

한의 온톨로지 기반 시맨틱 검색 시스템

Semantic Search System based on Korean Medicine Ontology

김상균, 박동훈, 김안나, 오용택, 김지영, 예상준, 김철, 장현철
한국한의학연구원 정보개발운영그룹

Sang-Kyun Kim(skkim@kiom.re.kr), Dong-Hun Park(pd23@kiom.re.kr),
AnNa Kim(ankim2012@kiom.re.kr), Yong-Taek Oh(ydydxor@kiom.re.kr),
Ji-Young Kim(helloqq@kiom.re.kr), Sang-Jun Yea(tomita@kiom.re.kr),
Chul Kim(chulnice@kiom.re.kr), Hyun Chul Jang(hcjang@kiom.re.kr)

요약

본 연구에서는 한의 온톨로지를 이용한 시맨틱 검색 시스템을 제안한다. 시맨틱 검색은 사용자의 검색어가 어떤 개념에 속하는지 인식하고 온톨로지에서 검색어와 연관된 정보들을 보여주는 기능을 제공하는 장점이 있다. 본 연구에서 구현한 시맨틱 검색 시스템 또한 이와 같은 시맨틱 검색 기능을 구현하였다. 뿐만 아니라 한의 온톨로지의 시맨틱을 분석해서 한의학적으로 의미 있는 검색 시나리오를 정의하고 이 시나리오를 수행하는 프로세스 및 사용자 인터페이스를 구현하였다. 이를 통해 단순히 온톨로지에서 연결된 정보를 검색해주는 것 뿐만 아니라 보다 의미 있는 검색 결과를 제공함으로써 한의 정보를 검색하는 사용자들이 유용한 검색 결과를 얻을 수 있도록 하였다.

■ 중심어 : | 한의학 | 온톨로지 | 시맨틱 검색 | 시맨틱 분석 |

Abstract

We in this paper propose a semantic search system based on Korean medicine ontology. Semantic search augments search results and improves search accuracy by understanding which concept denotes terms which users is trying to find. Our semantic search system also provides these semantic search capabilities. Moreover, search scenarios which is meaningful in Korean medicine are designed and implemented by analyzing the semantics of Korean medicine ontology. Therefore, our system can help users find the useful search results with respect to Korean medicine by providing the more meaningful information as well as the connected information in ontology.

■ keyword : | Korean Medicine | Ontology | Semantic Search | Semantic Analysis |

I. 서론

최근 차세대 웹으로써 시맨틱웹이 각광을 받고 있다. 시맨틱웹은 인터넷에 존재하는 리소스와 리소스들간의

관계에 대한 의미 정보를 컴퓨터가 이해할 수 있는 온톨로지로 표현함으로써 지식을 공유하고 재사용할 수 있는 프레임워크를 제공한다. 시맨틱웹에 대한 연구와 시맨틱웹 기술을 적용한 사이트는 계속적으로 증가하

* 본 연구는 한국한의학연구원 주요사업 "전통의학 지식정보 활용 시스템 개발(K12093)"의 지원을 받아 수행되었습니다.

접수번호 : #121004-002

접수일자 : 2012년 10월 04일

심사완료일 : 2012년 12월 10일

교신저자 : 장현철, e-mail : hcjang@kiom.re.kr

고 있으며, 또한 온톨로지들[1]간에 상호 연결을 통해 온톨로지의 공유 및 활용을 위한 노력들이 계속되고 있다. 온톨로지를 등록하고 공유하는 대표적인 사이트는 the Data Hub[2]가 있다. 이 사이트에는 누구든지 자신의 온톨로지를 등록할 수 있도록 오픈되어 있으며, 2012년 9월 현재 4천여 개의 온톨로지 및 데이터 셋이 등록되어 있다. 또한 다른 사람이 등록한 온톨로지를 검색하고 이용할 수 있는데, 대부분 RDF/XML 형태의 링크나 SPARQL 질의로 검색하게 되어 있다. 이는 시맨틱웹이 본래 기계가 이해하고 자동으로 처리 가능하도록 만들어졌기 때문이다. 하지만 사람들은 RDF 형태의 온톨로지보다는 검색을 통해서 정보를 얻기를 원한다. 특히 시맨틱웹에서의 검색 애플리케이션을 시맨틱 검색[3]이라고 한다. 구글과 같이 정보 검색 기술을 이용하는 일반 검색은 웹 상의 문서들을 색인한 후 사용자가 검색어를 입력하면 검색어와 일치하는 문서들을 찾아서 보여준다. 반면에 시맨틱 검색은 웹 데이터를 온톨로지화 구축하고 사용자가 검색어를 입력하면 우선 검색어가 무엇인지를 해석하고 검색어와 관련된 정보들을 보여준다. 예를 들어 “인삼”이라는 검색어를 입력하면 일반 검색에서는 인삼이 포함된 문서들 예를 들어 위키피디아의 인삼, 한국인삼공사등의 웹사이트가 검색된다. 반면에 시맨틱 검색에서는 인삼이 약재라는 것을 인식하고, 인삼의 학명, 효능 등 인삼과 관련된 정보를 보여준다. 이와 같이 시맨틱 검색은 일반 검색에 비해서 사용자가 원하는 것을 보다 정확하게 찾아서 연결된 관련 정보들을 보여주는 장점을 가진다.

