

## 전통 약재 정보의 효율적인 활용을 위한 온톨로지 시스템 개발

차승준<sup>01</sup> 유정연<sup>1</sup> 엄동명<sup>2</sup> 고유미<sup>2</sup> 이규철<sup>1†</sup>

<sup>1</sup>충남대학교 컴퓨터공학과

<sup>2</sup>한국한의학연구원

{junii<sup>01</sup>, jyyou<sup>1</sup>, kcleee<sup>1</sup>}@cnu.ac.kr

{haksan<sup>2</sup>, kkomii99<sup>2</sup>}@kiom.re.kr

### Development of Ontology System for Putting Oriental Medicine Informations to Efficient Use

Seung-Jun Cha<sup>01</sup> Jeong-Youn Yu<sup>1</sup>

Dong Myung, Eom<sup>2</sup> You Mi, Ko<sup>2</sup> Kyu-Chul Lee<sup>1</sup>

<sup>1</sup>Department of Computer Engineering, Chungnam National University

<sup>2</sup>Korean institute of Oriental Medicine

#### 요 약

Tim Berners-Lee가 웹을 만든 이후 인터넷은 정보유통에 혁명적인 변화를 가져왔다. 이들 변화중 한가지는 개인이 접촉할 수 있는 정보의 양은 무한히 많아지게 되었다는 것이다. 이렇게 많아진 데이터에서 사용자들이 원하는 지식 정보를 정확하고 빠르게 검색하는 것이 중요한 이슈로 부각되었다. 그 해결책의 일환으로 시맨틱 웹(Semantic Web)이 많은 주목을 받고 있는데 이는 웹 정보에 잘 정의된 "의미 정보"를 부여하고 자동화 에이전트나 검색 엔진 등과 같은 소프트웨어들이 이용하게 한 것이다. 한국 한의학 연구원에서 운영하고 있는 한의학 지식 정보자원 중 약재들에 관한 데이터도 그 수에 있어서 매해 증가하고 있다. 따라서 사용자들에게 정확한 정보를 제공해주기 위해 기존의 키워드 검색 시스템에 시맨틱 웹을 사용하였다. 시맨틱 웹의 핵심 기술인 온톨로지를 사용하여 데이터를 구축하고 관리하는 시스템을 개발함으로써 사용자가 원하는 정보를 정확하게 제공하면서 또한 한의학 지식 자체를 전달할 수 있게 되었다. 뿐만 아니라 자원의 현대화는 물론 고급화된 자원의 효율적 가치 활용도를 높일 수 있다.

#### 1. 서 론

1990년대 Tim Berners-Lee가 웹을 창시한 이래로, 인터넷은 정보유통에 혁명적 변화를 가져왔다. 이는 양방향이면서, 거의 무시할 수 있는 수준의 낮은 유통비용, 정보의 실시간 획득과 제공 가능, 개인이 직접 정보를 획득·가공·배포를 할 수 있는 등 정보유통의 모든 측면에서 한계를 극복했다

이에 따라 개인이 접촉할 수 있는 정보의 양이 무한히 많아지게 되었다. 따라서 사용자가 원하는 지식정보를 정확하고 빠르게 검색하여 의미있는 지식 정보를 제공할 수 있는 정보검색 시스템이 필요하게 되었다. 즉, 정보검색 시스템은 사용자 측면에서 다양한 정보 검색을 통해 의미 있는 지식 정보를 제공할 수 있도록 하는 매개체 역할을 담당할 수 있어야 한다.[1]

기존의 키워드의 형태에만 의존하는 고전적 방식의 정

보 검색 기술은 한계점으로 나타났다. 사용지가 원하는 정보를 정확하게 찾아내고 조직화하고 관리하는 것이 더욱 더 어려워지고 있으며 단 한번도 활용되지 않고 사장되는 정보도 증가하고 있다.[2]

이러한 문제점들에 대한 해결 방안으로 시맨틱 웹(Semantic Web)이 많은 주목을 받고 있다. 시맨틱 웹은 차세대 웹 기술로서 뿐만 아니라 정보 유통 분야에서 자동화되고 유기적으로 통합된 지식 정보 공유 체계의 기반 기술로서 널리 채택되어 적용되고 있다.[3]

현재 한국 한의학 연구원에서 운영하고 있는 한의학 지식 정보 자원 웹 서비스는 고문헌 정보, 한방의료정보, 약초나 문헌등의 이미지들을 사용자에게 제공하는 서비스이다. 이러한 서비스도 그 정보의 양이 무한히 많아지고 있지만 서명순, 카테고리에 의한 분류 검색, 사용자가 입력한 조건을 기반으로 한 조건 검색만을 제공하고 있는 한계점이 있다.

