한국컴퓨터정보학회 하계학술대회 논문집 제20권 제2호 (2012. 7)

# 지능형 검색 지원을 위한 관계형 데이터베이스와 온톨로지 병행 모델

서현석°, 안기홍*, 김수경*
°*한발대학교 컴퓨터공학과
e-mail: skinahitoda@gmail.com°, kimsk, khahn@{hanbat.ac.kr}

# Relational Database and Ontology parallel model for intelligent search support

Hyunseok Seo°, Kihong Anh*, Sukyoung Kim*
°*Computer Engineering, Hanbat National University

● 요 약 ●

본 연구는 사용자가 특정 전문지식에 대하여 검색하는데 있어 관계형 데이터베이스와 온톨로지를 결합해 보다 적합한 검색 결과를 반환하도록 하는 관계형 데이터베이스와 온톨로지 병행모델에 관한 것이다. 데이터나 정보 양의 급격한 증가는 검색 결과의 사용자 확신을 도리어 떨어트리는 big data 문제에 부딪히게 되었으며 모바일 기기의 사용 증가는 검색과 결과의 판단에 있어 인간의 관여를 줄이는 단순성을 높이는 것이 강조되고 있다. 따라서 본 연구는 고수준의 의사결정이 요구되는 분야에 있어서의 검색 성능을 높이기 위해 관계형 데이터베이스로 구성된 데이터에 온톨로지를 결합시켜 사용자에게 적합한 데이터를 반환할 수 있는 모델에 대해 지원해보고자 한다. 본 연구의 검증을 위해 전문 지식이 요구되는 의약품 분야의 데이터베이스를 기준으로 서비스를 제공하는 사이트에서의 검색을 통해 문제점을 제시하고 연구의 필요성을 제시한다.

**키워드:** 지식베이스(knowledge-base), bigdata, 관계형 데이터베이스(Relational Database), 온톨로지(Ontology)

## I. 서 론

### 1.1 연구의 필요성(introduction)

인터넷을 통한 데이터 검색을 하는데 있어서 사용자에게 가장 중요한 것은, 검색된 결과가 사용자에게 적합한 데이터를 보다 접근하기 쉽게 찾아주는 것이다. 웹 상에 산재된 비정형 데이터를 통해 정보 검색을 하는데 있어서 반환되는 데이터는 정확률과 재현율을 기반으로 처리된다. 여기서 정확률은 얼마나 사용자의 요구에 적합한 문서가 검색 반환 되었는가를 통해 판단이 되며, 이러한 정확률은 문서상에 나타나는 사용자의 검색질문에 따른 적합한 키워드를 얼마나 포함하고 있는가를 바탕으로 주로 결정된다. 재현율은 요구되는 정보에 얼마만큼 요구하는 정보와 일치하는지를 바탕으로 나타나며, 정보검색을 하는데 위의 두 요소가 100%로 충족되었을 때 최상의 검색 결과를 사용자에게 반환할 수 있을 것이다. 하지만 두 요소는 상반되는 관계로 100%로 충족되는 경우는 없다.

본 연구는 인터넷에 산재되어있는 데이터에 대해 어떻게 보다 사용자에게 적합한 검색 결과를 반환할 수 있는지를 목표로 하며, 이러한 문제는 인터넷에 나타는 폭발적인 데이터의 증가로 인해 더욱 중요한 요소로 나타나고 있다. 본 논문에서는 위와 같은 문제를 개선시키는데 있어 데이터마이닝과 온톨로지를 통한 지식베이스 구축이 어떠한 역할을 하는지에 대해서 다룬다.

### 1.2 연구의 목적

산재되어 있는 데이터에 대해서 필요한 목적에 따라 데이터를 분류한 지식베이스를 구축하면, 무의미하게 산재된 데이터보다 나은 검색 결과를 반환해 줄 수 있을 것이라는 가정을 바탕으로 본 연구는 시작한다. 전문지식에 해당하는 의약품으로 지식베이스를 초점으로 맞췄다. 이러한 지식베이스를 구축하는데 있어서 데이터마이닝 기술이 필수적인데, 여기서 데이터마이닝이란 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아주는 것을 의미한다. 또한 이러한 지식 베이스에 의미를 부여하는 온톨로지를 통해 기존의 단순 키워드를 통한 검색보다 나은 의미기반 검색에 대해서 제시해보고자 한다.

## II. 본 론

### 2.1 온톨로지 스키마 모델 디자인

온톨로지를 이용하면 데이터 간에 관계를 정의하는데 있어서 수많은 정보를 연결할 수 있다. 모델이 구축되면 데이터를 통한 검색 결과를 반환해주는데 있어서 관계를 통한 검색이 이루어질 수

있게 된다. 이러한 결과를 제공하기 위해서는 명시적인 온톨로지 모델을 설계해줘야 하며 명시적인 온톨로지 모델을 고려할 때 기준은 다음과 같다.