본 연구에서는 Jang[4]이 구축한 한의 온톨로지를 이용한 시맨틱 검색 시스템을 제안한다. Jang의 논문에서 언급한 것과 같이 한의학 개념 및 용어가 온톨로지화 모델링되었으나 이를 해석하기 위해서는 한의학에 대한 이해가 필요하다. 본 연구의 시맨틱 검색은 위에서 일반 시맨틱 검색처럼 검색어가 어떤 개념에 속하는지 인식하고 온톨로지화에서 검색어와 연관된 정보들을 보여주는 기능을 제공한다. 뿐만 아니라 한의학의 내재된 특성을 분석하고 해석하여 유용한 결과를 사용자들에게 제공할 수 있도록 하였다.

II. 기존 연구

국내외에서는 시맨틱 검색에 대해서 그 동안 많은 연구와 시스템들이 제안되고 구축되었다.

Mäkelä[5]는 다양한 논문 데이터베이스에서 “semantic AND search”라는 키워드로 검색해서 20개의 시맨틱 검색 연구를 찾았으며, 각 연구들을 분석하고 시맨틱 검색 방법들의 특징들을 분석하였다. 특히 이 논문에서는 시맨틱 검색 연구들에서 사용하는 시맨틱 검색 방법들을 다음과 같이 5가지로 정리하였다.

- RDF path traversal : RDF가 그래프 모델이기 때문에 RDF 그래프를 순회하면서 검색
- Keyword와 concept 매핑 : 입력된 검색 키워드가 어떤 개념에 속하는지 매핑 (자어어 문장 매핑 포함)
- Graph Patterns : SPARQL과 같이 RDF 트리플 셋 단위로 검색하는 방법
- Logics : OWL의 기술로직 또는 다른 규칙 언어를 이용해 추론하는 방법
- Fuzzy concepts, fuzzy relations, fuzzy logics : 온톨로지를 fuzzy annotation 방법으로 구축하고 fuzzy 검색을 이용하는 방법

Sujatha et al[6]은 Powerset을 비롯해서 8개의 시맨틱 검색 엔진을 조사하고 각 엔진의 특징을 분석하였다. 시맨틱 검색 엔진들은 사용자 질의에 대해서 빠르고 정확한 검색을 제공하며, 단순한 키워드 매칭보다는 정확한 데이터와 관련된 정보를 제공하는 장점을 가진다고 요약하였다.

Chen[7]은 중의학 온톨로지(TCM : Traditional Chinese Medicine)[8]를 기반으로 시맨틱 e-Science 시스템을 구축하였다. 중의학 온톨로지는 본 논문의 한의학 온톨로지와 유사하게 중국의 약재, 처방, 병증 등의 전통의학 지식을 포함하고 있다. e-Science 시스템은 온톨로지 스키마 매핑, 시맨틱 검색, SPARQL 인터페이스, 서비스 발견 및 통합 등의 모듈을 가지고 있는 통합 시스템이다. 이 중 시맨틱 검색 모듈에서는 구글과 비슷한 검색 인터페이스를 이용해서 검색어에 대한 단순한 온

톨로지 검색 기능을 제공하고 있다.

국내의 경우 상용 시스템으로는 네이트 시맨틱 검색 [9]과 솔트룩스의 IN2 플랫폼[10]이 존재한다. 네이트의 시맨틱 검색은 사용자가 검색어를 입력하면 검색어에 대한 검색 결과뿐만 아니라 검색어와 연관된 키워드를 보여준다. 연관 키워드는 단순히 검색어와 유사한 키워드가 아니라 의미적으로 연결된 키워드들을 제시함으로써 시맨틱한 검색 결과를 제공하고 있다. 솔트룩스의 IN2 플랫폼은 정보 검색, 자연어 처리, 텍스트 마이닝 등의 기술을 기반으로 구축된 상용 시맨틱 검색 플랫폼이다. 또한 시맨틱 검색과 관련해서 몇몇 연구들이 제안되었다. KISTI에서는 온톨로지를 저장, 검색하고 추천하는 기능을 제공하는 시스템인 OntoFrame[11]을 개발하였다. 특히 OntoFrame은 대용량 온톨로지에 대해서도 효율적인 성능을 보이고 있으며, 최근에는 시맨틱 서비스 조합[12] 기능도 제공하고 있다. 유[13]는 온톨로지 기반의 호텔 검색 시스템을 제안하였다. 이 시스템에서는 호텔과 관련된 지식을 정형화해서 호텔 온톨로지로 구축하였으며, 호텔 이용자들에게 시맨틱 검색을 제공하기 위해서 다양한 추론 규칙을 SWRL로 구축하였다. 이를 통해 여행자들이 자신의 기호에 적합한 호텔을 검색할 수 있는 시맨틱 호텔 검색 시스템을 구축하였다. 또한 송[14]과 오[15]는 개인 정보를 이용한 시맨틱 검색 시스템을 제안하였다. 송의 연구에서는 개인 정보를 RDF 프로파일의 형태로 구축하고 갱신하는 방법을 제안하였으며 이렇게 구축된 개인화 프로파일을 기반으로 콘텐츠를 추천 및 검색해주는 시맨틱 검색 시스템을 구축하였다. 오의 연구에서는 개인의 모바일 라이프 로그가 기록되는 시맨틱 네트워크에서 연관 정보를 검색하는 방법을 제안하였다. 특히 시맨틱 네트워크를 시각화하고 시각화된 네트워크에서 선택적인 연관 검색을 수행할 수 있도록 하였다.