더구나, 한의학 지식정보자원 웹 서비스는 같은 정보

† 교신저자

에 대해 사용자가 입력하는 방법에 따라 다른 결과를 제공할 수 있다. 예를 들어, 갈근은 감갈(甘葛), 분갈(粉葛), 건갈(乾葛), 갈분(葛粉)과 같은 다른 이름을 가지지만, 각각을 이용하여 검색한 결과들은 서로 다른 검색 결과들을 제공한다. 또한 감갈(甘葛)과 감갈(疳葛)과 같이 같은 발음을 가지나 서로 다른 의미를 가지는 경우에도 잘못된 검색 결과를 제공하는 문제점이 있다.[4]

이러한 문제점은 웹에서와 마찬가지로 시맨틱 웹을 통해 해결될 수 있다. 시맨틱 웹의 핵심 기술인 온톨로지를 사용한 온톨로지 검색 시스템은 의미 기반의 정보 검색 기능을 제공함으로써, 사용자가 원하는 정보를 보다 정확하게 제공하면서 또한 한의학 지식 자체를 전달할 수 있다. 이는 사용자에게 올바른 한의 지식 전달하고, 폭넓은 지식의 활용 기회를 제공하는데 매우 중요한 것이다.

이 논문에서는 이러한 전통 약재 정보들의 효율적인 활용을 위해 약재 정보들의 정보를 온톨로지로 작성하여 관리하는 시스템에 대한 내용이다

본 논문의 구성은 다음과 같다. 2장에서는 시맨틱 웹과 그와 연관된 온톨로지, OWL, 프로티지에 관해 배경을 설명하고 3장에는 이와 관련된 연구들에 대해 기술한다. 5장에서는 온톨로지 구축을 위한 약재 온톨로지 시스템 구성에 대해 설명하고 6장에서는 온톨로지 검색 시스템에 대해 알아보았다. 마지막 7장에서 결론을 도출하였다.

## 2. 배경

### 2.1. 시맨틱 웹(Semantic Web)

시맨틱 웹은 Tim Berners-Lee가 최초로 제안한 개념으로서 현재와 같은 웹 정보에 잘 정의된 "의미 정보"를 부여하고 자동화 에이전트나 검색 엔진 등과 같은 소프트웨어들이 이용하게 함으로써 고수준의 정보 서비스 혹은 정보 유통 체계의 고도화된 기능화와 자동화의 구현을 목표로 한다. 의미 부여란 컴퓨터 소프트웨어가 쉽게 이해할 수 있고 처리할 수 있는 형태로 정보 또는 자원의 내용에 대한 메타데이터를 선언적으로 기술하는 것이다(그림 1). [5]

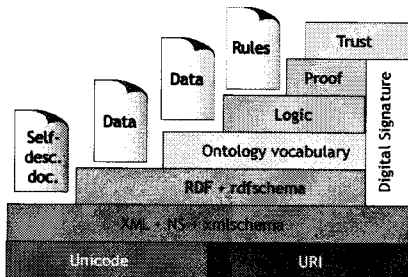


그림 1 시맨틱 웹 스택(Semantic Web Stack)

### 2.2. 온톨로지(Ontology)

시맨틱 웹에서 중요한 핵심 기술인 온톨로지란 특정 도메인(domain)내에서 공유되는 데이터들의 개념화한 형식적이고 명백한 규정이며, 이는 특정분야에서 사용되는 표준 어휘들의 모음이라고 할 수 있다.

