

알츠하이머 관련 논문을 대상으로 하는 온톨로지 기반 지식 표현 방법 연구[☆]

A Study on Ontology Based Knowledge Representation Method with the Alzheimer Disease Related Articles

이 재 호¹ 김 연 희² 신 현 경^{3*} 송 기 봉⁴
Jaeho Lee Younhee Kim Hyunkyung Shin Kibong Song

요 약

의료 분야에서는 질병의 진단과 치료를 목적으로 하는 지식베이스 구축에 관심이 높다. 이러한 목적의 지식베이스를 구축하는 데 가장 중요한 것은 정확하게 지식을 표현하는 것이다. 본 논문에서는 온톨로지를 이용해 최근 의료 분야에서 많은 관심을 받고 있는 알츠하이머 질병과 관련된 국내 논문들을 대상으로 지식을 표현하는 방법을 제안하였다. 본 논문에서 제안한 온톨로지 기반 지식 표현 방법은 저자, 발행기관 등과 같은 서지 정보에서 추출한 클래스들은 물론 논문의 제목, 초록, 키워드, 결론에서 추출한 연구 주제와 관련된 클래스들을 모두 정의하고 프로퍼티를 통해 클래스들간의 다양한 의미적 관계를 포함하고 있다. 그리고 클래스들간의 계층 관계와 프로퍼티의 이행적 특성도 포함하고 있기 때문에 이를 이용한 추론을 지원한다. 따라서 단순한 키워드 검색뿐만 아니라 의미에 기반을 둔 지식 검색이 가능하다. 또한 온톨로지 검색 언어인 SPARQL을 이용해 추론을 통한 지식 검색 요청을 보다 쉽게 표현할 수 있다.

☞ 주제어 : 알츠하이머 질병, 온톨로지, 지식 표현, 학술 논문 데이터베이스, 추론 기반 검색

ABSTRACT

In the medical field, for the purpose of diagnosis and treatment of diseases, building knowledge base has received a lot of attention. The most important thing to build a knowledge base is representing the knowledge accurately. In this paper we suggest a knowledge representation method using Ontology technique with the datasets obtained from the domestic papers on Alzheimer disease that has received a lot of attention recently in the medical field. The suggested Ontology for Alzheimer disease defines all the possible classes: lexical information from journals such as 'author' and 'publisher' research subjects extracted from 'title', 'abstract', 'keywords', and 'results'. It also included various semantic relationships between classes through the Ontology properties. Inference can be supported since our Ontology adopts hierarchical tree structure for the classes and transitional characteristics of the properties. Therefore, semantic representation based query is allowed as well as simple keyword query, which enables inference based knowledge query using an Ontology query language 'SPARQL'.

☞ keyword : Alzheimer disease, ontology, knowledge representation, database of domestic academic papers, inference based query

1. 서 론

의료 분야의 데이터는 환자의 질병 진단과 치료를 위해 중요하게 사용된다. 의학의 발달과 정보 처리 기술의 발달로 인해 의료 분야의 데이터 양이 기하급수적으로 증가하고 그 유형도 다양해지고 있다. 의료 분야의 데이터를 환자의 병증을 해석하고 치료하는데 활용하기 위해서는 방대하고 다양한 의료 데이터로부터 가치있는 지식을 추출하고 추출한 지식을 효율적으로 관리하는 것이 무엇보다 중요하다. 이러한 이유로 의료 분야에서는 지식을 효율적으로 관리할 수 있는 지식베이스(knowledge

¹ Department of Computer Education, Gyeongsin National University of Education

² Department of e-Business, Bucheon University

³ Department of Mathematical Science, Gachon University

⁴ Electronics and Telecommunications Research Institute

* Corresponding author (hyunkyung@gachon.ac.kr)

[Received 02 June 2014, Reviewed 08 June 2014, Accepted 18 June 2014]

☆ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 휴먼인 지환경사업본부-신기술융합형 성장동력사업의 지원을 받아 수행된 연구임(2013K000384).

base)의 구축에 많은 관심을 가져왔다[1, 2, 3, 4, 5].

지식베이스를 이용하면 저장되어 있는 지식은 물론 추론으로 유도되는 새로운 지식을 통해 비전문가인 사용자도 문제 해결을 위한 지식 기반의 판단이 가능하게 된다. 지식베이스 구축 시 가장 중요하게 고려해야 하는 것은 사용자에게 꼭 필요하고 이후 추론 과정에서도 가치 있게 활용될 수 있는 지식을 확보하는 것이다. 따라서 특정 도메인에서 제공되는 다양한 데이터들로부터 어떻게 지식을 획득하고, 획득한 지식을 어떻게 효율적으로 표현할 것인가에 관한 이론에 해당하는 지식 표현의 방법은 지식베이스의 성패를 좌우하는 중요한 요소라고 할 수 있다.

전 세계적으로 고령화 인구가 크게 증가하면서 예방이나 치료 방법이 명확하게 밝혀지지 않은 알츠하이머 질병에 대한 관심이 높아지고 있으며 의료 분야에서 알츠하이머와 관련한 예방 및 치료와 관련한 연구가 활발하게 진행되고 있다. 따라서 알츠하이머 질병과 관련한 연구 결과물로부터 지식을 추출하여 관리하고 이를 다양하게 활용하고자 하는 요구가 증가하고 있다[6].

본 논문에서는 의료 분야에서 최근 관심이 높은 알츠하이머 질병과 관련한 데이터를 대상으로 온톨로지를 이용한 지식 표현 방법을 제안하였다. 특히, 알츠하이머 질병과 관련하여 국내 학술지에 게재된 논문을 전문 지식의 수집 대상으로 선정하였다. 본 논문에서 제안한 지식 표현 방법은 논문으로부터 제공되는 지식의 표현에 온톨로지 기술을 활용하고 온톨로지 검색 언어를 추론에 활용함으로써 지식의 의미를 보다 정확하게 이해하고 단순한 키워드 검색이 아닌 의미 기반의 지식 검색을 지원한다.