(1) 설계된 온톨로지 모델의 표현 범위
(2) 온톨로지를 통해 제공할 명확하거나 암시적인 추론 범위
(3) 웹 또는 만연한 데이터들과 같은 오래된 데이터에 관한 인프라를 어떻게 통합할 것인가
(4) 온톨로지 표현 언어 및 추론엔진의 성능
(5) 데이터 사용을 위한 접근 및 변경에 대한 온톨로지 정보에 대한 실시간 접근성

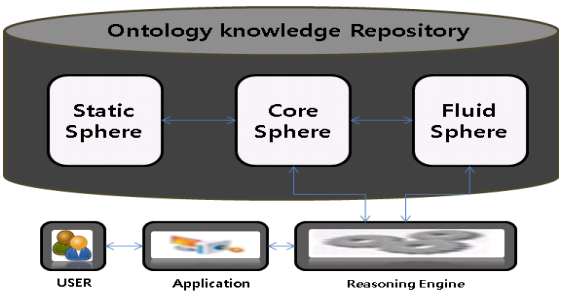


그림 1. 온톨로지 저장소 구조 및 레이아웃

위 기준에 따르면 온토롤지 저장소에 모델이 되는 스키마는 Static Sphere 영역에 저장되어 인스턴스의 수정, 추가와 관계없이 관리되고 사용자와 증상등에 관련되는 data는 Core Sphere 영역에 저장되 사용자에 관련된 동적인 상태 값들은 Fluid Sphere 영역에 저장되어 각각을 담당한다. 즉 질의 추론이 시작되는 Static Sphere에 있는 스키마를 통해 검색될 영역과 인스턴스의 속성들이 결정되고 이를 통해 Core Sphere와 Fluid Sphere영역의 인스턴스를 검색하게 된다.

본 연구에서 제시하는 기준은 지식베이스로 하는 의약품에 대한 데이터를 얼마나 온톨로지를 통해 적합하게 사용할 수 있는가를 기준으로 한다. 이를 통해 온톨로지 구조, 서비스 범위, 확장성을 결정하는데 사용한다. 그림 1의 경우 온톨로지 모델 프레임워크로 이를 기준으로 이를 바탕으로 상세 설계 모델을 구성하는데 참고하고자 한다.

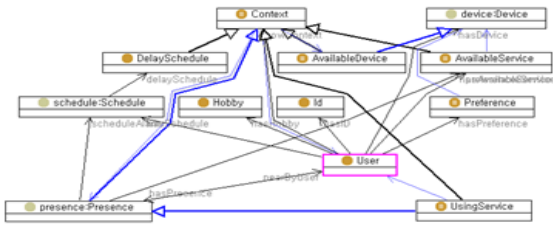


그림 2. 온톨로지에 관한 클래스 다이어그램

## 2.2 약품 지식 베이스 모델을 통한 대량으로 구축된 정보의 신뢰성 검증 모델

그림 3과 같이 전체적인 설계모델에 앞서 본 논문에서는 ontology와, Database에 주안점을 두고 모델을 설계하였다. 그 중에서도 중요한 모델 설계는 ontology로 초점을 맞추고 있는데, 사용자에게 데이터를 반환해주는데 있어서 가장 중요한 역할을 하고 있기 때문이다.

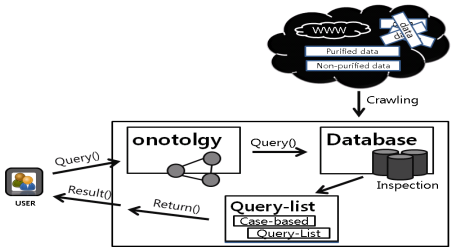


그림 3. 전체적인 설계 구조

나아가서 지식베이스는 의약품에 초점을 맞추었는데, 연구 제목에서도 나타나듯이 사용자를 위한 고수준 정보 검색을 지원하고자 한다. 이러한 연구 배경에 맞추어 데이터 모델을 의약품으로 지식베이스를 구축하도록 계획한 이유는 일반인이 얼마나 전문지식에 어려움 없이 접근할 수 있는가를 위한 이유이며, 나아가서 의약품을 검색하는데 있어 단순히 의약품의 모델을 통한 검색보다는 증상 및 효능에 따른 검색을 통하도록 하여, 자연어 특성에 맞는 언어간 관계 모델링을 통해 추론 엔진을 적용하기에 적합하다는 이유(그림-2)에서 선정하였다.

표 1. 의약품에 기재된 효능 및 이에 대한 유사어

| 의약품에 기재되어있는 효능 | 유사어(문장) |
|---|---|
| 감기 | 고뿔, 기침, 재치기 |
| 두통 | 머리 |

## 2.3 검증된 지식베이스를 통한 데이터베이스 모델

추론을 위한 추론 엔진을 만들기에 앞서서 선행되어야 하는 작업으로는 데이터의 수집과 이를 바탕으로 데이들 사이의 관계 모델링을 필요로 하는데, 이는 의약품 데이터 수집을 통한 관계형 데이터베이스를 구축하는 것으로부터 시작된다.