이와 같이 시맨틱 검색은 국내외에서 다양하게 연구되고 있으며, 다양한 도메인 지식에 대해서 구축되고 활용되고 있다. 하지만 현재까지 한의학 도메인에 대해서 구축된 연구나 시스템은 존재하지 않는다. 한의학 온톨로지와 유사한 중의학 온톨로지 및 시스템이 존재하긴 하지만 논문으로만 나와 있을 뿐 온톨로지가 공개

되어 있지 않고 시스템도 접속이 안 되어 활용이 불가능하다. 지식 도메인의 차이는 단순히 데이터가 다른 측면도 있지만 도메인 지식의 특성이 다르기 때문에 동일한 검색이나 추론 방법을 적용할 수 없다. 또한 도메인 사용자가 원하는 서비스 시나리오가 다르다. 예를 들어 한의학 도메인에서 한의사들은 환자가 호소하는 여러 증상들이 어떤 병증의 증상들인지 검색하는 것을 원한다. 또한 환자의 병을 치료하기 위해서 어떤 처방과 어떤 약제들이 적당한지 지능적으로 검색해줄기를 원한다. 전자의 경우 여러 개의 증상을 입력하고 검색하면 증상들을 치료하는 약제나 처방보다는 증상을 많이 수반하는 병증들을 보여주어야 한다. 증상은 병증뿐만 아니라 약제나 처방하고도 연결이 되어 있지만 실제 임상에서는 크게 의미가 없는 결과이기 때문이다. 후자의 경우 여러 가지 검색 시나리오가 존재할 수 있다. 우선 병증 검색의 경우와 유사하게 여러 약제들을 입력했을 때 약제들을 포함하는 처방을 검색할 수 있다. 이 경우에도 약제들과 치료 관계로 연결된 병증들이 존재하지만 임상에서는 거의 사용되지 않는 지식들이다. 또한 한의학에서는 약대(처방에서 주로 같이 사용되는 약제들에 대한 분석 정보)라는 것이 존재하는데 환자 처방 시 이 약대 정보를 고려해서 처방을 구성할 수 있다.

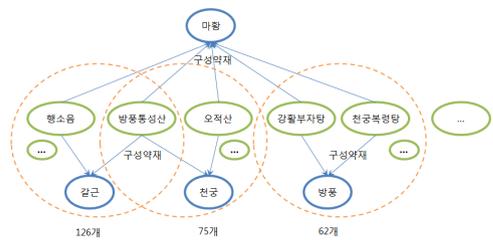


그림 1. 마황에 대한 약대 검색 예제

[그림 1]은 마황의 약대 정보 중의 일부를 그림으로 표현한 것이다. 본 연구에서 구축한 시맨틱 검색 시스템에서는 마황과 약대 관계인 약제가 갈근, 천궁, 방풍 등으로 검색되며 마황과 갈근은 126개의 처방에서, 마황과 천궁은 75개의 처방에서 마황과 방풍은 62개의 처방에서 같이 사용된 것으로 검색되었다.

이와 같이 한의학 도메인의 사용자들은 일반적인 시

맨틱 검색뿐만 아니라 한의학 도메인에서 의미가 있는 검색 시나리오를 요구한다. 다른 시맨틱 검색 시스템들은 한의학 온톨로지에 대해서 일반적인 시맨틱 검색 기능을 제공할 수 있지만 위와 같은 시맨틱 검색을 제공할 수는 없다.

III. 한의 시맨틱 검색 시스템

1. 시스템 구조

본 연구에서 구축한 한의 시맨틱 검색 시스템[16]의 시스템 구조는 [그림 2]와 같다. 온톨로지는 Jena TDB v0.8.10[17]에서 관리된다. Jena TDB는 시맨틱 웹 애플리케이션을 위한 자바 프레임워크인 Jena를 기반으로 RDF를 저장하고 질의하기 위한 방법을 제공한다. 한의 시맨틱 검색 시스템은 스프링 프레임워크 v3.0.5를 기반으로 하고 있다. 사용자 질의가 입력되면 검색어에 대한 온톨로지 클래스를 매핑시키고 SPARQL을 생성해서 Jena TDB에 질의를 수행한다. 검색된 결과는 결

과 생성기에서 처리되며 일부 정보는 구글 차트를 통해 차트로 보여진다. 또한 약재의 경우 약재 이미지 파일이 존재하며 약재 검색시 같이 보여준다.