본 논문은 다음과 같이 구성된다. 2장에서는 온톨로지를 기술하기 위한 표준 언어인 OWL에 대해 소개하고 온톨로지를 이용한 의료 데이터의 지식 표현과 관련한 기존 연구 내용과 온톨로지를 이용한 논문 검색과 관련한 기존 연구 내용을 제시하였다. 3장에서는 알츠하이머 관련 논문을 대상으로 하는 온톨로지 기반의 지식 표현 방법을 제안하였다. 4장에서는 본 논문에서 제안한 지식 표현 방법으로 온톨로지를 구축한 예와 의미에 기반을 둔 검색 방법을 소개하였다. 그리고 5장에서 결론을 맺는다.

2. 관련 연구

2.1 OWL

OWL(Web Ontology Language)은 시맨틱 웹(Semantic

Web)에서 온톨로지를 정형화된 형태로 기술하기 위해 사용하는 표준화된 언어이다[7, 8, 9]. 특정 도메인에 대한 지식표현 방법론으로서 온톨로지는 지식의 단위를 개념(concept)으로 보고 이를 표현하기 위한 데이터 구조로서 클래스(class)를 정의한다. 해당 개념의 상세 특징을 표현하기 위해 모든 클래스는 하부 데이터 구조로서 속성(property)을 정의하고 이를 포함한다. 속성은 또한 각 클래스 간의 연관 관계를 표현하기 위한 용도로도 사용된다. 그리고 지식 표현의 기본 단위인 정보 자원 개체(individual)를 특정 클래스의 타입으로 지정하고 정보 자원 개체의 특성이나 다른 정보 자원 개체와의 의미적 관계를 미리 정의한 프로퍼티를 이용해 기술한다.

OWL은 주어-서술어-목적어(subject-predicate-object) 구조의 트리플 문장들로 구성된 문서의 형태로 온톨로지를 기술하지만 주어와 목적어를 노드로, 서술어를 간선으로 나타내는 방향성 그래프 모델로도 표현이 가능하다.

2.2 온톨로지를 이용한 의료 데이터의 표현

의료 분야에서 지식 표현을 위해 온톨로지를 활용하려는 연구가 계속 진행되어 왔다. 온톨로지를 이용한 의료 데이터의 지식 표현 방법에 관한 기존의 연구 내용은 데이터 유형 따라 크게 세 가지로 분류할 수 있다.

첫 번째는 의학 용어들을 대상으로 온톨로지를 구축한 사례로 Gene Ontology, GALEN, UMLS 등이 대표적이다. Gene Ontology는 생물학 분야에서 주로 사용되는 용어와 용어들간의 관계를 온톨로지로 표현한 것이고, GALEN은 해부학, 생리학, 병리학 분야에서 사용되는 핵심적인 의학 용어들과 용어들간의 계층 관계를 온톨로지로 표현 것이다[1, 2]. 그리고 UMLS는 15개국 약 100여 개의 의료분야에 관련된 용어체계를 통합하여 온톨로지로 표현한 것이다[1].

두 번째는 임상 데이터나 문진표, 의료 관련 FAQ 등의 의료 데이터를 대상으로 하는 사례로 U-Health 온톨로지나 암정보 제공을 위한 온톨로지 기반 대화시스템 등이 대표적이다. U-Health 온톨로지는 임상 데이터와 문진으로부터 얻은 데이터들로부터 질병의 종류, 관련 증상, 요인, 대처 방법들에 관련한 지식을 추출하여 온톨로지로 표현한 것이다[5]. 암정보 제공을 위한 온톨로지 기반 대화시스템은 암 관련 정보 제공 웹 사이트의 FAQ 자료에 대한 질문 분석을 통한 분류 체계를 바탕으로 질문 온톨로지를 구축하여 질의 응답에 사용하는 것이다[1].

세 번째는 의료 분야의 논문이나 서적을 대상으로 하

는 사례로 한의학 온톨로지가 대표적이다. 한의학 분야에서는 한의학 관련 논문이나 서적의 내용이 주로 병증, 증상, 원인, 약재, 처방 등에 대한 것이므로 이들을 개념화하고 이들간의 의미적 관계를 온톨로지로서 표현하였다 [3, 4, 10].

2.3 온톨로지를 이용한 논문 검색

일반적인 논문 검색은 키워드를 기반으로 단순한 문자열 매칭을 통해 이루어지기 때문에 사용자가 제시한 검색 조건과 관련이 없는 결과가 반환되거나 반환된 결과가 너무 많아 사용자가 다시 검색 결과를 확인해야 하는 과정이 필요하다. 반면에 논문 검색에 온톨로지를 이용하게 되면 키워드 기반 검색의 문제점을 해결하고 추론을 이용한 의미 기반 검색이 가능하다[10, 11, 12, 13].

하지만 온톨로지를 이용한 논문 검색과 관련한 기존 연구에서는 논문의 제목, 저자, 발행기관 등의 서지 정보를 대상으로만 온톨로지를 구축하여 논문 자체가 가지고 있는 내용에 대해서는 의미 검색이 가능하지 않거나[10, 11], 반대로 논문의 내용을 대상으로 온톨로지를 구축하였으나 참고 문헌, 분류 등 다양한 서지 정보는 고려하지 않는 경우[12, 13]가 많다.

3. 알츠하이머 논문 온톨로지 설계

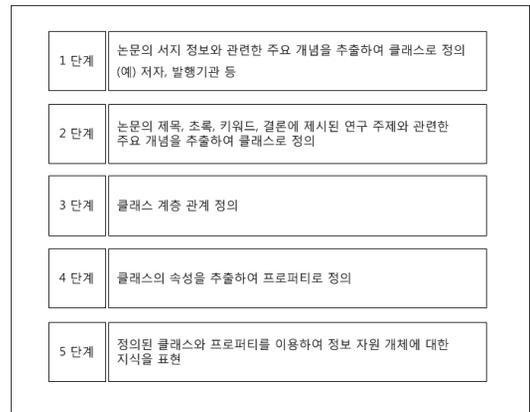
지식을 사람과 컴퓨터가 모두 이해할 수 있으면서도 정확하고 효율적인 관리가 가능하도록 표현하기 위해서 생성 규칙(production rule), 의미 네트워크(semantic network), 프레임 시스템(frame system), 온톨로지 등의 다양한 지식 표현 기법들이 사용되어 왔다.

