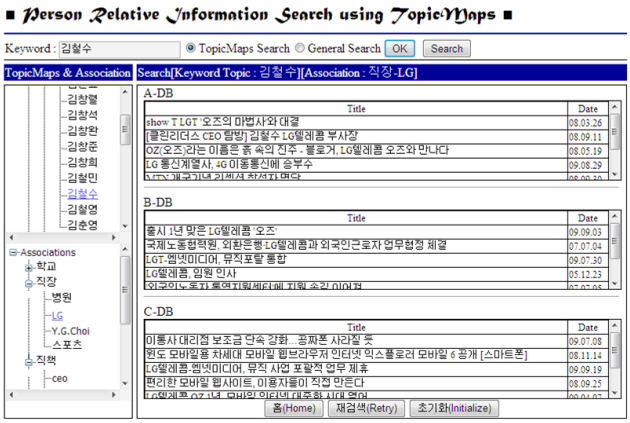



(그림 2) 토픽맵의 예

(그림 2)는 토픽맵의 루트 토픽은 TopicMaps이고, 토픽은 하위토픽을 가질 수 있다. 이때 상위토픽은 토픽클래스라 하고 실선으로 표기된 직사각형으로 표현하고, 하위토픽은 토픽 인스턴스로 점선으로 표기된 직사각형으로 표현된다. 그리고 마름모로 표기된 도형은 토픽사이의 연관관계를 표현하는 것으로 연관관계에 포함되는 토픽들에 대한 관계 역할을 가지고 있다. 표현된 토픽맵은 MDR의 표준 항목을 통하여 작성 되어 있다. MDR은 기존 시스템들의 표준을 제공하는 것으로 (그림 2)에서 보면 기존 시스템에서 사용하는 필드명이 '성명', 'name', '이름'으로 동일한 데이터임에도 불구하고 서로 다르게 표현되는데 이를 MDR의 표준인 '이름'으로 하여 서로 매핑시켜 준다. 그리고 각 시스템에 대한 연결은 ML을 이용한다. 이렇게 기존 시스템과의 이질적 문제와 접근에 관한 문제는 MDR을 이용하여 해결하고 키워드의 모호성 문제와 지식 표현은 토픽과 연관관계를 통하여 해결하게 한다.

## 3. 적용사례

2장에서 제시한 방법에 따라 테스트용 데이터베이스를 구축하여 운용한 적용사례는 (그림 3)과 같다. (그림 3)은 사용자가 검색할 사람의 이름을 키워드로 하여 입력하면 토픽맵에서 입력된 사람의 이름에 해당하는 토픽을 열어준다. 그리고 입력된 키워드는 토픽맵의 입장에서는 하나의 토픽이므로 토픽의 연관관계를 나타내주므로 단순한 이름에 의한 검색이 아닌 이름과 연관관계를 동시에 적용한 검색이 되도록 한다. (그림 3)에서 표현된 인터페이스는 서론에서 언급한 사람 막연한 사람 '김철수'를 검색한 것이 아니라 그 이름의 연관관계 중 직장이 'LG'인 것에 대해 검색한 것이다. 그에 따른 검색의 결과는 '김철수'라는 사람이 'LG'의 부사장이므로, 부사장과 관련된 정보들이 검색된 것을 확인할 수 있다. 단순히 '김철수'란 이름을 검색했을 때 보다는 훨씬 정확한 정보를 얻을 수 있다.




(그림 3) 인물검색 적용 사례

## 4. 결론

적용사례에서 확인할 수 있듯이 TMDR을 이용한 검색은 기존 키워드 검색의 단점인 키워드가 포함된 모든 정보를 검색하는 것보다 효율적인 것을 확인할 수 있다. 본 시스템에서 제안하는 TMDR은 이름을 이용한 검색뿐만 아니라 정확한 검색을 요구하는 모든 분야에서 적용가능하며, 데이터웨어하우스와 같이 방대한 정보를 효율적으로 수집하기 위한 시스템에서도 효율적이다. 그리고 본 시스템을 적용하면 검색을 위한 초기 검색속도는 일반 키워드 검색 속도보다는 약간 떨어질 수 있으나 검색된 결과의 정확성으로 인해 검색된 자료의 건수가 압축되기 때문에 실제적으로는 검색효율이 향상될 수 있다. 향후 이 시스템은 데이터웨어하우스의 ETL 분야를 위해 적용할 수 있는 방안에 대해서 연구해 보고자 한다.

## 참고문헌

[1] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, Modern information retrieval, Addison Wesley, 1999.

[2] Members of the Topic Map.org Authoring Group, "XML Topic Maps(XTM) 1.0", www.TopicMaps.org/xtm/, 2001.8.

[3] Ray Gates, "Introduction to MDR-Tutorial on ISO/IEC 11179" Metadata Open Forum 2004, Xian, May 17, 2004.

[4] 정계동, 황치곤, TMDR 기반의 실시간 데이터 통합 환경 설계, 한국해양정보통신학회, Vol.13, No.9, 2009.