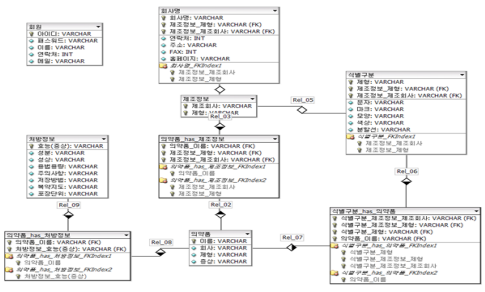

그림 4. 의약품 기반 관계형 데이터 베이스 모델

그림 4의 데이터 베이스 모델링에서 알 수 있듯이 의약품을 하나 검색하기 위한 키워드는 다양하게 존재하지만, 일반 사용자가 정확하게 해당 데이터에 대해서 파악하고 검색하는데는 어려움이 따른다, 이를 고려하여 의약품 하나의 모델을 ①의약품, ②처방정보, ③식별구분으로 나누었지만, 사용자가 해당 데이터에 어떤 값이 들어있나 알아야한다는 전제조건이 해결되는 것은 아니다. 이러한 문제는 사용자가 데이터에 접근하는데 있어서 시작부터 접근성을 떨어뜨리며, 이러한 문제는 데이터에 따른 적합한 온톨로지 모델링을 통해 극복이 가능한 문제다.

## 2.4 일반 자식베이스 모델의 추론 실험

의약품을 검색하는데 있어서 사용자는 다음과 같은 상황에 있을 수 있는데, ① 자신의 병명을 정확하게 안다. ② 자신의 증상은 알지만 병명은 모른다. ③ 자신의 병명과 증상을 전혀 모르지만 몸이 어떤 상태인지 안다. 3가지 상황으로 분류하였을 때 자신이 복용하는 약에 대한 정보를 알고 싶을 때 각 분류 별로 필요한 정보는 다음과 같다(표-2).

표 2. 검색 방법에 따른 상황 분류

| 상황 | 필요한정보 |
|---|---|
| ① | 병명에 대한 키워드 |
| ② | 증상에 대한 키워드 |
| ③ | 증상에 대한 설명 |

①, ②상황에 대한 원하는 정보검색 결과는 키워드 입력을 통해 쉽게 얻어낼 수 있지만, ③의 경우에는 문장이 필요하게 되고 이를 통한 자연어처리를 통한 키워드 도출이 필요하다. 이러한 상황에 대해 한국 약학정보원에서 위의 상황을 가정하고 검색을 해본 결과는 다음과 같다. ①의 경우에는 검색한 결과에 대해 정보 검색이 가능하였다(그림 5).


그림 5. ① 의 상황에 대해 '감기' 를 검색한 결과

감기의 증상인 기침에 대해서 입력할 경우 이러한 증상에 대해서 검색 결과 또한 ①의 상황과 같이 잘 이루어 졌다(그림 6)


그림 6. ② 의 상황에 대한 '기침' 을 검색한 결과

①, ② 상황에 대해서는 검색이 원활하게 잘 이루어 졌지만, ③의 상황에서는 사용자가 자신이 어떤 약을 복용해야하는지에 대해 원하는 결과를 찾고자 할 때, 사용자는 자신의 증상을 설명할 수 있다. 하지만 증상에 대한 결과로는 문장, 또는 주요 증상에 대한 설명이 있는데, 그림 7에서 알 수 있듯이, 예로 ①의 상황을 문장으로 표현하거나, 단순히 속이 좋지 않은 상태를 나타내는 말로 '매스꺼움'을 입력했을 때 의약품에 대한 정보는 물론 결과조차 검색되지 않음을 알 수 있다.


그림 7. 상황 '①' 을 문장으로 표현했을 때 결과

## III. 결 론

위와 같은 경우 만약 온톨로지 모델링을 통해 적용하여, 사용자가 검색을 하는데 있어 단순 키워드가 아닌 의미가 부여된 문장에 대한 검색을 할 때 더 적절한 데이터를 반환할 수 있을 것이다. 이러한 결과는 데이터가 많아질수록 더 효율적인 해결방법이 될 수 있을꺼라 예상한다. 추후 본 연구의 지속을 위해 의약품에 적합한 온톨로지 모델링을 통한 적용을 통해 이를 검증하고자 한다.

## 사 사

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구 사업입니다.(No.20120004360)

## 참고문헌

[1] Dean Allemang, "Semantic Web for Working Ontologist", Schtec Media, pp3-10,

[2] SPICE Ontology Definition of User Profiles, Knowledge Information and Services pp.3-5, 2006.

[3] MobiLife, http://www.ist-mobilife.org,2008.

[4] Chantal Taconet, Zakia Kazi-Aoui, "Context-awareness and Model Driven Engineering", E-Commerce application scenario. ICDIM 2008.

[5] Bright, M.W., Hurson, A.R., Pakzad, S.H. "Automated resolution of semantic terogeneity in multidatabases", ACM Transaction on Database Systems 19, pp.212-253, 1994.

[6] Guarino, N., Masolo, C., Verete, G. “Ontoseek: Content-based access to the web”, IEEE Intelligent Systems 3, pp.70-80, 199

[7] Kim su-kyoung*, Ahn ki-hong** A Study of Cyber Medicine Guider based on Smart Phone using Medicine Semantic Social Network and Image Matching