일반적으로 많이 사용되는 생성 규칙에 의한 표현은 단순한 패턴 매칭을 통해 규칙을 검색하기 때문에 같은 의미라도 다르게 표현된 규칙에 대해 다른 해석을 하게 되는 문제가 있다. 이러한 문제를 해결하기 위해 의미 네트워크나 프레임 시스템 등의 방법을 사용하기도 하지만 정형화된 지식의 표현이나 의미에 기반을 둔 지식 검색에 대한 고려가 부족하다. 반면, 온톨로지는 다른 지식 표현 기술들에 비해 OWL과 같은 표준화된 언어를 이용해 정형화된 표현이 가능하고 다양한 의미적 관계를 표현할 수 있어 의미에 기반을 둔 지식 검색 측면에서 장점을 가지고 있다. 또한 온톨로지를 지원하는 SPARQL이라는 검색 언어가 별도로 존재하기 때문에 간단한 질의문

작성으로도 추론을 통한 검색을 보다 쉽게 지원할 수 있다. 따라서 본 논문에서는 알츠하이머 관련 논문에 대한 지식 표현을 위해 온톨로지를 활용하였다.

3.1 온톨로지 구축 과정

본 논문에서는 그림 1에서 제시한 5단계를 통해 국내 학술지에 게재된 알츠하이머 질병 관련 논문들을 대상으로 지식을 표현하기 위한 온톨로지를 구축하였다.



(그림 1) 온톨로지 구축 과정
(Figure 1) ontology construction process

그림 1의 첫 번째 단계에서는 알츠하이머 논문에 대해 제목, 저자, 발행기관, 논문지명, 발행년도 등에 대한 서지 정보를 추출하여 클래스로 정의한다. 서지 정보는 논문을 게재할 때 기본적으로 제공되는 정보로 쉽게 추출이 가능하기 때문에 일반적인 논문 검색에서 검색 조건으로 많이 사용된다. 따라서 알츠하이머 논문을 위한 온톨로지를 구축하기 위해서는 서지 정보와 관련한 개념들을 추출하고 이를 클래스로 정의하는 것이 가장 먼저 수행되어야 한다. 본 논문에서는 국내에서 논문 검색 시 일반적으로 많이 사용하는 DBpia[14]에서 일반 검색과 상세 검색을 위해 요청하는 서지 정보와 관련한 검색 조건들을 분석하고 그 외에 추가적으로 검색에 활용할 수 있는 서지 정보 관련 조건들을 추출하여 클래스로 정의하였다.

두 번째 단계에서는 알츠하이머 논문의 제목, 초록, 키워드, 결론에서 제시하는 연구 주제와 관련하여 주요 개념들을 추출하고 클래스로 정의한다. 논문의 본문에 비

해서 제목, 초록, 키워드, 결론에서는 연구 결과가 간단하고 명확하게 제시된다. 따라서 본 논문에서는 논문의 제목, 초록, 키워드, 결론에서 제시하는 연구 결과를 대상으로 논문의 연구 주제를 온톨로지로 표현함으로써 서지 정보만을 이용한 논문 검색에서는 제공하기 어려운 연구 주제 기반의 검색이 가능하도록 하였다.

세 번째 단계에서는 이전 단계를 통해 정의된 클래스들간에 성립하는 계층 구조를 추출하여 온톨로지에 정의한다.

네 번째 단계에서는 이전 단계를 통해 정의된 클래스에 대해 리터럴 데이터(literal data)로 기술되는 속성이나 다른 클래스와의 의미적 관계를 기술하는 속성을 추출하고 온톨로지에 프로퍼티로 정의한다.

다섯 번째 단계에서는 실제 알츠하이머 논문들에서 정보 자원 개체를 추출하고 정의된 클래스와 프로퍼티를 이용하여 정보 자원 개체들이 제공하는 지식을 온톨로지에 기술한다.

3.2 알츠하이머 논문 온톨로지

3.1절에서 제시한 5단계 구축 과정을 거쳐 본 논문에서 설계한 알츠하이머 논문 온톨로지에는 25개의 클래스가 정의되어 있으며 논문 전체를 의미하는 ‘ADPaper’ 클래스가 최상위 클래스이다. 25개의 클래스는 크게 서지 정보와 관련된 클래스와 연구 주제 정보와 관련된 클래스로 구분된다.

표 1은 서지 정보와 관련하여 정의된 16개의 클래스에 대해 각각의 이름과 의미를 설명한다. 본 논문에서는 DBPia에서 일반 검색과 상세 검색을 위해 사용자에게 요청하는 서지 정보와 관련한 검색 조건들을 분석해 키워드, 저자, 논문지, 발행기관, 분류, 언어, 발행년도의 7가지 개념을 추출하여 세분화하고 논문지 등급 등과 같이 추가적으로 검색에 활용할 수 있는 개념을 추가하여 클래스로 정의하였다.

표 2는 연구 주제와 관련하여 정의된 9개의 클래스에 대해 각각의 이름과 의미를 설명한다. 본 논문에서는 국내 논문지에 게재된 500편의 논문들을 수집하고 연구 주제를 분석하여 관련질병, 원인, 증상, 치료법, 예방법, 검사방법, 진단방법, 환자가족에 대한 8가지 연구 주제를 추출하고 클래스로 정의하였다. 그리고 8가지 연구 주제에 대한 클래스는 연구 주제를 의미하는 ‘Subject’ 클래스의 하위 클래스로 정의하였다.