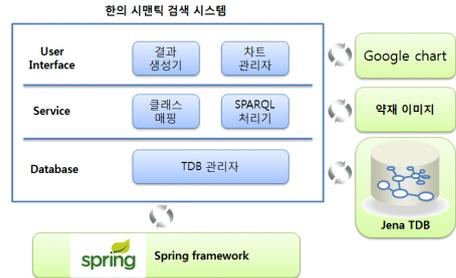
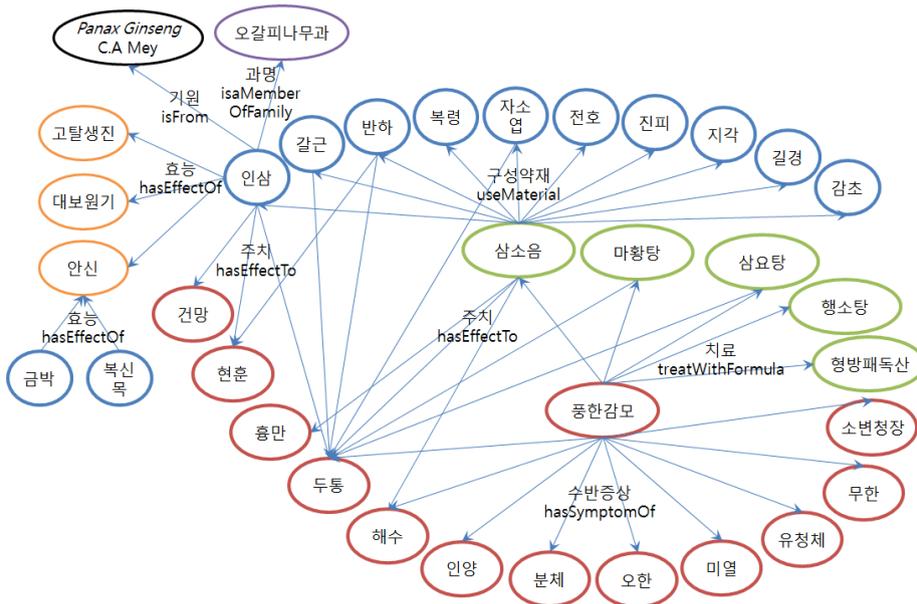


그림 2. 한의 시맨틱 검색 시스템 구조

2. 시스템 특징

본 연구에서 구축한 시맨틱 검색 시스템은 한의학의 내재된 의미를 해석해서 유용한 검색 결과를 제공할 수 있도록 하였다. 이러한 목적을 위해서 본 연구의 시스템은 다음과 같이 4가지 특징을 가진다.



(기원명을 제외하고 인스턴스의 한글명만 표시함. 프로퍼티명은 영문을 사용하며 프로퍼티의 한글명은 영문을 해석한 이름임. 붉은 타원은 병증과 증상, 초록 타원은 처방, 파란 타원은 약재, 주황 타원은 효능, 보라색 타원은 과명, 검은색 타원은 기원 클래스를 표시함)

그림 3. 한의 온톨로지 RDF 그래프 예제

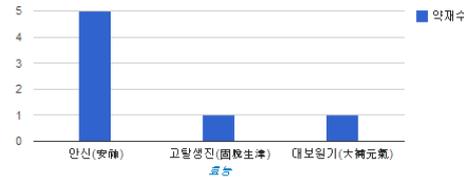
약재 >	인삼(人蔘)								
연관 처방 >									
연관 병증 >									
	<p>설명 두릅나무과(오가과 五加科 : Araliaceae)에 속한 다년생초본(多年生草本)인 인삼(人蔘) Panax ginseng C. A. Meyer.의 뿌리로 재식(栽植) 4~6년 후 가을에 줄기와 잎이 마를 때 캐어 새긴한 것을 인삼(人蔘)이라 하고 수근(鬚根)을 제거하고 쪼서 건조(乾燥)한 것을 홍삼(紅蔘)이라 함</p> <p>기원명 Panax ginseng C.A.Meyer [NCBI Link]</p> <p>과명 오갈피나물과(Araliaceae)</p> <p>분류 보기약(補氣藥)</p> <p>귀경 비(脾), 폐(肺), 심(心)</p> <p>성 미온(微溫)</p> <p>미 감(甘), 미고(微苦)</p> <p>이명 元蔘, 黃石, 地精, 人蔘, 血蔘, 遼蔘, 人蔘, 金井玉蘭, 潯州人蔘, 人蔘, 蔘, 清河人蔘, 土蔘, 百濟蔘, 白蔘蔘, 靑蔘, 人身, 羊角蔘, 鞭蔘丹, 獨勝, 孩兒蔘, 新羅人蔘, 上黨人蔘, 羅蔘, 大人蔘, 上黨蔘, 鞭蔘, 遼丹, 紫團蔘, 白物, 遼東蔘, 海蔘, 玉蔘, 仁蔘, 白蔘蔘, 黃蔘, 人蔘, 人蔘</p> <p>약용부위 뿌리, 根, 根莖</p> <p>효능 안신(安神), 고탈생진(固脫生津), 대보원기(大補元氣)</p>								
	<p>효능별 유사 약재</p>  <table border="1"> <caption>효능별 유사 약재</caption> <thead> <tr> <th>효능</th> <th>약재수</th> </tr> </thead> <tbody> <tr> <td>안신(安神)</td> <td>5</td> </tr> <tr> <td>고탈생진(固脫生津)</td> <td>1</td> </tr> <tr> <td>대보원기(大補元氣)</td> <td>1</td> </tr> </tbody> </table> <p>안신(安神) : 인삼(人蔘), 복신목(茯神木), 금박(金箔), 합환화(合歡花), 영사(靈砂) 고탈생진(固脫生津) : 인삼(人蔘) 대보원기(大補元氣) : 인삼(人蔘)</p>	효능	약재수	안신(安神)	5	고탈생진(固脫生津)	1	대보원기(大補元氣)	1
효능	약재수								
안신(安神)	5								
고탈생진(固脫生津)	1								
대보원기(大補元氣)	1								