온톨로지 구축에는 표준화된 언어가 필수적인데, 온톨로지 언어 표준으로는 W3C(World-Wide-Web Consortium)의 OWL(Web Ontology Language)가 대표적이다

### 2.3. OWL(Web Ontology Language)

OWL은 온톨로지를 구축하기 위한 표준 언어로 W3C 워킹 그룹에 의해 개발되었다. OWL을 사용하여 구축된 온톨로지는 사람과 컴퓨터가 모두 이해 및 활용이 가능하다. 즉, OWL 온톨로지는 컴퓨터로 직접 처리할 수 있는 애플리케이션을 구현하는데 활용될 수 있도록 서례된 언어이다. [6]

### 2.4. 프로티지(Protege)

프로티지는 OWL 기반의 공개된 대표적인 편집이로, 다른 편집기들보다 OWL의 다양한 엘리먼트들을 사용자 요구사항에 맞게 잘 정의할 수 있게 해 주고, 또한 다양한 플러그인을 통하여 원하는 기능들을 추가시킬 수 있는 확장성도 겸비한 편집기이다. [7][8]

## 3. 관련연구

한의학 분야에서 온톨로지를 기반으로 한 검색 시스템 개발은 아직 개발 초기에 있다. 최근들어, 국내에서는 2005년에 한의학 논문 검색 시스템 개발을 통해 한의학 진단 시스템 개발을 통한 시도가 이루어져 왔다. 특히 인터넷 건강 정보 게이트웨이 시스템은 질병과 증상에 대해 계층 구조의 온톨로지를 검색할 수 있다.(그림 2) [9]

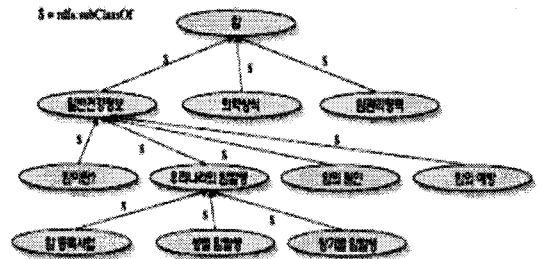


그림 2 인터넷 건강 정보 게이트웨이 시스템에서 제공하는 온톨로지 예

국외에서는 단순 약 및 처방 지식 추출 시스템에 대한 개발만이 이루어지고 있다. 한의학 분야와는 달리 생물학 학과 서양 의학에서 온톨로지를 기반으로 한 검색 시스템 개발은 이미 오래 전부터 진행되어 왔다. 생물학 분야에서는 방대한 양의 생물 데이터를 체계적으로 표현하고 관리하는 온톨로지를 구축하고 생물 정보를 검색하는 작업을 진행해왔다. 의학 분야 역시 질병을 분류함으로써 의학 정보에 대한 지식을 구축하고 다양하게 정의된 의학 정보들을 통합하고, 이를 검색하고 활용한 의료 관련 시스템들을 개발함으로써 의료 기술을 개발해 왔다. 특히 MeSH는 NLM(The US National Library of Medicine)에서 개발한 것으로 생물학자 뿐만 아니라 의료학자들이 의학 문헌들을 인덱스 및 분류하고 검색할 수 있도록 정의한 온톨로지이다.(그림 3) [10]

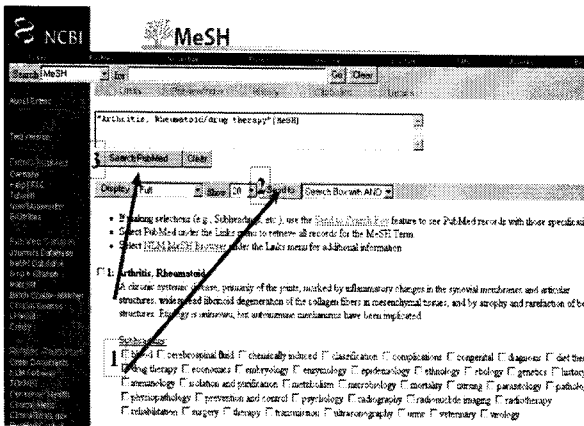


그림 3 MeSH를 이용한 논문 검색

하지만 현재 약재 정보에 관해서는 한국한의학연수원에서 적은 양의 온톨로지만이 시험적으로 구현되어 있고, 또한 현재 구축된 온톨로지는 토픽맵(Topic Maps) 기반으로 구축되어 있어 다른 시맨틱 웹의 표준인 OWL로 구축된 온톨로지와 호환되지 못한다. 따라서 전통 약재분야의 지능형 지식 검색을 위해 보다 폭넓은 지식 구축 및 시스템 개발이 필요하게 되었다

