

온톨로지 기반의 연구자 모델링 기법과 연구자 네트워크 구축 기법

문현정[†], 전인하^{**}, 우용태^{***}

요 약

본 논문에서는 온톨로지 기반의 연구자 모델링 기법을 제안하고, 이를 이용하여 연구자간의 다양한 연결 속성을 이용한 연구자 네트워크를 구성한다. 연구자 모델링을 위한 온톨로지 스키마는 HR-XML을 기반으로 연구자 중심의 모델로 확장하여 생성하고 연구자 온톨로지 스키마와 인스턴스는 OWL를 이용하여 생성한다. 연구자 네트워크는 학연과 같은 정적인 연결속성과 공동 논문, 프로젝트와 같은 동적인 연결속성별로 가중치를 부여하여 연구자간 친밀도를 표현한다. 친밀도는 연구자별로 연구자 네트워크가 다르므로 방향성을 가진다. SWRL를 이용하여 연구자간의 직접 관계를 추론하기 위한 규칙을 정의하고, racer 추론기를 이용하여 온톨로지 규칙을 추론한다. 본 연구에서 제안한 모델은 연구자간의 협업 모델이나 공동 프로젝트 수행을 위한 전문가 집단을 동적으로 검색하는데 응용할 수 있다.

A Researcher Model based on Ontology and a Social Network Construction Technique

Hyeon Jeong Mun[†], In Ha Jun^{**}, Yong Tae Woo^{***}

ABSTRACT

In this paper, we propose a researcher modeling technique based on ontology and construct social network for researchers using diverse relational properties. User ontology schema is created by extending the existing HR-XML model for a researcher model. User ontology schema and instance are created by OWL. We compose social network model for efficient cooperation between researchers using static relational properties such as educational background and dynamic relational properties such as co-authors and co-workers, etc. Closeness has direction because researcher network is differently configured by the researchers. We define inferencing rules using SWRL and inference ontology rules using racer inference machine to compose direct relationships between researchers. The proposed model for researchers can be applied to the cooperation model for researchers by retrieving common expert group dynamically.

Key words: ontology(온톨로지), social network(사회연결망), user modeling(사용자 모델링), researcher network(연구자 네트워크)

※ 교신저자(Corresponding Author): 우용태, 주소: 경남 창원시 사림동 9(641-773), 전화: 055)213-3813, FAX: 055)282-8218, E-mail: ytwoo@changwon.ac.kr
접수일: 2008년 10월 21일, 완료일: 2009년 4월 21일
[†] 정회원, 하이브레인넷 수석연구원
(E-mail: hjmun@hibrain.net)

^{**} 하이브레인넷 선임연구원

(E-mail: ihjun@hibrain.net)

^{***} 정회원, 창원대학교 컴퓨터공학과 교수

※ 이 논문은 2004년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.(KRF-2004-050-D00016)

1. 서 론

사회 연결망(social network)은 사람들이 일련의 관계에 의해 연결된 관계망으로, 사회학 분야에서 연구를 진행해 왔다[1,2]. 최근에는 온라인상에서 인맥 관리나 정보 교환을 위하여 사회 연결망을 구축하고 있으며, 온톨로지를 이용한 사회 연결망에 대한 연구도 진행되고 있다. FOAF(Friend Of A Friend)에서는 시맨틱 웹 기술을 적용하여 사용자간의 관계를 확장하여 친구 정보를 관리하고 이를 기반으로 사회 연결망을 구성하였다[3]. Junichiro Mori 등은 웹 마이닝 기법을 이용하여 사용자들의 정보와 관계 정보를 추출하여 사회 연결망을 구성하였다[4].

효율적인 사회 연결망 구성을 위하여 인적 자원을 효율적으로 저장, 관리 그리고 표현하기 위한 사용자 모델링 기법이 필요하다. 사용자 모델링에 관한 온톨로지 기반의 기존의 연구 방법은 RDF를 기반으로 한 ONTO-LOGGING[5], 일반 사용자 정보를 개인 정보와 감정 상태 등의 여러 가지 관점으로 관리하기 위한 GUMO(General User Modeling Ontology)[6], 인재 활용을 위해 인적 자원을 체계적으로 관리하기 위한 HR-XML[7] 등과 같은 기법이 제시되었다.

이러한 사회 연결망중에서 연구자 네트워크는 학력, 경력, 프로젝트, 논문 등과 같이 연구자들이 가지는 고유 정보를 이용하여 연구자를 연결하여 구성된 네트워크를 의미한다. 이러한 연구자 네트워크는 연구자간의 지식 공유, 연구자 검색 그리고 협업 연구를 지원하기 위해 구성한다. Junichiro Mori 등은 문서에서 연구자명이 동시에 나타난 빈도수에 기반하여 연구자간 관련성을 분석하였다[4]. 성원경 등은 협업 연구 지원을 위한 연구자 네트워크를 논문 정보 기반으로 구성하였다[8]. POLYPHONET 시스템에서는 연구자명과 키워드의 동시 출현 빈도수를 분석하고 키워드 집합이 유사한 연구자를 네트워크로 연결하였다[9].