그림 4. “인삼” 검색어에 대한 검색 결과

첫째, 한의 온톨로지의 시맨틱을 분석해서 의미 있는 검색 결과를 제공한다. 예를 들어 한의 온톨로지의 약재 클래스는 기원, 과명, 효능, 주치(치료병증)등의 프로퍼티를 가지는데, 약재 인삼을 검색했을 때 인삼의 기원, 과명, 효능, 주치(치료병증)들만 보여주는 것이 아니라 인삼의 효능 또는 주치를 가지는 다른 약재 또는 처방들도 추가적으로 보여줌으로써 인삼과 유사한 약재나 처방을 알 수 있게 한다. [그림 3]과 [그림 4]에서 보면 인삼의 효능인 안신, 고탈생진, 대보원기 그리고 주치인 건망, 현훈, 두통 등을 보여줄 뿐만 아니라 금박, 복신목등도 안신의 효능을 가지며 갈근, 반하, 자소엽등도 두통의 주치를 가지는 것을 알 수 있다. 이러한 시맨틱 검색 기능은 4장의 한의 온톨로지 시맨틱 분석에서 상세하게 기술한다.

둘째, 사용자가 검색어를 입력하면 온톨로지에서 검

색어를 포함하는 인스턴스 이름들을 추천해준다. 이는 일반 검색엔진의 검색어 추천과 유사한 기능이지만, 단순히 많이 사용되는 검색어가 아니라 온톨로지에 존재하는 인스턴스를 추천해주는 점이 다르다.

셋째, 검색어가 온톨로지의 어떤 개념(클래스)에 속하는지를 분석하고 이에 따라 검색 결과를 보여준다. 즉, 검색어가 온톨로지 인스턴스와 일치하는 경우, 예를 들어 [그림 4]와 같이 인삼을 검색하면 인삼이 약재라는 것을 인식하고 인삼에 대한 상세정보를 바로 보여준다. 이는 일반 검색과 다른 시맨틱 검색 기능으로써 사용자가 원하는 정보를 정확하게 검색해서 연관된 정보를 같이 보여줄 수 있다. 하지만 병증 클래스의 두통과 증상 클래스의 두통이 존재하는 경우와 같이 두 개의 검색 결과가 존재하는 경우에는 사용자가 어떤 것을 검색할지 선택하도록 하였다.

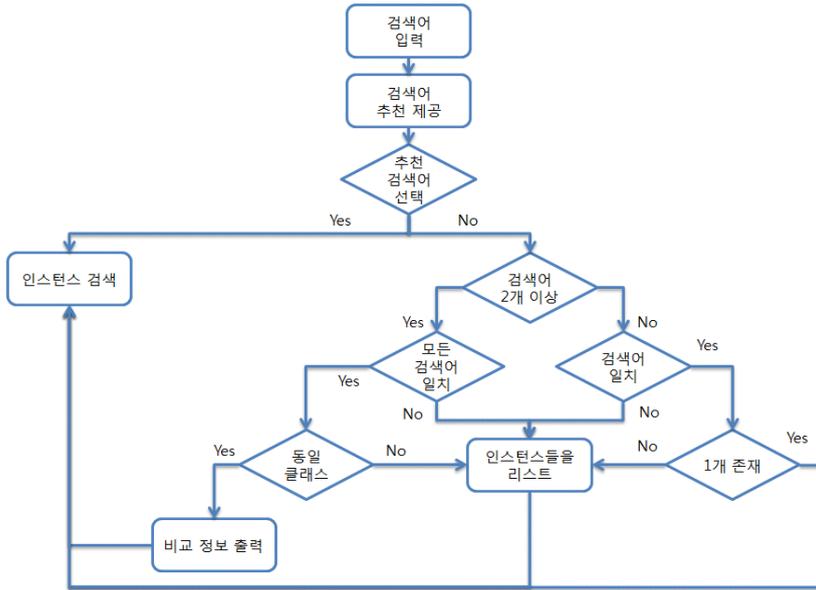


그림 5. 시맨틱 검색 프로세스

넷째, 검색어가 여러 개이고 모두 같은 클래스의 인스턴스인 경우 인스턴스의 각 필드를 테이블 형태로 비교한 결과를 보여준다. 일반 검색에서는 검색어가 여러 개인 경우 검색어가 동시에 나오거나 많이 나오는 문서들을 보여주지만, 시맨틱 검색에서는 정확한 검색 결과를 얻을 수 있기 때문에 인스턴스 검색 결과에서 각 항목들을 비교해서 검색이 가능하다.

를 보여준다.

3. 검색 프로세스 및 인터페이스

한 의 시맨틱 검색 시스템의 검색 프로세스는 [그림 5]와 같다. 사용자가 검색어를 입력하면 우선 검색어를 포함하는 인스턴스 이름이 온톨로지에 존재하는지 검사한다. 만약 존재하면 [그림 6]과 같이 인스턴스의 한글명(한자명)[출전], 클래스의 형태로 최대 15개까지 검색어로 추천해준다.