#### 4. 약재 온톨로지 시스템 구성

한의학 연구원에서 현재 쓰이는 약재 데이터는 총 16개의 속성을 가지고 있다. OWL에서는 크게 세가지의 타입으로 온톨로지를 구성되는데 그것들은 Class, Property, Individual이다. 따라서 약재 온톨로지 데이터

도 세가지의 구성요소로 데이터를 설계하였다.(그림 4)

Class는 기본적으로 약재 및 속성을 구성하는 기본 단위로 16가지의 속성에 맞추어 하나씩의 클래스들로 구성된다. 또한 각각의 클래스는 그 특징에 따라 계층구조를 나타낼 수 있는데 이것은 서브클래스(SubClass)로 정의하였다.

Property는 앞에서 구성한 클래스들 간의 관계를 정의하거나 혹은 클래스의 데이터를 정의한 것이다. 전자는 오브젝트 프로퍼티(ObjectProperty), 후자는 데이터 프로퍼티(DataProperty)로 정의되었다.

Individual은 실제적으로 약재에 대한 정보를 가지고 있는 것으로 앞에서 정의한 Class와 Property를 통해서 각각의 속성에서 정의된 값을 연결하였다.

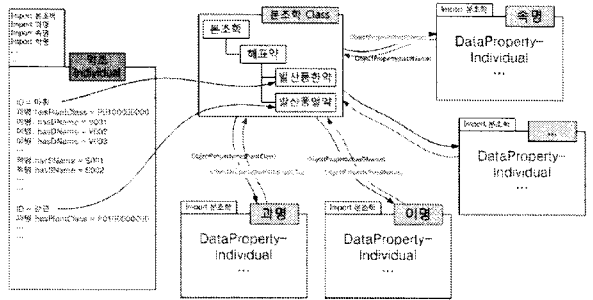


그림 4 약재 온톨로지의 구성도

약재 온톨로지 시스템의 구성하기 위해 다음과 같은 오픈 소스 및 에디터를 사용하였다.

- 이클립스(Eclipse) 3.2
- 프로티지(Protege) 3.3 API

이클립스는 자바를 저작하는데 사용되는 널리 사용되는 도구로 본 시스템의 코딩을 하는데 사용하였다. 프로티지는 온톨로지 저작도구로 제공하는 API를 사용해 구성된 OWL파일을 내장 파서인 Jena를 사용하여 유용성을 검사하였고 또한 OWL 파일을 수정하는데 사용되었다.

#### 5. 온톨로지 관리 시스템

온톨로지 관리 시스템은 구축되어있는 약재 온톨로지 파일들에 대한 데이터를 효율적으로 관리할 수 있게 하는 시스템이다, 이는 기존에 구축된 온톨로지와, 그것들의 관계성을 바탕으로 약재 데이터 및 속성들의 값을 추가, 수정, 삭제가 가능하도록 구현하였다.

추가과정은 약제 데이터에 대한 추가로 크게 4가지로 분류되어 적용된다. 우선 첫 번째로는 텍스트만 입력하는 필드로서 추가하려는 약제데이터의 기본 정보인 한글 이름, 한자이름, 라틴어가 그것에 속한다. 기존값에서 검색하는 부분은 분류, 과명, 금기, 수치법과 같이 기존에 입력된 데이터에서 추가하지 않고 단지 읽어와서 선택하는 것이다. 새로운 값을 입력하는 것이라 기존의 값과는 상관없이 새로운 값을 추가적으로 입력하는 것으로 이름, 학명, 속명, 약성가가 그것에 속한다. 마지막 남은 속성들은 기존에 구성된 데이터를 선택하여 사용할 수 있지만 추가적으로 해당 속성값에 데이터를 추가할 수 있다.