하지만 현재까지 제시된 사용자 모델링 기법과 연구자 네트워크는 다음과 같은 문제점이 있다. 먼저, 기존의 RDF를 이용한 사용자 모델링 기법은 연구자간 연결속성 및 연결강도를 추론하기 어렵다. 즉, 이 모델은 개인의 신상정보와 취미와 같은 개인적인 성향의 정보를 관리하기 위한 사용자 모델로서, 연구자 고유의 정보를 적절하게 반영하기 어렵다. 또한, 연

구자간에 동적으로 변화하는 연결 관계를 반영한 네트워크 구성이 어렵다. 기존의 연구자 네트워크는 공저 논문과 같이 단순한 형태의 연구자 정보만을 이용하여 관계를 정의하는 관계로 연구자간의 적절한 협업 모델 구성이 어렵다.

본 논문에서는 온톨로지를 이용하여 연구자의 다양한 정보를 체계적으로 관리하기 위한 사용자 모델링 기법을 제안한다. 그리고 연구자간의 효율적인 협업 연구를 지원하기 위해 연구자 네트워크를 구축하기 위한 기법을 제안한다. 연구자 네트워크에서 연구자 네트워크별로 연구자간 연결 강도를 표현하기 위해 방향성 있는 친밀도를 제안한다. 본 연구는 크게 연구자 정보를 표현하기 위한 연구자 온톨로지 생성 과정과 연구자간의 친밀도를 가중치로 적용한 연구자 네트워크 구축 과정으로 구성된다.

먼저, 연구자 온톨로지 생성은 연구자 온톨로지 스키마 생성, 연구자 온톨로지 인스턴스 생성, 연구자 온톨로지 규칙 정의 과정으로 구성된다. 연구자 모델링을 위한 온톨로지 스키마는 인적자원관리를 위한 표준 스키마로 제안된 HR-XML을 기반으로 연구자 중심의 모델로 확장하여 OWL로 생성한다. 연구자 네트워크는 연구자간의 연결을 위하여 학연, 논문, 프로젝트와 같은 연결속성별로 친밀도를 부여하여 구성한다. 이 때, 친밀도는 연구자별 연구자 네트워크를 고려하여 계산하므로 동일 연구자라 할지라도 친밀도가 달라진다. 본 연구에서 제안한 연구자 네트워크는 연구자 네트워크의 방향성과 연구자의 속성 정보를 친밀도로 반영한 새로운 형태의 연구자 네트워크이다.

2. 관련 연구

2.1 사용자 모델링 기법에 관한 연구

기존의 사용자 모델링 기법에 대한 연구는 주로 사용자정보관리, 인적자원관리, 학습자 정보관리, 지식관리모델 등과 같은 다양한 분야에서 진행되고 있다. 먼저, GUMO는 사용자에 대한 광범위하고 구체적인 정보를 관리하기 위한 온톨로지 기반의 사용자 모델링 기법이다. GUMO에서는 이름, 나이, 생일 등의 개인 정보와 함께 심장 박동 수, 수영 실력, 감정 상태 등과 같은 세부적인 정보를 OWL을 이용하여 온톨로지로 관리한다[6]. 하지만 이 모델은 특정 용

용분야에서 이러한 정보를 구체적으로 관리하기 어려운 문제점이 있다. LIP는 학습 시스템에서 학습자 정보를 교환하기 위해 제안된 모델이다[10]. PAPI는 개인정보모델과 수행평가 정보모델을 위해 제안된 모델이다[11]. ONTO-LOGGING은 KMS(Knowledge Management System)분야에서 사용하기 위해 제안한 사용자 모델링 기법으로 LIP를 참조하여 RDF 기반의 사용자, 도메인 그리고 로그 온톨로지를 생성하였다[5].

HR-XML은 인적자원을 체계적으로 관리하기 위하여 HR-XML Consortium에서 e-biz간의 인적자원을 교환하기 위한 XML 표준안으로 제안된 모델이다. 이 모델에서는 다양한 인적자원 관리를 위해 구인구직에 관련한 75개 이상의 서로 연관된 XML 스키마의 라이브러리로 구성되어 있다. HR-XML에서는 인적자원관리를 위한 표준 XML 스키마를 제공하여 채용시스템에서 사용자 모델 구축을 위해 참조할 수 있다[7].

2.2 사회 연결망 구성에 관한 연구

사회 연결망 구성에 관한 기존의 연구는 시맨틱 웹의 발전과 함께 사용자 관계에 의미를 부여하거나 사용자간의 협업을 효과적으로 지원하기 위한 방향으로 연구되고 있다. 초기 사회 연결망에서 사용자 관계는 RDF가 가지는 정보 표현력의 한계로 인해 사용자간의 상·하위 관계 정도만 표현할 수 있었다. FOAF는 사용자 정보와 사용자간의 관계를 관리하기 위해 제안한 RDF 기반의 대표적인 사용자 모델로 사용자간의 관계를 명시적으로 입력하거나 웹마이닝 기법에 의해 사용자간의 관계를 추출하여 사회 연결망을 구성할 수 있다[3].

그리고 사용자간 연관성의 정도에 따라 가중치를 부여하여 사회 연결망을 구성하기 위한 연구도 진행되고 있다. Junichiro Mori 등은 웹 마이닝 기법을 사용하여 사용자간 연관성을 분석하여 사회 연결망을 구성하였다[4]. 국내에서는 특정 조직내에서 사용자간의 지식 공유를 위해 사회 연결망을 분석하기 위한 연구도 진행되었다[12]. 최근에 연구자간의 협업 연구를 지원하기 위하여 연구자 정보를 관리하기 위한 OntoFrame-K를 제안하였다[8]. OntoFrame-K에서는 연구자 네트워크를 구성하기 위해 논문의 공저자 수, 피인용수를 이용하여 공저 논문의 수에

따라 연구자간의 연결 강도를 분석하였다. 그리고 연구자 네트워크를 이용하여 심사자 추천시 심사자 추천 및 배제를 위한 연구도 있었다[13].