사용자가 추천된 인스턴스를 선택하면 바로 해당 인스턴스를 검색해서 상세 정보를 보여주며, 아닌 경우 입력된 검색어가 공백을 구분자로 해서 몇 개가 입력되었는지 체크한다. 만약 검색어가 1개인 경우 검색어가 온톨로지의 인스턴스명과 일치하고 정확히 한 개만 존재하는 경우에는 해당 인스턴스를 검색해서 상세 정보

인삼		검색
인삼(人蔘) - 약재		
인삼강활산(人蔘羌活散) - 처방		
인삼강활산(人蔘羌活散)[동의방제와처방해설] - 처방		
인삼강활산(人蔘羌活散)[東醫寶鑑] - 처방		
인삼강활산(人蔘羌活散)[東醫寶鑑] - 처방		
인삼강활산(人蔘羌活散)[東醫寶鑑] - 처방		
인삼강활산(人蔘羌活散)[東醫寶鑑] - 처방		
인삼강활산(人蔘羌活散)[東醫寶鑑] - 처방		
인삼개위탕(人蔘開胃湯) - 처방		
인삼개위탕(人蔘開胃湯)[동의방제와처방해설] - 처방		
인삼개위탕(人蔘開胃湯)[東醫寶鑑] - 처방		
인삼개위탕(人蔘開胃湯)[東醫寶鑑] - 처방		
인삼견터판(人蔘健脾丸)[비계내과환] - 처방		
인삼고(人蔘膏) - 처방		
인삼고(人蔘膏)[동의궤계내과환] - 처방		

그림 6. 온톨로지 검색어 추천 화면

[그림 4]는 온톨로지의 약재 클래스의 인삼을 검색한 결과 화면이다. [그림 4]와 같이 검색 결과 화면은 크게 두 부분으로 나뉘어진다. 왼쪽 메뉴는 검색된 인스턴스를 프로퍼티값으로 가지는 클래스를 보여주며, 메뉴를 클릭하면 검색된 인스턴스를 프로퍼티값으로 가지는 인스턴스들을 보여준다. 각 클래스마다 프로퍼티가 다르기 때문에 검색된 인스턴스의 타입에 따라서 메뉴 항목도 다르게 나타난다. [표 1]은 검색된 인스턴스의 클래스에 따라서 나오는 메뉴 항목을 나타낸 것이다.

표 1. 검색어 클래스 타입별 메뉴 항목

검색어 클래스	메뉴 항목
약재	약재, 연관 처방, 연관 병증
병증	병증, 연관 처방, 연관 약재
처방	처방, 연관 약재, 연관 병증
효능	효능, 연관 약재, 연관 처방
증상	증상, 연관 병증

약재 클래스인 인삼을 검색하는 경우 [그림 4]와 같이 약재, 연관 처방, 연관 병증의 세 개의 메뉴가 나온다. 약재 메뉴는 인삼의 상세 정보를 보여주며, 연관 처방 메뉴를 클릭하면 인삼을 구성약재로 가지는 처방 리스트, 연관 병증 메뉴를 클릭하면 인삼을 치료약재로 가지는 병증 리스트를 볼 수 있다.

그리고 검색 결과 화면에서 가운데 부분은 검색된 인스턴스가 가지는 프로퍼티값들을 보여준다. [그림 4]에서 인삼의 검색 결과인 설명, 기원명, 과명, 분류, 귀경, 성, 미, 이명, 약용부위, 효능, 주치등은 모두 약재 클래스의 프로퍼티들이다. 약재의 경우 약재의 사진이 같이 보여지는데 이는 온톨로지에 있는 것은 아니며 서버에 별도로 저장된 사진을 링크로 보여주게 된다.

이와 같은 [그림 4]의 결과는 [그림 5]의 인스턴스 검색 단계에서 이루어지는데 [그림 7]은 이 단계의 세부 프로세스를 보여준다.

우선 입력된 검색어를 가지고 온톨로지 인스턴스의 한글명, 한자명, 영문명에서 동일한 이름이 있는지 검사한다. 만약 존재하면 해당 인스턴스의 프로퍼티와 프로퍼티값을 검색한다. 또한 검색한 인스턴스가 약재 클래스의 인스턴스이면 약대 정보를 추가로 검색하고 처방 클래스의 인스턴스이면 구성 약재와 분석과 유사 처방

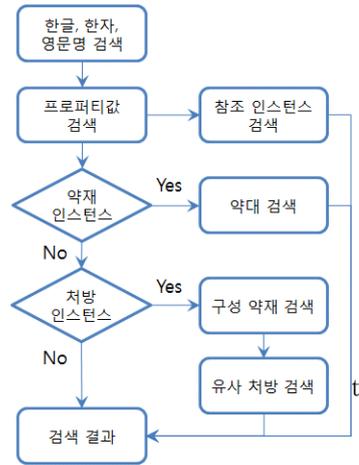


그림 7. “인스턴스 검색” 세부 프로세스

검색을 추가로 수행한다. 약대 검색 단계에서는 [그림 1]과 같이 약재와 연결된 처방 그리고 처방과 연결된 약재를 모두 검색해서 검색된 결과를 보여준다. 구성 약재 검색 단계에서는 처방 구성약재가 가지는 분류, 귀경, 성, 미를 추가로 검색하고 이를 분류해서 차트로 보여줄 수 있도록 구성한다. 유사 처방 검색 단계에서는 처방의 구성약재를 가지는 다른 모든 처방을 검색해서 구성약재가 일치하는 개수대로 처방을 정렬하게 된다.