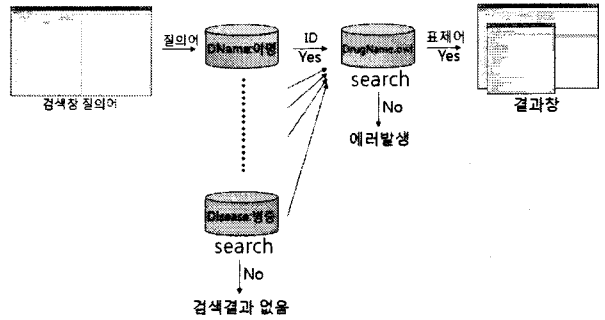


그림 6 온톨로지 검색 시스템 알고리즘

온톨로지 검색 시스템에서 제공하는 검색의 방법은 크게 단순검색과 상세검색으로 나누어져 있다

단순 검색은 특정 검색값은 검색하는 것으로서 정의되어 있는 모든 클래스에 질의를 하지만 사용자가 알고자 하는 값은 그것에 대한 약제명(표제어)의 값이기 때문에 검색한 질의어에 대한 표제어의 정보들을 반환해주고, 사용자는 검색된 정보들 중에 원하는 값을 선택하여 내용을 확인할 수 있다.

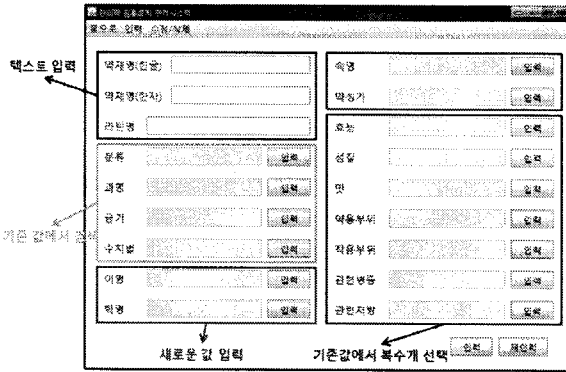


그림 5 약제 온톨로지 입력 시스템

수정/삭제는 뒤에서 설명될 온톨로지 검색 시스템을 사용하여 수정/삭제하고자 하는 데이터를 찾은 후, 추가과정과 같은 단계를 거쳐서 선택된 데이터를 수정 / 삭제한다.

## 6. 온톨로지 검색 시스템

약제 온톨로지 검색 시스템은 구축되어 있는 약제 온톨로지를 이용하여 의미 기반으로 약제 정보를 검색함으로써, 사용자가 원하는 정보를 보다 정확하게 검색하여 제공하는 시스템이다. 온톨로지 검색시스템은 입력받은 질의어에 대하여 정의된 클래스의 속성 값을 통해 값을 비교하게 되고 검색된 값을 모아서 결과값으로 전달하게 된다.

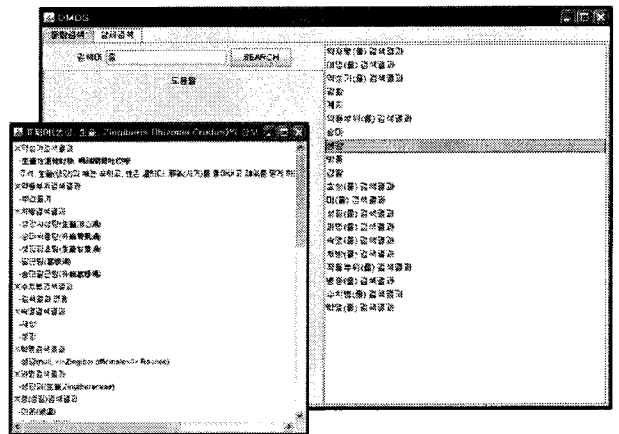


그림 7 통합검색의 예

상세검색은 약제 온톨로지에 정의된 특정 클래스에 질의를 하는 것으로, 하나의 클래스 뿐만 아니라 여러개의 클래스에 질의하여 그 결과값을 AND/OR 연산을 통해 더욱 정확한 결과값을 도출할 수 있다. 결과값은 단순검색과 동일하게 약제명(표제어)에 대한 검색 결과를 제공한다.