2.3 기존 연구의 문제점

기존 사용자 모델링 기법은 일반 사용자를 대상으로 매우 구체적인 정보를 관리하거나 학습, 지식 관리, 인적 자원 관리 등과 같은 분야에서 사용자를 관리하기 위한 것이 대부분이다. HR-XML에서는 구인, 구직자를 위한 정보는 구체적으로 관리하고 있으나, 연구자의 전공 분야, 연구 정보를 관리하기 위한 부분은 부족한 부분이 있다. 사용자간 관계는 관계의 유무만을 사용자가 명시적으로 표현하므로 사용자간 연결 강도를 표현하는 의미적인 관계 표현이 어렵다.

연구자 네트워크는 연구자별로 연구자 네트워크가 다르게 형성되므로 연구자간 연결 관계는 방향성에 따라 연구자간 친밀도가 달라질 수 있다. 즉, A 연구자를 기준으로 협업을 수행한 연구자와 B 연구자를 기준으로 협업을 수행한 연구자가 다르므로 A 연구자 기준에서 B 연구자에 대한 친밀도와 B 연구자 기준에서 A 연구자에 대한 친밀도는 달라질 수 있다. 하지만 기존의 연구자 네트워크에서 연구자간 관계는 논문 정보와 같은 단일 항목만을 고려하므로 연구자간 관계에 대한 방향성과 방향성에 따른 가중치는 고려하지 않고 있다. 또한, 논문의 피인용자수를 고려하여 연구자간의 협업을 위한 연구자 그룹을 검색하기 위한 연구도 있으나, 이는 연구자간의 직접적인 연결 관계를 이용한 연구자간의 연관 관계보다는 논문간의 연관 관계를 표현하는데 적합하다고 볼 수 있다.

3. 온톨로지 기반의 연구자 모델링 기법

본 논문에서는 연구자간의 협업 연구를 지원하기 위한 온톨로지 기반의 연구자 모델링 기법을 제안하고 연구자간의 친밀도를 반영한 연구자 네트워크를 구성한다. 본 연구는 크게 연구자 정보를 관리하기 위한 연구자 온톨로지 생성과 연구자간의 학연, 공저 논문, 공동 프로젝트 정보를 기준으로 친밀도를 분석하는 연구자 네트워크 구축 과정으로 구성된다. 연구자 온톨로지는 기존의 HR-XML을 참조하여 연구자

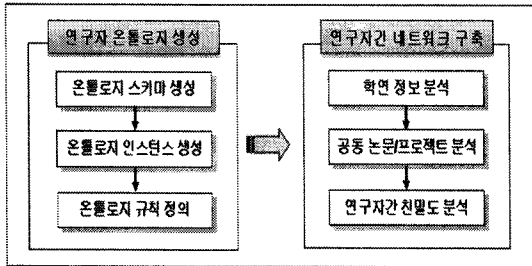


그림 1. 온톨로지 기반의 연구자 모델링 전체 개념도

중심의 정보 관리를 위한 연구자 모델로 확장하여 OWL로 생성하였다. 그림 1은 본 연구에서 제시한 온톨로지 기반의 연구자 모델링과 연구자 네트워크 구축 기법의 전체적인 개념도이다.

3.1 연구자 온톨로지 생성

본 연구에서는 온톨로지 언어인 OWL에 의해 연구자 온톨로지를 생성하는 방법을 제시한다. 연구자 온톨로지 생성은 연구자 온톨로지 스키마 생성, 연구자 온톨로지 인스턴스 생성, 그리고 연구자 온톨로지 규칙 정의 과정으로 구성된다. 연구자 온톨로지 스키마는 인적자원관리를 위한 표준 스키마로 제안된 HR-XML 스키마를 참조하여 연구자 정보를 중심으로 정의한다. 연구자 온톨로지 스키마와 인스턴스는 OWL를 이용하여 생성한다. 그리고 연구자간의 관계를 추론하기 위하여 SWRL를 이용하여 규칙을 정의한다.

3.1.1 연구자 온톨로지 스키마 생성

연구자 온톨로지 스키마 생성 과정은 연구자 온톨로지 스키마 정의와 연구자 온톨로지 스키마 생성 과정으로 구성된다. 먼저, 연구자 온톨로지 스키마는 HR-XML 스키마를 확장하여 온톨로지 클래스와 클래스간의 관계를 정의한다. 본 연구에서 제안한 연구자 온톨로지 스키마는 HR-XML에서 구직자의 정보를 관리하기 위한 Candidate 스키마의 하부 스키마 중에서 CandidateProfile과 Resume 스키마를 참조한다. CandidateProfile과 Resume 스키마는 인적자원, 경력, 이력, 소속협회, 특허실적, 자격사항, 수상사항 등의 연구자 정보를 표현할 수 있는 속성들로 구성되어 있다.

본 연구에서는 HR-XML 스키마에서 연구자 협업 정보기반의 연구자 친밀도를 분석하기 위해 연구자의

전공분야를 저장하는 MajorArea 스키마와 연구자가 수행한 프로젝트 정보를 저장하는 ProjectHistory 스키마를 추가한다. 그리고 Resume 스키마에서 논문 및 저·역서의 정보를 저장하는 PublicationHistory 스키마를 확장한다. 본 연구에서는 연구자 온톨로지 스키마를 생성하기 위하여 150개의 클래스와 클래스간의 관계 표현을 위한 117개의 object property 그리고 리터럴 값을 가지는 200개의 data property를 정의한다. 그림 2는 본 연구에서 정의한 연구자 온톨로지 스키마의 일부를 나타낸 그림이다.

