

---

# 멀티 온톨로지 기반의 키워드 연관성을 이용한 전문가 검색 시스템

정계동\* · 황치곤\*\* · 최영근\*\*

The Expert Search System using keyword association based on Multi-Ontology

Kye-dong Jung\* · Chi-gon Hwang\*\* · Young-keun Choi\*\*

---

이 논문은 2011년도 광운대학교 연구비를 지원받았음

---

## 요 약

본 연구는 연구논문 및 저자 프로파일을 기반으로 상호 협력이 가능한 전문가 검색 시스템을 구축한다. 제안한 방법론은 다음과 같다. 첫째, 입력 키워드와 가장 연관성 높은 키워드를 검색하기 위한 가중치 부여 기법을 제안하고, 둘째, 이 기법을 통해 전문가를 효율적으로 검색하는 방안을 제안한다. 우선적으로 논문에서 키워드와 저자 프로파일을 추출하고, 이를 통하여 전문가를 검색할 수 있도록 한다. 이것은 소셜 네트워크의 여러 분야에서 활용할 수 있다. 이러한 정보는 여러 시스템에 분산되어 있다. 이렇게 분산된 데이터를 통합하기 위한 기술로 멀티 온톨로지를 이용하는 기법을 제안한다. 멀티 온톨로지는 메타 온톨로지, 인스턴스 온톨로지, 로케이션 온톨로지와 연관관계 온톨로지 구성되고, 연관관계 온톨로지는 동적으로 키워드 연관관계 분석을 통해 구축된다. 이 멀티 온톨로지를 이용하여 전문가 망을 제공하고, 이것은 키워드의 연관관계 추적을 통한 전문가 검색이 가능하도록 한다. 이를 통하여 전문가들의 연구물을 확인할 수 있도록 제공함으로써 세부 전문분야를 확인할 수 한다.

## ABSTRACT

This study constructs an expert search system which has a mutual cooperation function based on thesis and author profile. The proposed methodology is as follows. First, we propose weighting method which can search a keyword and the most relevant keyword. Second, we propose a method which can search the experts efficiently with this weighting method. On the preferential basis, keywords and author profiles are extracted from the papers, and experts can be searched through this method. This system will be available to many fields of social network. However, this information is distributed to many systems. We propose a method using multi-ontology to integrate distributed data. The multi-ontology is composed of meta ontology, instance ontology, location ontology and association ontology. The association ontology is constructed through analysis of keyword association dynamically. An expert network is constructed using this multi-ontology, and this expert network can search expert through association trace of keyword. The expert network can check the detail area of expertise through the research list which is provided by the system.

## 키워드

소셜 네트워크, 멀티 온톨로지, 연관 온톨로지, 전문가 망, XMDR(eXtended MetaData Repository)

## Key word

Social Network, Multi Ontology, Association Ontology, Expert Network, XMDR(eXtended MetaData Repository)

---

\* 정회원 : 광운대학교 컴퓨터과학과(교신저자, gdchung@kw.ac.kr)