검색어를 포함하지만 정확히 일치하는 인스턴스가 존재하지 않거나 검색어와 일치하지만 온톨로지에 검색어와 동일한 이름을 가지는 인스턴스가 1개 이상 존재하는 경우, 그리고 검색어가 2개 이상인데 모두 온톨로지에서 동일 클래스의 인스턴스가 아닌 경우에는 [그림 8]와 같이 해당하는 인스턴스들을 리스트로 보여주고 사용자가 어떤 것을 검색할 건지 선택하도록 한다.

특히, 검색어가 2개 이상 입력되는 경우에는 모두 온

Everything >

두통(頭痛)

<http://sem.kiom.re.kr/sem:medicine#D1頭痛>

설명 : 머리가 아픈 병증(病症).

출처 : 간계내과학

연관 병증 >

두통(頭痛)

<http://sem.kiom.re.kr/sem:medicine#S1頭痛>

설명 : 머리가 아픈 증상(症狀).

그림 8. 검색어와 일치하는 인스턴스가 1개 이상인 예제

서 보여준다.

- 처방에서 같이 사용되는 약재 검색
 한의학에서는 약대라고 해서 처방에서 주로 같이 사용되는 약재들에 대한 지식들이 존재한다. 즉, 여러 처방들을 보니 인삼은 어떤 약재들과 많이 사용된다는 것인데, 처방의 종류가 셀 수 없이 많기 때문에 약대에 대해서 명확하게 정리를 할 수는 없다. 대신, 본 연구에서는 온톨로지의 처방을 분석해서 약재를 검색했을 때 약대 정보까지도 제공한다.
- 2개 이상의 약재로 구성되는 처방 검색
 위의 유사한 약재 구성을 가지는 처방 검색과 처방에서 같이 사용되는 약재 검색 기능은 모두 하나의 약재나 처방을 검색했을 때 나오는 결과이다. 이와 달리 처음부터 어떤 약재들을 구성약재로 가지는 처방을 검색하고 싶을 때 검색창에 입력하고 검색하면 바로 검색 결과를 보여준다.
- 2개 이상의 증상을 수반하는 병증 검색
 두 개 이상의 증상을 검색하면 입력한 증상들을 수반하는 병증들을 검색할 수 있다.

IV. 결론

본 연구에서는 한의 온톨로지를 기반으로 시맨틱 검색 시스템을 설계하고 구현하였다. 시맨틱 검색은 사용자의 검색어가 어떤 개념에 속하는지 자동으로 해석하고 온톨로지를 이용해 정확한 검색 결과를 제공하는 장점이 있다. 본 연구에서 구현한 시맨틱 검색 시스템 또한 이와 같은 일반적인 시맨틱 검색 기능을 구현하였다. 뿐만 아니라 한의 온톨로지에 존재하는 한의학적인 시맨틱을 분석하고 이에 따른 검색 시나리오를 추가로 정의함으로써 한의학을 알고 있는 사용자들에게 유용한 검색 결과를 제공할 수 있도록 하였다.

한의 시맨틱 검색 시스템이 실제 사용자들에게 알려져서 활용도가 높아지려면 시스템의 기능도 중요하지만 검색 시스템의 성능도 중요한 요소가 된다. 검색 시스템의 성능은 여러 가지 요소로 분석이 가능하지만 중요한 요소 중의 하나는 데이터를 저장하고 검색하는 데

이터베이스의 성능일 것이다. 한의 시맨틱 검색 시스템은 현재 RDF 저장소의 오픈 소스 중 하나인 Jena TDB를 이용하고 있다. 하지만 다른 시맨틱 검색 시스템들은 각기 다른 저장 검색 시스템을 이용하고 있기 때문에 직접적인 비교는 힘들다. 향후에는 이에 대한 연구가 보완되어야 할 것이다.

한의 온톨로지는 사람을 치료하는 의료 지식을 가지고 있다. 현재 온톨로지는 한의학에 대한 시맨틱을 내포하고 있지만 실제 사람 특히 한의사가 가지고 있는 만큼의 지식을 체계적으로 가지지 않는다. 또한 온톨로지에 나오는 지식은 모든 환자에게 적용될 수 있는 것이 아니기 때문에 한의사의 확인을 받아야 한다. 따라서 본 연구에서 제공하는 시맨틱 검색 결과는 온톨로지로 구축된 한의 지식을 검색 및 공유하고 활용하기 위해서 사용해야 할 것이다.

참 고 문 헌

- [1] T. R. Gruber, "A translation approach to portable ontologies," Knowledge Acquisition, Vol.5, No.2, pp.199-220, 1993.
- [2] <http://thedatahub.org/>
- [3] R. Guha, Rob McCool, and Eric Miller, "Semantic Search," Proceedings of the 12th International Conference on World Wide Web, pp.700-709, 2003.
- [4] H. Jang, J. Kim, S. K. Kim, C. Kim, S. H. Bae, A. Kim, D. M. Eom, and M. Y. Song, "Ontology for medicinal materials based on traditional Korean medicine," Bioinformatics, Vol.26, No.18, pp.2359-60, 2010.
- [5] <http://www.seco.tkk.fi/publications/2005/makela-semantic-search-2005.pdf>.
- [6] Anusree. ramachandran and R. Sujatha, "Semantic search engine:A survey," Int. J. Comp. Tech. Appl., Vol.2, No.6, pp.1806-1811, 2011.
- [7] Xuezhong Zhou, Zhaohui Wu, Aining Yin,

Lancheng Wu, Weiyu Fan, and Ruen Zhang, "Ontology development for unified traditional Chinese medical language system," *Artificial Intelligence in Medicine*, Vol.32, pp.15-27, 2004.