(표 1) 서지 정보 클래스

(Table 1) bibliographic information class

클래스 이름	의미
ADPaper	논문 전체를 개념화한 클래스
Contents	논문의 내용과 관련한 전체 정보를 개념화한 클래스
Publication	논문의 발행과 관련한 전체 정보를 개념화한 클래스
Title	논문 제목을 개념화한 클래스
Keyword	논문 키워드를 개념화한 클래스
Author	논문 저자를 개념화한 클래스
FirstAuthor	제1저자를 개념화한 클래스
CorrespondingAuthor	교신저자를 개념화한 클래스
CoAuthor	공동저자를 개념화한 클래스
Classification	연구 주제와 관련한 분류 항목을 개념화한 클래스
Publisher	논문 발행기관을 개념화한 클래스
Journal	논문이 발행된 논문지를 개념화한 클래스
Grade	논문 등급을 개념화한 클래스
KCIRecord	학진등재지를 개념화한 클래스
KCICandidate	학진등재후보지를 개념화한 클래스
KCIUnRecord	비등재지를 개념화한 클래스

(표 2) 연구 주제 클래스

(Table 2) subject of research class

클래스 이름	의미
Subject	논문과 관련한 연구 주제를 개념화한 클래스
RelatedDisease	알츠하이머 관련 질병을 개념화한 클래스
Etiology	알츠하이머의 원인을 개념화한 클래스
Symptom	알츠하이머의 증상을 개념화한 클래스
Treatment	알츠하이머의 치료법을 개념화한 클래스
Prevention	알츠하이머의 예방법을 개념화한 클래스
Examination	알츠하이머의 검사방법을 개념화한 클래스
Diagnosis	알츠하이머의 진단방법을 개념화한 클래스
PatientFamily	알츠하이머 환자의 가족을 개념화한 클래스

표 3은 본 논문에서 제안한 알츠하이머 논문 온톨로지에 정의된 클래스들의 계층 관계를 설명한다. 표 3에서

제시한 생리학 분야를 의미하는 ‘Physiology’ 클래스와 병리학 분야를 의미하는 ‘Pathology’ 클래스는 표 1과 표 2에 제시한 알츠하이머 논문 클래스에는 현재 포함되어 있지 않다. 이들은 ‘Classification’ 클래스의 하위 클래스로서 필요에 따라 확장 가능하며 한국연구재단에서 제시한 의약학 분야의 분류 기준을 클래스로 정의한 것이다.

(표 3) 클래스 계층 관계
(Table 3) class hierarchy

상위 클래스 이름	하위 클래스 이름
Author	FirstAuthor
	CorrespondingAuthor
	CoAuthor
Grade	KCIRecord
	KCICandidate
	KCIUnRecord
Classification	Physiology
	Pathology
Subject	RelatedDisease
	Etiology
	Symptom
	Treatment
	Prevention
	Examination
	Diagnosis
	PatientFamily

표 4는 ‘ADPaper’ 클래스에 속한 프로퍼티를 설명한다. 프로퍼티는 클래스의 특성을 기술하기 위해 리터럴 데이터를 값으로 가지는 데이터 프로퍼티(DataProperty)와 다른 클래스와의 의미적 관계를 기술하기 위한 객체 프로퍼티(ObjectProperty)로 구분된다. ‘ADPaper’ 클래스의 프로퍼티 중 ‘referTo’ 프로퍼티는 이행적(transitive) 특성을 가지는 것으로 정의하였다. 만약 A 논문과 B 논문이 ‘referTo’ 관계가 있고 B 논문과 C 논문이 ‘referTo’ 관계가 있다면 이에 기반을 두고 A 논문과 C 논문이 연관 관계에 있음을 추론하여 논문 검색 시 이용할 수 있다.

표 5는 ‘Contents’ 클래스에 속한 프로퍼티를 설명한다. ‘Contents’ 클래스는 논문의 초록, 작성 언어, 사사 정보를 표현하기 위한 3개의 데이터 프로퍼티와 5개의 객체 프로퍼티를 가지고 있다.

(표 4) ADPaper 클래스의 프로퍼티
(Table 4) properties of ADPaper class

프로퍼티 이름	의미	유형
referTo	다른 ADPaper 클래스와의 인용 관계를 개념화한 프로퍼티	객체
hasContents	Contents 클래스와의 포함 관계를 개념화한 프로퍼티	객체
hasPublication	Publication 클래스와의 발행 관계를 개념화한 프로퍼티	객체

(표 5) Contents 클래스의 프로퍼티
(Table 5) properties of Contents class

프로퍼티 이름	의미	유형
abstract	논문의 초록 내용을 기술하는 프로퍼티	데이터
language	논문의 작성 언어를 기술하는 프로퍼티	데이터
acknowledgement	논문의 사사 정보를 기술하는 프로퍼티	데이터
hasTitle	Title 클래스와의 관계를 개념화한 프로퍼티	객체
hasAutor	Author 클래스와의 관계를 개념화한 프로퍼티	객체
hasKeyword	Keyword 클래스와의 관계를 개념화한 프로퍼티	객체
hasSubject	Subject 클래스와의 관계를 개념화한 프로퍼티	객체
hasClassified	Classification 클래스와의 관계를 개념화한 프로퍼티	객체

(표 6) Title 클래스의 프로퍼티
(Table 6) properties of Title class

프로퍼티 이름	의미	유형
hasKoreanTitle	논문의 국문제목을 기술하는 프로퍼티	데이터
hasEnglishTitle	논문의 영문제목을 기술하는 프로퍼티	데이터

표 6은 ‘Title’ 클래스에 속한 프로퍼티를 설명한다. ‘Title’ 클래스는 논문의 국문 제목과 영문 제목을 표현하기 위한 2개의 데이터 프로퍼티만을 가지고 있다.

표 7은 ‘Author’ 클래스에 속한 프로퍼티를 설명한다. ‘Author’ 클래스는 저자의 국문 이름, 영문 이름, 소속을 표현하기 위한 3개의 데이터 프로퍼티만을 가지고 있다.