3.1.2 연구자 온톨로지 인스턴스 생성

연구자 온톨로지 인스턴스 생성은 URI 기반의 인스턴스 생성을 위한 전처리 과정과 온톨로지 스키마를 이용한 인스턴스 생성 과정으로 구성된다. 먼저, 인스턴스 생성을 위한 전처리 과정은 저자명, 논문명, 프로젝트명과 같이 중복 처리될 수 있는 데이터에 대하여 고유한 식별번호를 부여하는 과정이다. 논문이나 프로젝트 데이터는 논문 제목, 게재지 등과 같은 속성 값을 이용하여 식별번호를 생성한다. 논문이나 프로젝트의 공동 연구자인 경우, 동명이인의 문제가 발생할 수 있으며, 이는 논문이나 프로젝트 URI를 이용하여 해결한다. URI 기반의 사용자 데이터는 사용자 온톨로지 스키마를 참조하여 OWL 기반의 인스턴스로 생성하였다. 다음 그림 3은 연구자 온톨로지 인스턴스 생성 과정을 나타낸 것이다.

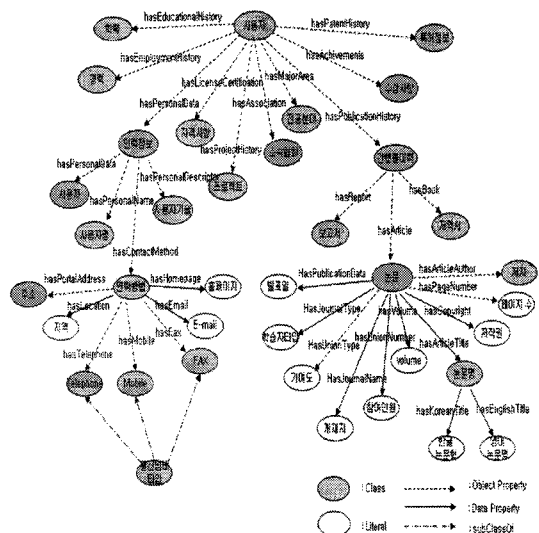


그림 2. 연구자 온톨로지 스키마의 일부

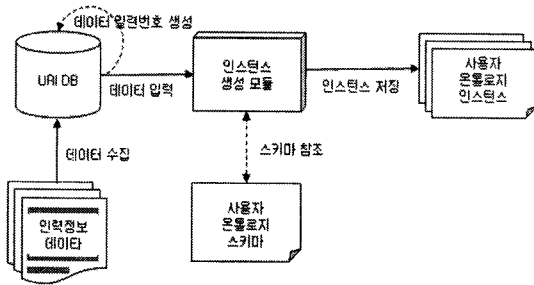


그림 3. 연구자 온톨로지 인스턴스 생성 과정

3.1.3 연구자 온톨로지 규칙 정의

본 연구에서 제안한 온톨로지 규칙은 SWRL을 이용하여 연구자간의 의미적인 추론 관계를 표현하기 위하여 정의한 규칙이다. SWRL은 OWL 지식베이스에 유사 Horn 규칙(horn-like rules)을 포함시켜 OWL의 axiom을 확장할 수 있게 하는 RuleML(Rule Markup Language)이다. 온톨로지 규칙은 연구자간의 연결속성 규칙과 연구자간의 직접 관계 추출 규칙으로 정의한다.

1) 연구자간의 연결속성 규칙 정의

연구자간의 연결속성 규칙은 연구자 클래스와 명시적인 연관 관계가 없는 클래스간의 객체속성 관계에 대하여 이행성 규칙(transitive rule)과 역관계 규칙(inverseOf rule)을 이용하여 암묵적인 연관관계를 정의한 규칙이다. 이 규칙은 공동 연구 논문이나 프로젝트 정보를 검색하기 위한 규칙과 이를 이용하여 공동 연구자를 검색하기 위한 규칙으로 정의한다. 그리고 지도교수가 동일한 연구자를 검색하기 위한 연결속성 규칙도 정의한다. 다음 규칙은 SWRL로 정의한 연구자간의 연결속성 규칙이다.

[규칙 1] 연구정보 연결속성 규칙

$hasResultHistory(?x, ?y) \wedge hasResult(?y, ?z) \rightarrow wasResultCreator(?x, ?z)$
 의미 : 연구자 ?x는 ?y라는 연구실적 이력을 가지고 ?y라는 연구실적 이력이 ?z라는 연구실적을 가지면, 연구자 ?x는 ?z라는 연구실적의 연구자이다.

[규칙 2] 연구 실적정보 연결속성 규칙

$wasResultCreator(?x, ?y) \rightarrow hasResearcher(?y, ?x)$
 의미 : 연구자 ?x가 연구실적 ?y를 보유하면, 연구실적 ?y의 연구자는 ?x이다.

[규칙 3] 지도교수 연결속성 규칙

$hasStudent(?x, ?y) \rightarrow wasAdvisor(?y, ?x)$
 의미 : 연구자 ?x는 ?y 연구자를 학생으로 가지면, 연구자 ?y의 지도교수는 ?x이다.