[8] Huajun Chen, Yuxin Mao, Xiaoqing Zheng, Meng Cui, Yi Feng, Shuiguang Deng, Aining Yin, Chunying Zhou, Jinming Tang, Xiaohong Jiang, and Zhaohui Wu, "Towards Semantic e-Science for Traditional Chinese Medicine," *BMC Bioinformatics*, Vol.8, Suppl 3, S6, 2007.

[9] <http://happy.nate.com/nateAllNew/nateAllNew.html>

[10] <http://semanticwiki.salflux.com/index.php/IN2%20Platform>

[11] 성원경, 이승우, 한선화, 정한민, 김평, 이미경, 박동인, "시맨틱 웹 기술 기반 정보서비스 시스템 OntoFrame", 한국지능시스템학회 춘계학술대회, 2008.

[12] 정한민, 이미경, 류범중, "온톨로지 기반 사용자 제시 조건을 이용한 시맨틱 서비스 조합", 한국콘텐츠학회지, 제9권, 제10호, pp.32-39, 2009.

[13] 유동희, 서용무, "시맨틱 웹 기술을 이용한 온톨로지 기반 호텔 검색 시스템", 한국전자거래학회지, 제13권, 제4호, pp.71-92, 2008.

[14] 송장우, 김종훈, 정경용, 류중경, 이정현, "시맨틱 웹에서 개인화 프로파일을 이용한 콘텐츠 추천 검색 시스템", 한국콘텐츠학회지, 제8권, 제1호, pp.318-327, 2008.

[15] 오근현, 김용준, 조성배, "시각화 기반 모바일 라이프 로그 시맨틱 네트워크 연관 검색 시스템", 정보과학회논문지:컴퓨팅의 실제 및 레터, 제16권, 제6호, pp.727-731, 2010.

[16] <http://tkm.kiom.re.kr/search2012/index.do>

[17] <http://jena.apache.org/documentation/tdb/index.html>

저 자 소 개

김 상 균(Sang-Kyun Kim)

정회원



- 1999년 2월 : 충남대학교 정보통신공학과(공학사)
- 2001년 2월 : 충남대학교 컴퓨터공학과(공학석사)
- 2008년 2월 : 충남대학교 컴퓨터공학과(공학박사)
- 2007년 12월 ~ 현재 : 한국한의학연구원 선임연구원
<관심분야> : 시맨틱웹, 데이터베이스

박 동 훈(Dong-Hun Park)

정회원



- 2008년 2월 : 부산대학교 컴퓨터공학과(공학석사)
- 2012년 2월 ~ 현재 : 한국한의학연구원 연구원
<관심분야> : 시맨틱웹, 정보검색

김 안 나(Anna Kim)

정회원



- 2009년 2월 : 우석대학교 한의학과(한의학사)
- 2011년 8월 : 우석대학교 한의학과(한의학석사)
- 2009년 8월 ~ 현재 : 한국한의학연구원 연구원
<관심분야> : 한의분초·방제, 한의진료지원

오 용 택(Yong-Taek Oh)

정회원



- 2008년 2월 : 대구한의대학교 한의학과(한의학사)
- 2010년 2월 : 대구한의대학교 한의학과(한의학석사)
- 2010년 3월 ~ 현재 : 한국한의학연구원 연구원
<관심분야> : 한의학, 온톨로지

김 지 영(Ji-Young Kim)

정회원



- 2005년 2월 : 서울대학교 중어중문학과(문학사)
 - 2008년 2월 : 서울대학교 경영대학원(경영학석사)
 - 2012년 2월 : 부산대학교 한의학전문대학원(한의무석사)
 - 2012년 3월 ~ 현재 : 한국한의학연구원 연구원
- <관심분야> : 한의학 생리·병리

예 상 준(Sang-Jun Yea)

정회원



- 2002년 2월 : 한국과학기술원 전산학과(공학사)
 - 2004년 2월 : 한국과학기술원 전산학과(공학석사)
 - 2008년 5월 ~ 현재 : 한국한의학연구원 연구원
- <관심분야> : 영상처리, 한의약콘텐츠

김 철(Chul Kim)

정회원



- 1998년 2월 : 한국과학기술원 산업공학과(공학사)
 - 2000년 2월 : 한국과학기술원 산업공학과(공학석사)
 - 2009년 8월 : 원광대학교 한의정보학과(공학박사)
 - 2006년 7월 ~ 현재 : 한국한의학연구원 선임연구원
- <관심분야> : RFID/USN, 한의약정보화

장 현 철(Hyun Chul Jang)

정회원



- 1996년 8월 : 충남대학교 컴퓨터공학과(공학사)
 - 1999년 2월 : 충남대학교 컴퓨터공학과(공학석사)
 - 2007년 2월 : 충남대학교 컴퓨터공학과(공학박사)
 - 2008년 5월 ~ 현재 : 한국한의학연구원 책임연구원
- <관심분야> : 의료정보, 시맨틱, 온톨로지