(표 7) Author 클래스의 프로퍼티
(Table 7) properties of Author class

프로퍼티 이름	의미	유형
hasKoreanName	저자의 국문 이름을 기술하는 프로퍼티	데이터
hasEnglishName	저자의 영문 이름을 기술하는 프로퍼티	데이터
hasAffiliation	저자의 소속을 기술하는 프로퍼티	데이터

표 8은 ‘Publication’ 클래스에 속한 프로퍼티를 설명한다. ‘Publication’ 클래스는 논문의 페이지, 발행년도, 발행월, 발행권호, 논문 파일 경로를 표현하기 위한 6개의 데이터 프로퍼티와 2개의 객체 프로퍼티를 가지고 있다.

(표 8) Publication 클래스의 프로퍼티
(Table 8) properties of Publication class

프로퍼티 이름	의미	유형
hasPage	논문의 페이지 정보를 기술하는 프로퍼티	데이터
hasYear	논문의 발행년도를 기술하는 프로퍼티	데이터
hasMonth	논문의 발행월을 기술하는 프로퍼티	데이터
hasVol	논문의 발행권을 기술하는 프로퍼티	데이터
hasNo	논문의 발행호를 기술하는 프로퍼티	데이터
hasURI	논문의 파일이 존재하는 경로나 논문의 웹페이지 주소를 기술하는 프로퍼티	데이터
hasPublisher	논문을 발행한 기관인 Publisher 클래스와의 관계를 개념화한 프로퍼티	객체
publishedBy	논문의 발행지인 Journal 클래스와의 관계를 개념화한 프로퍼티	객체

표 9는 ‘Journal’ 클래스에 속한 프로퍼티를 설명한다. ‘Journal’ 클래스는 논문의 이름을 표현하기 위한 1개의 데이터 프로퍼티와 ‘Grade’ 클래스와의 의미적 관계를 표현하기 위한 1개의 객체 프로퍼티를 가지고 있다.

(표 9) Journal 클래스의 프로퍼티
(Table 9) properties of Journal class

프로퍼티 이름	의미	유형
hasJournalName	논문지 이름을 기술하는 프로퍼티	데이터
hasGrade	논문지 등급을 의미하는 Grade 클래스와의 관계를 개념화한 프로퍼티	객체

표 10은 ‘Keyword’ 클래스에 속한 프로퍼티를 설명한다. ‘Keyword’ 클래스는 논문에 제시된 키워드와 관련이 있는 연구 주제를 표현하기 위해 ‘Subject’ 클래스와 관계를 맺고 있는 1개의 객체 프로퍼티를 가지고 있다.

(표 10) Keyword 클래스의 프로퍼티
(Table 10) properties of Keyword class

프로퍼티 이름	의미	유형
relatedTo	논문 키워드와 관련된 Subject 클래스와의 관계를 개념화한 프로퍼티	객체

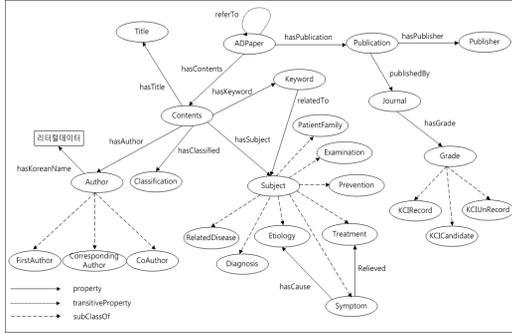
표 11은 ‘Symptom’ 클래스에 속한 프로퍼티를 설명한다. ‘Symptom’ 클래스는 알츠하이머 증상과 그것의 원인 또는 치료 방법을 연관시키기 위한 2개의 객체 프로퍼티를 가지고 있다.

(표 11) Symptom 클래스의 프로퍼티
(Table 11) properties of Symptom class

프로퍼티 이름	의미	유형
hasCause	알츠하이머 증상의 원인이 되는 Etiology 클래스와의 관계를 개념화한 프로퍼티	객체
relieved	알츠하이머 증상을 완화시키는 Treatment 클래스와의 관계를 개념화한 프로퍼티	객체

그림 2는 본 논문에서 설계한 알츠하이머 논문 온톨로지를 그래프 형태로 모델링한 결과를 보여준다. 클래스는 타원으로, 리터럴 데이터는 직사각형으로, 프로퍼티는 직선 화살표로 표현하였다. 그리고 클래스들간의 계층 관계를 의미하는 subClassOf 관계는 점선 화살표로 표현하였다. 클래스들간의 의미적 관계를 쉽게 파악할 수 있도록 그림 2는 온톨로지에 정의된 클래스들과 객체 프로

퍼티들을 중심으로 표현하였다. 데이터 프로퍼티는 4장에서 제시한 온톨로지 구축 예제와 검색 예제에서 사용된 것들만 표현하였다.



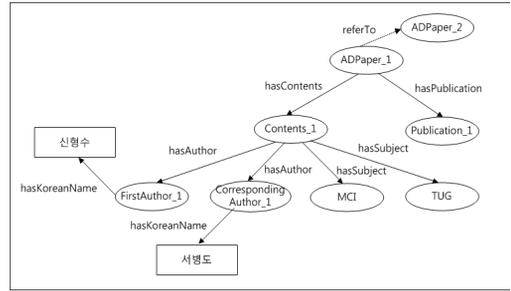
(그림 2) 알츠하이머 논문 온톨로지의 그래프 표현
(Figure 2) a graph model for Alzheimer disease paper ontology

4. 알츠하이머 논문 온톨로지 구축

4.1 온톨로지 구축 예

3장에서 설계한 온톨로지를 이용해 알츠하이머 논문을 대상으로 지식 표현이 가능하다. 본 절에서는 알츠하이머 질병과 관련하여 수집한 500편의 논문 중 [6]의 논문을 대상으로 3장에서 설계한 온톨로지를 이용해 실제로 지식 표현을 수행한 예를 제시하였다.