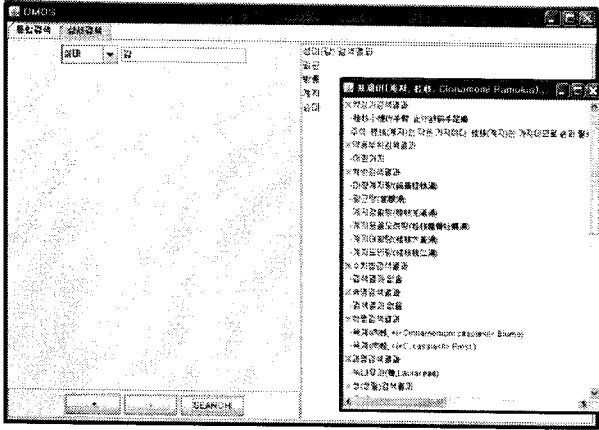


그림 8 온톨로지 상해검색의 예

## 7. 결론

웹 발전에 따른 정보량의 증가로 인해 효율적인 정보 검색의 필요성이 대두되고 있다. Tim Berners-Lee가 제시한 시맨틱 웹은 이러한 문제점을 지능화된 웹을 이용하여 해결하고자 하였다.

한의학에서 사용되는 수많은 약재정보들에 대해서 효율적인 검색을 위해 시맨틱웹을 사용하였다. 이는 기존에 쓰이던 약재 정보들을 온톨로지 언어인 OWL로 변환하여 데이터를 구성하였고, 또한 이렇게 구성된 온톨로지 파일들을 효과적으로 관리, 검색하기 위한 시스템을 구현하였다.

이러한 시스템의 구현으로 전통 지식 데이터베이스의 가치를 극대화 할 수 있다. 즉, 한의학 분야의 전통지식 데이터베이스에 언어공학적 의미 분석 단계의 기술 중 하나인 온톨로지 개념을 도입함으로써 자원의 현대화는 물론 고급화된 자원의 효율적 가치 활용도를 높일 수 있다. 또한 분산된 자원의 효율적인 활용방안으로 쓰일 수 있다. 분야별로 세분화되어 서로 분산 구축되어 있는 자원을 온톨로지 구축을 통해 통합함으로써 실제적인 자원의 활용이 가능한 최대한의 환경을 구현할 수 있다. 마지막으로 전통 지식의 통합된 서비스를 설계할 수 있다. 이는 향후 한의학 전반적인 분야의 자원을 단계별로 체계화하여 관리함으로써 사용자 중심의 통합되고 현대화된 전통 의학 정보의 서비스가 가능하도록 할 것이다.

## 8. Acknowledge

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성지원사업 (IITA-2005-C1090-0502-0016)의 연구비 지원으로 수행하였습니다.

## 9. 참고문헌

- [1] 최호섭, 옥철영, "정보검색 시스템과 온톨로지", 정보과학회지 22권 4호, 2004
- [2] Vishnu Challam and Susan Gauch, "Contextual Information Retrieval Using Ontology-Based User Profiles", Information and Telecommunication Technology Center
- [3] 전중홍, "차세대 웹 기술을 적용한 지식관리시스템 아키텍처에 관한 연구" 한국 전자거래학회지, 8권, 4호, 2003
- [4] 이현실, 이두영, "온톨로지 기반 한의학 처방 지식관리시스템 설계에 관한 연구", 정보관리학회지, 제20권, 제1호, pp. 341-371, 2003
- [5] T.Berners-Lee, J. Hendler, O. Lassila " The Semantic Web", Scientific American, 2001
- [6] OWL Web Ontology Language Reference, <http://www.w3g.org/TR/owl-ref>, 2004
- [7] 임형신, 황윤영, 엄동명, 이규철, "OWL 기반 온톨로지 도구 비교분석", 한국한의학연구원 논문집, 12/1, 2006
- [8] 박경모, 박중현, "Protege를 이용한 한의학의 구조화된 증상 입력을 위한 온톨로지 개발", 제20차 대한의료정보학회 춘계학술대회 논문집, 2003
- [9] 인터넷 건강정보 게이트웨이 시스템 개발, 한국보건사회연구원, 2003
- [10] Medical Subject Heading - MeSH Browser, <http://www.nlm.nih.gov/mesh>