\*\* 정회원 : 광운대학교 컴퓨터과학과

접수일자 : 2011. 09. 05

심사완료일자 : 2011. 10. 10

## I. 서 론

온라인 소셜 네트워크는 웹이 활성화되기 시작한 2000년 초에 등장했다. 이것은 사용자들의 오락적 혹은 비즈니스적인 인적 네트워크를 구성하는 데 도움을 주었다. 이것은 오프라인 소셜 네트워크와 달리 소셜 네트워크 형성을 다른 비즈니스 모델을 구현하는 데 필요한 수단으로 활용하기 시작하였다. 검색 포털 사이트에서는 이용자 중 일부만이 논문의 정확한 서지사항을 직접 입력하여 자료를 검색하며, 대부분의 이용자는 키워드 검색을 통한 반복적인 시행착오를 거쳐서 적합한 논문을 찾게 되기 때문이다. 그리고 온라인 쇼핑물, 검색 포털 등에 널리 사용되어 온 연관관계 분석을 통한 키워드 추천 방법이 일부 학술정보 포털 서비스에도 최근 적용되고 있다. 그러나 이 방법은 키워드가 서로 연관성이 있다는 것만 알려줄 뿐 키워드 간의 연결 분포와 중심 키워드는 제시하지 못하고 있다[1]. 현재 서비스되고 있는 대부분의 소셜 네트워크 사이트는 사용자와 관련된 정보 자원 및 개체들 간의 관계에 대하여 명확하게 정의하고 있지 않아 의미적 관계를 발견하지 못하는 한계가 있다. 이러한 문제점은 사회적 개체와 이들 간의 관계에 관련된 도메인 온톨로지(ontology), 적절한 규칙(rule), 그리고 추론 메커니즘을 통해 새로운 정보를 추론하여 새로운 관계나 이전에 존재하지 않았던 새로운 개념을 발견해냄으로써 극복될 수 있다[2][3].

본 논문에서는 특정 키워드를 연구 내용으로 하는 연구자를 지능적으로 추천하여 효율적으로 소셜 네트워크를 구축할 수 있는 방안으로 연구자들의 망을 구축하여, 이를 전문가 망이라 하고, 온톨로지를 이용한 지능적 기법으로 XMDR[4]을 적용하고자 한다. XMDR에서 제공하는 온톨로지를 이용하여 데이터의 이질성 관계를 해결하고, 키워드의 연관관계를 분석하여 전문가 망과 연계한다. 이를 통해 얻을 수 있는 효과는 다음과 같다. 먼저, 비전공 분야의 사용자가 특정 분야의 연구 결과물을 찾기 위한 초기 검색 키워드에 대한 로드맵을 제공할 수 있다. 둘째, 사용자가 입력한 키워드에 대한 분석으로 키워드들 간의 연관관계를 제공할 수 있다. 셋째, XMDR의 온톨로지를 활용하여 특정 키워드에 대한 동의어의어, 이음동의어, 유사어와 같은 연관관계와 이질성으로 인해 발생하는 문제를 해결할 수 있

다. 넷째, 사용자가 요구하는 키워드를 기준으로 키워드 간의 연관관계를 제공하는 방법으로 가중치 부여 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장은 관련연구로 키워드 연관성, 데이터 마이닝, 추천시스템 그리고 분산 시스템의 연관성 분석을 통한 데이터 통합을 위한 저장소인 XMDR에 대한 기존 연구들을 간략하게 기술한다. 3장은 제안 모델과 연구 범위, 그리고 제안하는 방법론을 기술한다. 4은 제안한 기법에 대한 적용과 평가에 대해 기술한다. 그리고 5장은 본 연구의 결과 및 한계 그리고 향후 연구에 대해 기술한다.

## II. 관련연구

본장에서는 본 논문을 위해 필요한 시맨틱 소셜 네트워크, 추천 기법과 XMDR에 대해 본다. 최근 소셜 네트워크에 시맨틱 웹 기술을 이용해서 시맨틱 소셜 네트워크를 구축하려는 여러 연구가 진행되어 왔다. 이러한 기술로는 FOAF[5]이 있다. FOAF는 시맨틱 웹 기술을 통하여 관계성을 확장함으로써 소셜 네트워크를 구축하는 기술이고, XFN은 하이퍼링크에 특정 속성을 추가함으로써 개인과의 관계를 표현할 수 있도록 하여 소셜 네트워크를 구축하는 기술이다. 이러한 기술을 바탕으로 최근 소셜 네트워크에 시맨틱 웹 기술을 이용해서 시맨틱 소셜 네트워크를 구축하려는 여러 연구가 진행되어 왔다[6]. 기존의 온톨로지가 개념-인스턴스의 양자(bipartite) 모델이었다면, 여기에 사용자가 추가된 3자(tripartite) 모델을 이용해서 시맨틱 소셜 네트워크 온톨로지를 구축함으로써 시간의 변화에 따른 사용자의 활동을 모델링할 수 있도록 하였다[7].