2) 연구자간의 직접 관계 추출 규칙 정의

연구자간의 직접 관계 추출 규칙은 동일한 연결속성 정보를 가지는 연구자간의 연관관계를 표현한 규칙이다. 즉, 이 규칙은 학연, 논문 또는 프로젝트 정보를 공통 연결속성으로 가지는 연구자를 검색하기 위한 규칙이다. 다음 규칙은 SWRL로 정의한 연구자간의 직접 관계 추출 규칙이다.

[규칙 4] 연구 실적에 의한 공동연구 관계 추출 규칙

$wasResultCreator(?x, ?y) \wedge wasResultCreator(?z, ?y) \wedge differentFrom(?x, ?z) \rightarrow hasCoResearcher(?x, ?z)$
 의미 : 서로 다른 연구자 ?x와 ?z가 논문이나 프로젝트 ?y를 공동으로 연구한 경우, ?x와 ?z는 서로 공동연구 관계이다.

[규칙 5] 동일 지도교수에 의한 학연 관계 추출 규칙

$wasAdvisor(?x, ?y) \wedge wasAdvisor(?z, ?y) \wedge differentFrom(?x, ?z) \rightarrow sameAdvisor(?x, ?z)$
 의미 : 서로 다른 연구자 ?x와 ?z에 대해 연구자 ?x의 지도교수가 ?y이고, 연구자 ?z의 지도교수가 ?y이면, 연구자 ?x와 ?z의 지도교수는 동일하다.

3.2 연구자간의 친밀도 가중치를 반영한 연구자 네트워크 구축

3.2.1 연구자별 연구자 네트워크 구성에 따른 방향성 관계

연구자 네트워크는 협업을 목적으로 하므로 연구자간 연결 유무 뿐만 아니라 친밀도를 표현하는 것이 효과적이다. 친밀도는 연구자간 공동 연구한 논문이나 프로젝트 정보를 이용하여 정의할 수 있다. 하지만 연구자별로 연구자 네트워크가 다르므로 친밀도는 같은 연결 관계에 있다 하더라도 방향성에 따라 달라진다.

다음 그림 4는 A 연구자와 B 연구자를 중심으로 공저한 논문 수를 나타낸 것이다. A 연구자 입장에서 네트워크로 연결된 연구자는 B, D, H, C이고 B 연구자와 공저한 논문의 수가 가장 많으므로 친밀도가

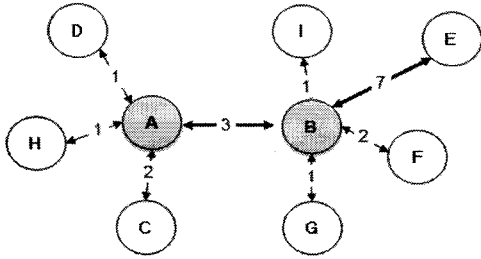


그림 4. 공저 논문으로 연결된 A, B 연구자 네트워크

가장 높다고 볼 수 있다. B 연구자 입장에서는 네트워크로 연결된 연구자는 A, I, E, F, G이고 E 연구자와 가장 많은 공저 관계에 있으므로 A 연구자와는 친밀도가 중간 수준 정도로 볼 수 있다. 본 논문에서는 연구자별 연구자 네트워크를 고려하여 연구자간 친밀도를 차별적으로 부여한다.

3.2.2 연구자간 친밀도 계산

본 논문에서 제안한 연구자간의 친밀도는 학연과 같은 정적 연결속성과 논문이나 프로젝트와 같은 동적 연결속성에 의해 계산된다. 먼저, 정적 연결속성은 학연처럼 연구자간의 관계 유무를 통해 추론 가능한 속성으로 학연 정도는 학사, 석사, 박사 과정 3가지로 구분하여 학연의 겹치는 정도에 따라 0~1사이의 값으로 정규화한다. 그리고 동적 연결속성은 공동 논문이나 프로젝트 수에 의해 추론 가능한 속성으로 연구자 네트워크별 최대수를 기준으로 0~1사이의 값으로 정규화한다. 본 논문에서는 연구자간의 정적 연결속성과 동적 연결속성을 추론하여 친밀도를 가중치로 반영하여 연구자 네트워크를 구성한다.

임의의 연구자와 직접 관계에 있는 연구자 집합 R은 $R = \{r_1, r_2, \dots, r_m\}$ 이라 정의한다. 이 때, 임의의 연구자 r_i 와 r_j 간의 친밀도는 학연, 공동 연구 논문 및 프로젝트의 수를 0~1 사이의 값으로 정규화하여 합한 값으로 정의한다. 친밀도 계산식은 다음 식(1)과 같다.