그림 3은 본 논문에서 설계한 알츠하이머 논문 온톨로지에 기반을 두고 참고문헌 [6]의 논문에서 추출한 지식 중 일부를 그래프 모델로 표현한 것이다. 추출한 지식에서 정보 자원 개체는 타원으로 표현하고 정보 자원 개체의 프로퍼티는 직선 화살표로 표현하였다. 그리고 리터럴 데이터로 표현된 정보 개체의 프로퍼티 값은 직사각형으로 표현하였다. 일반적으로 정보 자원 개체는 URI로 식별되는데 본 논문에서는 정보 자원 개체를 식별하면서 동시에 어떤 클래스 타입에 해당하는지를 명시적으로 나타내기 위해 클래스의 이름과 출현 순서에 따라 고유 아이디를 부여하였다. 그림 3에서 “경도인지장애와 알츠하이머 치매 여성노인의 하지운동기능 비교”라는 제목을 가진 참고문헌 [6]을 의미하는 ‘ADPaper_1’ 정보 자원 개체는 알츠하이머 논문 온톨로지서 정의한 ‘ADPaper’ 클래스 타입에 해당하고 가장 첫 번째로 출현한 것임을 알 수 있다.



(그림 3) 알츠하이머 논문 온톨로지를 이용한 지식 표현 결과를 나타낸 그래프 모델
(Figure 3) a graph model for knowledge representation using Alzheimer disease paper ontology

그림 4는 그림 3에서 제시한 그래프 모델을 OWL 언어로 기술한 결과를 보여준다.

```

<owl:Thing rdf:ID="ADPaper_1"/>
<owl:Thing rdf:about="#ADPaper_1">
  <rdf:type rdf:resource="#ADPaper"/>
  <hasContents rdf:resource="#Contents_1"/>
  <hasPublication rdf:resource="Publication_1"/>
  <referTo rdf:resource="ADPaper_2"/>
</owl:Thing>
<owl:Thing rdf:ID="ADPaper_2"/>
<owl:Thing rdf:ID="Contents_1"/>
<owl:Thing rdf:about="#Contents_1">
  <rdf:type rdf:resource="#Contents"/>
  <hasAuthor rdf:resource="#FirstAuthor_1"/>
  <hasAuthor rdf:resource="#CorrespondingAuthor_1"/>
  <hasSubject rdf:resource="#MCI"/>
  <hasSubject rdf:resource="#TUG"/>
</owl:Thing>
<owl:Thing rdf:ID="Publication_1"/>
<owl:Thing rdf:ID="FirstAuthor_1"/>
<owl:Thing rdf:about="#FirstAuthor_1">
  <rdf:type rdf:resource="#FirstAuthor"/>
  <hasKoreanName>신형수</hasKoreanName>
</owl:Thing>
<owl:Thing rdf:ID="CorrespondingAuthor_1"/>
<owl:Thing rdf:about="#CorrespondingAuthor_1">
  <rdf:type rdf:resource="#CorrespondingAuthor"/>
  <hasKoreanName>서명도</hasKoreanName>
</owl:Thing>
<owl:Thing rdf:ID="MCI"/>
<owl:Thing rdf:about="#MCI">
  <rdf:type rdf:resource="#RelatedDisease"/>
</owl:Thing>
<owl:Thing rdf:ID="TUG"/>
<owl:Thing rdf:about="#TUG">
  <rdf:type rdf:resource="#Diagnosis"/>
</owl:Thing>
    
```

(그림 4) 알츠하이머 논문 온톨로지를 이용한 지식 표현 결과를 기술한 OWL 문서

(Figure 4) an OWL document for knowledge representation of using Alzheimer disease paper ontology

본 논문에서는 스탠포드 대학에서 개발한 Protege 4.3 온톨로지 모델링 도구를 이용하여 알츠하이머 논문 온톨로지를 구축하였다.

4.2 지식 표현에 대한 검색 예

온톨로지를 이용하여 표현된 알츠하이머 논문에 대한 지식은 온톨로지 검색 언어인 SPARQL을 이용해 검색이 가능하다. SPARQL은 SQL 문법 구조와 유사하면서도 온톨로지가 제공하는 추론의 기능을 포함하고 있어 의미 기반의 검색을 쉽게 지원할 수 있는 장점이 있다.

그림 5, 그림 6, 그림 7은 4.1절에서 제시한 참고문헌 [6]의 논문에 대한 지식 표현 결과에 대해 다음과 같은 세 가지 유형의 검색 질의를 SPARQL로 표현한 것이다.

(1) 제1저자가 ‘신형수’인 논문의 국문제목을 검색
 기존의 논문 온톨로지와는 달리 본 논문에서 제안한 논문 온톨로지는 저자 클래스에 대해 제1저자, 교신저자, 공동저자의 하위 클래스를 포함하고 있다. 따라서 상위 클래스인 저자 클래스의 ‘hasKoreanName’ 프로퍼티를 제1저자가 상속받기 때문에 제1저자에 대한 검색을 그림 5와 같이 SPARQL로 표현할 수 있다.

```
SELECT ?result
WHERE {
    ?aTitle rdf:type :Title ;
            :hasKoreanTitle ?result .
    ?bContents rdf:type :Contents ;
              :hasTitle ?aTitle ;
              :hasAuthor ?cAuthor .
    ?cAuthor rdf:type :FirstAuthor ;
            :hasKoreanName "신형수" .
}
```

(그림 5) 지식 표현에 대한 첫 번째 검색 예
 (Figure 5) a first example of retrieval for knowledge representation

(2) 알츠하이머 진단방법에 대해 소개하는 논문의 제목과 그 논문에서 제시하는 진단방법이 무엇인지를 검색

기존의 논문 온톨로지와는 달리 본 논문에서 제안한 논문 온톨로지는 제목, 초록, 키워드, 결론에서 추출한 연구 주제와 관련한 클래스를 포함하고 있다. 따라서 알츠하이머 진단방법과 관련한 주제를 포함하고 있는 논문에

대한 의미 기반의 검색이 가능하며 이를 그림 6과 같이 SPARQL로 표현할 수 있다.

```
SELECT ?titleResult ?diagnosisResult
WHERE {
    ?aTitle rdf:type :Title ;
            :hasKoreanTitle ?titleResult .
    ?bContents rdf:type :Contents ;
              :hasTitle ?aTitle ;
              :hasSubject ?diagnosisResult .
    ?diagnosisResult rdf:type :Diagnosis .
}
```

(그림 6) 지식 표현에 대한 두 번째 검색 예
 (Figure 6) a second example of retrieval for knowledge representation

(3) “경도인지장애와 알츠하이머 치매 여성노인의 하 지운동기능 비교”라는 제목의 논문이 인용한 다른 논문의 주제를 검색

기존의 논문 온톨로지와는 달리 본 논문에서 제안한 논문 온톨로지는 논문 클래스들간의 인용(referTo) 관계와 같은 의미적 관계를 정의하고 있다. 따라서 사용자가 제시한 조건에 해당하는 논문이 인용하고 있는 다른 논문들에 대한 검색이 가능하며 이를 그림 7과 같이 SPARQL로 표현할 수 있다. 특히, 인용 관계는 이행적 특징을 가지고 있기 때문에 이러한 특징에 기반을 둔 추론 검색도 추가적인 조인 연산을 통해 가능하다.

```
SELECT ?result
WHERE {
    ?aTitle rdf:type :Title ;
            :hasKoreanTitle "경도인지장애와 ..." .
    ?bContents rdf:type :Contents ;
              :hasKoreanTitle ?aTitle .
    ?cPaper rdf:type :ADPaper ;
            :hasContents ?bContents ;
            :referTo ?dPaper .
    ?dPaper rdf:type :ADPaper ;
            :hasContents ?eContents .
    ?eContents rdf:type :Contents ;
              :hasKoreanTitle ?fTitle ;
              :hasSubject ?result .
    ?result rdf:type :Subject .
}
```

(그림 7) 지식 표현에 대한 세 번째 검색 예
 (Figure 7) a third example of retrieval for knowledge representation

4.3 평가

본 논문에서 제안한 알츠하이머 논문 온톨로지의 정성적 평가를 위해 4가지 측면에서 3개의 기존 연구와 비교한 결과를 표 12에 제시하였다.

본 논문에서 제안한 온톨로지와의 비교를 위해 기존 연구 결과물 중에서 논문을 대상으로 설계된 3개의 온톨로지들을 선정하였다. 선정한 첫 번째 온톨로지는 참고문헌 [10]에서 제시한 한의학 분야의 논문을 대상으로 설계된 온톨로지이다. 두 번째 비교 온톨로지는 참고문헌 [12]에서 제시한 논문 온톨로지 연구 분야에 제한을 두지 않고 설계된 것이다. 세 번째 비교 온톨로지는 참고문헌 [13]에서 제시한 컴퓨터 공학 분야의 논문을 대상으로 설계된 온톨로지이다.

(표 12) 관련 논문과의 정성적 비교
(Table 12) qualitative comparison with related papers

항목	제안 온톨로지	비교 온톨로지1	비교 온톨로지2	비교 온톨로지3
서지 정보와 관련된 지식 표현	지원함	지원함	지원함	논문 제목만 고려함
논문 내용과 관련된 지식 표현	지원함	지원함	주제별 분류만 고려함	주제별 분류만 고려함
OWL/SPARQL	지원함	고려하지 않음	지원함	지원함
프로퍼티의 의미적 관계	지원함	고려하지 않음	고려하지 않음	고려하지 않음

본 논문에서 제안한 알츠하이머 논문 온톨로지는 DBPia에서 일반 검색과 상세 검색을 위해 요청받는 서지 정보와 관련된 검색 조건들을 모두 수용하였고 논문 검색에 추가적으로 활용할 수 있는 서지 정보 관련 조건들을 추출하여 온톨로지를 구축하였다. 따라서 제안 온톨로지는 다른 비교 온톨로지들에 비해 논문의 서지 정보와 관련한 클래스들의 개수가 많아 검색 시 활용하면 사용자들의 검색 만족도를 향상시킬 수 있다.

본 논문에서 제안한 알츠하이머 논문 온톨로지는 논문의 제목, 초록, 키워드, 결론에 기술된 연구 내용과 관련하여 주요 개념들을 추출해 클래스로 정의하였다. 반면에 비교 온톨로지2와 비교 온톨로지3은 논문에 기술된 연구 내용은 고려하지 않고 단순히 논문의 연구 주제를 공학, 컴퓨터, 인공지능 등의 분야로 분류할 수 있는 개

념들만 클래스로 정의하였기 때문에 논문에 기술된 내용에 기반을 둔 검색을 지원할 수 없다.

본 논문에서 제안한 알츠하이머 논문 온톨로지는 다른 비교 온톨로지들과는 달리 클래스들간의 의미적 관계뿐만 아니라 이행적 특성과 같은 프로퍼티들간의 의미적 관계도 포함하고 있기 때문에 이러한 관계에 기반을 둔 추론 검색을 지원할 수 있다.

5. 결론

의료 분야에서는 질병의 진단과 치료를 목적으로 다양한 형태의 의료 데이터로부터 지식을 추출하여 지식베이스를 구축하려는 시도가 계속 진행되어 왔다. 지식베이스를 구축하는데 있어 가장 중요한 것은 정확하게 지식을 표현하는 것이라 할 수 있다.

본 논문에서는 최근 의료 분야에서 많은 관심을 받고 있는 알츠하이머 질병과 관련하여 국내 학술지에 게재된 논문들을 대상으로 온톨로지를 이용해 지식을 표현하는 방법을 제안하였다. 본 논문에서 제안한 알츠하이머 논문 온톨로지는 25개의 클래스를 포함하고 있으며 크게 서지 정보에서 추출한 개념들을 정의한 클래스들과 논문의 제목, 초록, 키워드, 결론에서 추출한 연구 주제와 관련한 개념들을 정의한 클래스들로 구분할 수 있다.