(정의 1) 연구자간의 친밀도 계산식

$$Closeness(cr_x) = \alpha + \sum_{y=1}^n \frac{a_{xy}}{a_y}, a_y = \max_{1 \leq x \leq m} (a_{xy}) \quad (1)$$

- cr_x : x번째 연구자 쌍
- α : 정적 연결속성의 값
- a_{xy} : x번째 연구자 쌍의 y번째 동적 연결속성의 값
- a_y : y번째 동적 연결속성의 최대값

표 1. R1464 연구자와 연결된 연구자와의 친밀도 계산

연구자1	연구자2	공저논문		공동 프로젝트		학연		친밀도
		수	가중치	수	가중치	수	가중치	
R1464	R2606	0	0.0	2	0.5	0	0.0	0.5
R1464	R4026	2	0.09	0	0.0	0	0.0	0.09
R1464	R0828	3	0.13	1	0.25	0	0.0	0.38
R1464	R0328	8	0.35	4	1.0	0	0.0	1.35
R1464	R5205	23	1.0	0	0.0	3	1.0	2

다음 표 1은 연구자 R1464를 기준으로 직접 연결 관계로 연결된 연구자들의 논문, 프로젝트, 학연 정도 및 가중치, 친밀도를 나타낸 것이다. 연구자 R1464와 공저한 논문 수가 가장 많은 연구자는 R5205이므로 논문 가중치가 1이 되고, 이를 기준으로 다른 연구자와의 논문 가중치를 분석하여 정규화한다. 공동으로 수행한 프로젝트수가 가장 많은 연구자는 R0328이므로 프로젝트 가중치가 1이 되고, 이를 기준으로 다른 연구자와의 프로젝트 가중치를 분석하여 정규화한다. 또한 R5205연구자와는 학사, 석사, 박사 과정의 학연정보가 일치하므로 가중치가 1이 되어, R1464와 R5205연구자와의 친밀도가 2로 가장 높다고 볼 수 있다.

4. 실험 결과 및 고찰

4.1 실험 방법

본 논문에서 제안한 기법의 효율성을 검증하기 위해서 학술진흥재단의 연구자 중에서 공개를 허용한 1,409명의 컴퓨터 관련 연구자 정보와 3,885건의 논문 데이터와 프로젝트 경력을 대상으로 실험하였다. 먼저, 연구자 모델을 구축하기 위하여 연구자 온톨로지 스키마를 생성하였다. 그리고 컴퓨터 관련 전공자의 인적사항, 지도교수, 논문 및 프로젝트 정보를 이용하여 연구자 온톨로지 인스턴스를 생성하였다. 연구자 온톨로지 스키마와 인스턴스를 생성하기 위한 온톨로지 기술 언어는 OWL를 사용하였고, 온톨로지 편집도구는 Protege 3.2를 이용하였다. 연구자 온톨로지 규칙을 정의하기 위해서 SWRL을 사용하였고, 온톨로지 규칙 추론은 racer 추론기를 이용하였다.

4.2 연구자 네트워크 구성에 따른 친밀도 비교

4.2.1 연구자간 연결속성 규칙에 기반한 직접 관계 추출

본 연구에서는 연구자간의 친밀도를 계산하기 위하여 직접 관계 추출 모듈을 생성하였다. 연구자간 직접 관계 추출 모듈은 Jena와 SPARQL을 기반으로 생성한다. 연구자 온톨로지에 저장된 1,409명간의 직접 관계를 분석한 결과, 한 연구자가 평균 2명 이상의 연구자와 직접적인 관계로 연결되는 것을 알 수 있다.

공저 논문 수에 의해 연구자 네트워크를 구성한 기존의 연구에서는 1,631명이 직접 관계로 연결되었지만, 본 연구에서 제안한 정적 연결속성과 동적 연결속성을 이용하여 관계를 구성한 경우에는 2,287명이 직접 관계로 연결되었다. 또한 기존 연구에서는 단순히 공저 논문 수에 의해 연구자 네트워크를 구성한다. 본 연구에서는 학연, 논문, 프로젝트와 같은 다양한 연결속성에 의해 연구자별로 동적으로 연구자 네트워크를 구성한다. 표 2는 연결속성별 데이터 및 연결 관계수를 나타낸다. 전체 연구자 수 1,409명중에서 학력 정보를 가진 연구자 수는 1,391명이고, 논문 및 프로젝트 정보는 각각 3,885, 5,812건이다. 학연 기반의 연구자간 직접 연결 관계수는 595건이고, 논문 기반의 직접 연결 관계수는 1,631건, 프로젝트 기반의 직접 연결 관계수는 181건이다. 학연, 논문, 프로젝트 기반의 직접 연결 관계수는 2,287건이다.

4.2.2 연구자 네트워크의 방향성 관계에 따른 논문 친밀도 비교

친밀도는 공저 논문 수처럼 임의의 두 연구자간에 고정된 값을 가지지 않고 특정 연구자를 중심으로 관련된 네트워크에 따라 서로 다른 값을 가진다. 다음 그림 5는 임의의 두 연구자간의 공저 논문 수와 본 연구에서 제안한 연구자 네트워크 방향성 관계에 따른 친밀도를 비교한 결과이다.

친밀도는 (식 1)에 의해 특정 연구자가 연결 관계에 있는 연구자 중 공저한 최대 논문 수를 기준으로

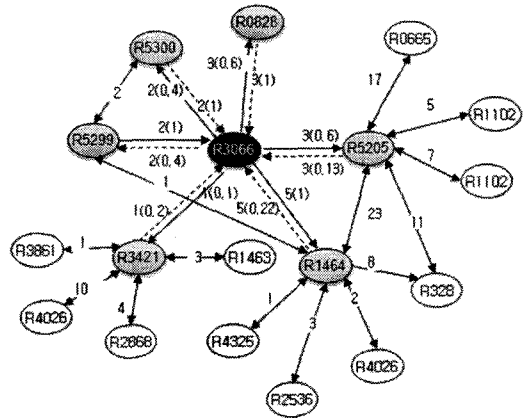


그림 5. 연구자 네트워크의 방향성 관계에 따른 논문 친밀도

임의의 연구자와 공저한 논문 수를 정규화하여 계산한다. 즉, R3066 연구자와 R1464 연구자의 네트워크에서 R3066 연구자의 최대 공저 관계는 5편의 논문을 공저한 R1464이므로 R3066연구자와 R1464연구자에 대한 친밀도는 1이 된다. R1464 연구자는 모든 논문을 R5205와 공저한 관계로 R3066 연구자에 대한 친밀도는 0.22가 된다. 그림 5에서 간선위의 숫자는 논문 수를 의미하고, ()의 값은 방향성 관계에 따른 논문 친밀도를 나타낸다.

4.2.3 연구자간의 친밀도를 반영한 연구자 네트워크 구성

표 3은 본 논문에서 제안한 다양한 연결속성을 이용하여 연구자간의 직접 연결 관계와 친밀도를 분석한 결과이다. 1번의 경우는 프로젝트 연결 속성에 의해 직접 연결 관계를 구성한 것이고, 2번 경우는 학연에 의해서도 직접 연결 관계를 구성할 수 있음을 나타낸다. 3번 경우는 논문 수는 동일하나 프로젝트 친밀도에 의해 전체 친밀도가 높아짐을 보여주고, 4번은 학연 친밀도에 의해 전체 친밀도가 높아짐을 보여준다. 마지막으로 5번의 경우는 논문 수는 동일하나 프로젝트와 학연 관계에 의해 친밀도가 높아짐을 보여준다.

본 실험에서는 연구자 온톨로지를 기반으로 연구자간의 다양한 연결속성을 분석하여 연구자 네트워크를 구성한다. 연구자 네트워크는 본 연구에서 제안한 직접 관계 추론 규칙을 이용하여 논문, 프로젝트, 학연과 같은 연결속성에서 직접 관계를 분석하여 친밀도를 부여한 네트워크이다. 다음 그림 6은 연구자

표 2. 연결속성별 데이터 및 연결 관계수

연결속성 수	학연	논문	프로젝트	전체
데이터 수	1,391(명)	3,885(건)	5,812(건)	1,409(명)
연결 관계수	595(건)	1,631(건)	181(건)	2,287(건)

표 3. 다중 연결속성을 고려한 직접 연결 관계 및 친밀도 비교

No	분석기준	연구자1	연구자2	공저논문		공동프로젝트		학연		친밀도
				수	가중치	수	가중치	수	가중치	
1	프로젝트로 연결한 경우	R0010	R0252	0	0.0	2	1.0	0	0.0	1.0
2	학연으로 연결한 경우	R0147	R5319	1	0.1	0	0.0	0	0.0	0.1
3	프로젝트 가중치의 친밀도에 대한 영향	R1464	R0328	8	0.35	1	1.0	0	0.0	1.35
		R1464	R0665	8	0.35	0	0.0	0	0.0	0.35
4	학연 가중치의 친밀도에 대한 영향	R0689	R2062	2	1.0	0	0.0	3	1.0	2.0
		R0689	R0010	2	1.0	0	0.0	0	0.0	1.0
5	프로젝트, 학연 가중치의 친밀도에 대한 영향	R4495	R3017	2	0.67	0	0.0	0	0.0	0.67
		R4495	R3703	2	0.67	1	1.0	3	1	2.67

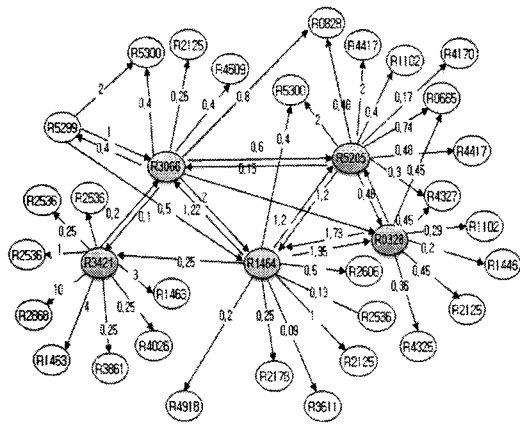


그림 6. 연구자간의 친밀도 가중치를 반영한 연구자 네트워크 구성 일부

R0328, R1364, R3066, R3421, R5205를 기준으로 연구자 네트워크를 구성한 것의 일부이다.

이러한 연구자 네트워크는 특정 연구자를 중심으로 동적으로 구성하여, 연구자간의 협업 모델이나 공동 프로젝트 수행을 위한 연구자 집단을 검색하는데 응용할 수 있다.

5. 결론

본 논문에서는 온톨로지 기반의 연구자 모델링 기법을 제안하고, 이를 이용하여 연구자간의 다양한 연결속성을 이용한 연구자 네트워크를 구성한다. 연구자 모델링을 위한 온톨로지 스키마는 HR-XML을 기반으로 연구자 중심의 모델로 확장하여 OWL로 생성한다.

연구자 네트워크는 연구자간의 효율적인 협업 연

구를 위해 학연과 같은 정적인 연결속성과 공동 논문, 프로젝트와 같은 동적인 연결속성별로 가중치를 부여하여 구성한다. SWRL를 이용하여 연구자간의 직접 관계를 추론하기 위한 규칙을 정의하고, racer 추론기를 이용하여 연구자간 직접 관계를 추론한다. 본 연구에서 제안한 연구자 네트워크는 연구자간에 다양한 연결속성을 가중치로 적용하여 연구자간에 상세한 친밀도를 반영하여 보다 현실적인 연구자 네트워크를 구성할 수 있다.