기존 논문 온톨로지와는 달리 본 논문에서 제안한 논문 온톨로지는 서지 정보와 연구 내용을 모두 고려하여 온톨로지를 설계하였기 때문에 서지 정보를 이용한 기본적인 검색은 물론 연구 내용과 관련한 검색이 가능하다. 그리고 29개의 프로퍼티를 정의하여 클래스의 속성은 물론 다른 클래스와의 의미적 관계를 온톨로지에 정의하였고 이행적 특성과 같은 프로퍼티들간의 의미적 관계도 정의함으로써 의미에 기반을 둔 검색을 지원할 수 있다. 그리고 온톨로지 검색 언어를 추론에 활용함으로써 지식의 의미를 보다 정확하게 이해하고 단순한 키워드 검색이 아닌 의미 기반의 지식 검색을 지원할 수 있다.

향후에는 본 논문에서 제안한 온톨로지 기반 지식 표현 방법에 따라 지식베이스를 구축하고 SPARQL과 같은 전문적인 검색 언어를 모르는 사용자도 쉽게 지식을 검색할 수 있는 시스템을 구축하는 연구를 진행하고자 한다.

참 고 문 헌(Reference)

- [1] Sookyong Lee, Youngmoon Chae, Keonwook Kang, Kyungmo Park, Chan Park, and Gonhee Kim, "The Development of Ontology-Based Dialogue System for Efficient Cancer Information Supply," Journal of The Korea Society of Health Informatics and Statistics, Vol. 30, no. 2, 2005, pp. 31-39.
- [2] Gene Ontology, <http://www.geneontology.org/>
- [3] Sangkyun Kim, Hyunchul Jang, Jinhyun Kim, Sangjun Yea, Chul Kim, Dongmyung Eum, and Miyoung Song, "A Study on Reasoning based on Herb and Formula Ontologies," Journal of The Society of Korean Medical Classics, Vol. 22, no. 3, 2009, pp. 97-105.
- [4] Kyungsil Moon and Suhyun Park, "Oriental Medical Ontology for Personalized Diagnostic Services," Journal of Korea Society of Computer Information, Vol 15, no. 1, 2010, pp. 23-30.
- [5] Byoungwon Min and Yongsun Oh, "Improvement of Personalized Diagnosis Method for U-Health," Journal of Korea Contents Association, Vol. 10, no. 10, 2010, pp. 54-67.
- [6] Hyungsoo Shin and Byungdo Seo, "Comparison of Lower Extremities Motor Function between Elderly Women with Mild Cognitive Impairment and Alzheimer's Dementia," Journal of the Korean data analysis society, Vol. 13, no. 2, 2011, pp. 659-668.
- [7] OWL 2 Web Ontology Language Primer, <http://www.w3.org/TR/owl-primer/>
- [8] Seung Hoon Choi, "Feature Configuration Validation using Semantic Web Technology," Journal of Korean Society for Internet Information, Vol. 11, no. 4, 2010, pp. 107-117.
- [9] Do-Heon Jeong, Myunggwon Hwang, Minhee Cho, Hanmin Jung, Soyong Yoon, Kyungsun Kim, and Pyung Kim, "Ontology and Text Mining-based Advanced Historical People Finding Service," Journal of Korean Society for Internet Information, Vol. 13, no. 5, 2012, pp. 33-43.
- [10] Youmi Ko and Dongmyung Eom, "The study on the design of Korean Medical Article Retrieval System Supporting Semantic Navigation based on Ontology," Korean Journal of Oriental Medicine, Vol. 11, no. 2, 2005, pp. 35-52.
- [11] Bokmoon Jung, "The Design and Implementation of a Thesis Search System Based on Ontology Inference," Master Thesis Graduate School of Korea National University of Education, 2006.
- [12] Bokkeun Sun, Dahyun We, and Kwangrok Han, "A Study on Paper Retrieval System based on OWL Ontology," Journal of Korea Society of Computer Information, Vol 14, no. 2, 2009, pp. 169-180.
- [13] Jungsuk Mun, "Computer Engineering Papers Detection System using Ontology," Master Thesis Graduate School of Kyungsoong University, 2007.
- [14] DBPia, <http://www.dbpia.co.kr>

◎ 저 자 소 개 ◎



이 재 호

1989~1996년 한국전자통신연구원 선임연구원
2013~현재 (사)한국인터넷정보학회 이사
2014~현재 (사)한국창의정보문화학회 회장
1996~현재 경인교육대학교 컴퓨터교육과 교수
관심분야 : 정보교육, 융합영재교육, 교육용 콘텐츠 개발
E-mail : jhlee@gin.ac.kr



김 연 희

2000년 홍익대학교 컴퓨터공학과 졸업(학사)
2002년 홍익대학교 대학원 컴퓨터공학과 졸업(석사)
2006년 홍익대학교 대학원 컴퓨터공학과 졸업(박사)
2007년~현재 부천대학교 e-비즈니스과 강의전담교수
관심분야 : 시맨틱 웹, 온톨로지, XML, 모바일 데이터베이스, 의료 데이터베이스.
E-mail : yhkim@bc.ac.kr



신 현 경

2002년 State University of New York at Stony Brook 대학원 응용수학(박사)
2007년~현재 가천대학교 수학교육정보학과 부교수
관심분야 : Natural language processing, Machine learning, Knowledge representation.
E-mail : hyunkyung@gachon.ac.kr



송 기 봉

1988년 서강대학교 물리학과 졸업(학사)
1995년 서강대학교 대학원 물리학과 졸업(석사)
1999년 서강대학교 대학원 물리학과 졸업(박사)
2001년~현재 한국전자통신연구원(ETRI) 책임연구원
현재 의료인지융합연구단 단장
E-mail : kbsong@etri.re.kr