본 논문에서 제안한 기법의 효율성을 검증하기 위해 1,409건의 컴퓨터 관련 연구자와 3,885건의 컴퓨터 관련 논문 데이터와 프로젝트 경력을 대상으로 실험하였다. 연구자 온톨로지에 저장된 1,409명간의 직접 관계를 분석한 결과, 연구자는 평균 2명 이상의 다른 연구자와 학연 또는 논문에 의해 직접 관계로 연결되는 것을 알 수 있다. 공저 논문에 의해 연구자 네트워크를 구성한 기존의 연구에서는 1,687명이 직접 관계로 연결되었지만, 본 연구에서 제안한 정적 연결속성과 동적 연결속성을 이용하여 관계를 구성한 경우에는 2,287명이 직접 관계로 연결할 수 있다.

본 연구에서는 연구자별로 연구자 네트워크가 다르게 형성되는 점을 고려하여 관점에 따라 연구자간 친밀도가 다르게 계산됨을 보인다. 또한, 학연처럼 정적인 연결속성과 공저 논문이나 프로젝트처럼 시간에 따라 변화하는 동적인 연결속성을 이용하여 다면적인 친를 구성할 수 있음을 보인다. 결과적으로 연구자별 네트워크내에서 다양한 연결속성을 가중치로 적용할 경우 연구자간에 상세한 친밀도를 반영하여 보다 현실적인 . 본 논문에서 제안한 기법은 전문가나 연구자들의 인적자원관리 및 연구 집단 검색

을 통한 협업 모델 구성에 응용할 수 있으리라 기대된다. 그리고 지식시장을 활성화하기 위한 지식거래 시스템에서 유사 전공자의 검색 및 커뮤니티 구성에 기초적인 정보를 제공할 수 있으리라 본다.

참 고 문 헌

[1] S. E. Seibert, M. L. Kraimer, and R. C. Liden, "A Social Capital Theory of Career Success," *The Academy of Management Journal*, Vol. 44, No.2, pp. 219-237, 2001.

[2] 김지수, "디지털 인맥과 인간 관계망(human network)," *정보통신정책*, 제16권, 제16호, pp. 1-19, 2004.

[3] L. Razemertia, "User Models and User Modeling in Knowledge Management Systems," *Proc. of the User Modeling Conference*, Springer-Verlag, pp. 213-217, 2003.

[4] D. Heckmann, B. Brandherm, M. Schmitz, T. Schwartz, and B.M. von Wilamowitz-Moellendorf, "Gumo - the general user model ontology," *Lecture Notes in Computer Science* 3538, pp. 428-432, 2005.

[5] HR-XML Consortium, HR-XML Consortium Library, <http://www.hr-xml.org/>

[6] T. Finin, L. Ding, and L. Zou, "Social Networking on the semantic Web," *Journal of the Learning Organization*, Vol.12, No.5, pp. 418-435. 2005.

[7] J. Mori, Y. Matsuo, K. Hasida, and M. Ishizuka, "Web Mining Approach for a User-centered Semantic Web," *Workshop on End User Aspects of the Semantic Web*, 2005.

[8] 성원경, 정한민, 박동인, "OntoFrame-K:협업 연구 지원을 위한 시멘틱 웹 기반 지식정보 공유·유통 플랫폼," *한국정보과학회논문지*, 제 33권, 제4호, pp. 65-74. 2006.

[9] Y. Matsuo, J. Mori, M. Hamasaki, K. Ishida, T. Nishimura, H. Takeda, K. Hasida, M. Ishizuka, "POLYPHONET: An advanced so-

cial network extraction system from the Web," *Proc. of the 15th Int'l conference on World Wide Web*, pp. 397-406, 2006.

[10] IMS, IMS learner information package specification, <http://www.imsproject.org/profiles/index.cfm>

[11] IEEE, draft standard for learning technology. public and private information (papi) for learners (papi learner), <http://ltsc.ieee.org/wg2/>

[12] 서치중, 서의호, "Knowledge Network 구축 수단으로서의 Social Network Analysis에 관한 연구," *한국경영과학회 학술대회논문집 논문집*, pp. 317-322, 2002.

[13] 이정연, 이재운, 정한민, 강인수, 신숙경, "확률적 온톨로지와 연구자 네트워크를 이용한 심사자 자동 추천에 관한 연구," *한국정보관리학회 정보관리학회지*, 제 24권, 제 3호, pp. 43-65, 2007.



문 현 정

1994년 한국방송대학교 전자계산학과 이학사
 1996년 창원대학교 전자계산학과 이학석사
 2003년 창원대학교 컴퓨터공학과 공학박사
 2004년~2007년 창원대학교 연구교수

2007년~현재 하이브레인넷 수석연구원
 관심분야 : KDD, 온톨로지마이닝, 시멘틱웹



전 인 하

2005년 고신대학교 전산수학과 이학사
 2007년 창원대학교 컴퓨터공학과 공학석사
 2007년~현재 하이브레인넷 선임연구원

관심분야 : 온톨로지 모델링, 온톨로지 추론시스템, 암호화, 네트워크



우 용 태

- 1982년 경북대학교 전자공학과
공학사
- 1984년 경북대학교 전자공학과
공학석사
- 1995년 경북대학교 전자공학과
공학박사
- 1987년~현재 창원대학교 컴퓨터
공학과 교수

관심분야 : 데이터마이닝, 온톨로지마이닝, 시맨틱웹