추천 시스템에서 사용되는 추천 기법으로는 크게 협업필터링, 내용기반 추천, 등의 두 가지 형태로 분류할 수 있다. 협업필터링(Collaborative filtering) 추천 기법은 사용자 사이의 연관성을 기반으로, 선호도 또는 구매 패턴이 유사한 고객군을 분류하고, 유사한 고객군에 속하는 다른 사람들이 선호하는 상품을 추천하는 방법이다[8]. 그러나 신상품의 경우에는 선호도를 예측할 수 없으므로 추천이 불가능하게 되는 신상품 추천 문제가

발생한다[9]. 한편 정보 검색과 정보 필터링으로부터 발전된 내용기반(Content-based)추천 기법은 상품이 지닌 속성을 파악하고 고객의 구매 이력을 바탕으로 고객이 구매한 상품의 속성들과 유사한 상품을 추천하는 방법으로서 제품 정보를 중심으로 추천하는 방법이다. 하지만 이러한 추천 방식은 사용자의 과거 선호도를 학습하여 추천 아이টে를 결정하기 때문에, 선호도가 자주 변하거나 새로운 내용이 지속적으로 생성되는 환경에서는 추천의 성능이 떨어지며 다른 영역에 대한 추천이 어렵다[10].

XMDR은 MDR을 보다 더 확장한 개념으로 이전의 MDR에 전문 용어, 온톨로지, 시맨틱 웹 관리 기능을 확장한 것이다. XMDR의 각 구성 요소들은 ISO/IEC 11179-3에서 제안한 데이터의 속성 명세를 따른다. 실제 데이터는 표현의 차이로 서로 다른 데이터로 인식될 수 있다.

이러한 데이터들의 연관성과 의미를 파악하여 온톨로지 생성한 것이다[4]. 본 연구에서는 XMDR을 구성하는 멀티 온톨로지를 접근하고 관리하기 위한 에이전트들의 집합을 DIMs(Data Integration Managers)라 하고, 이것을 전문가 망을 구축하기 위해 전문가들의 관계성을 분석하여 관리하는데 이용하고, 키워드의 연관성을 분석을 통하여 사용자가 원하는 분야에 대한 전문적 검색을 지원하고, 분산된 데이터들 간의 이질적 문제를 해결하고자 한다.

### III. 키워드 연관성을 이용한 전문가 검색 시스템

#### 3.1. 멀티 온톨로지 기반의 전문가 검색시스템

본 논문에서 제안되는 전체 시스템 구조는 그림 1과 같다. 이것은 서비스 영역(Service Area)와 멀티 온톨로지 영역(Multi Ontology Area)로 나뉜다. 서비스 영역은 그림 1과 같이 세 개의 레이어로 구성되고, 멀티 온톨로지 영역은 4개의 온톨로지 구성한다. 각 세부 사항은 다음과 같이 구성된다.

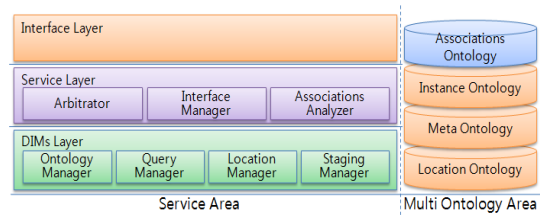


그림 1. 멀티 온톨로지 기반의 DIMs 시스템의 구성  
Fig. 1 The Composition of DIMs System based on Multi Ontology

#### 3.1.1. 서비스 영역(Service Area)

서비스 영역은 DIMs 계층(DIMs Layer), 서비스 계층(Service Layer), 인터페이스 계층(Interface Layer)로 구성된다. 멀티 온톨로지를 접근하고 관리함으로써 사용자가 분산된 데이터들을 하나의 시스템을 접근하는 것처럼 사용하게 하는 에이전트들의 집합을 DIMs라 한다. 이것은 분산된 다중 시스템을 하나의 시스템과 같이 접근할 수 있도록 각 시스템에 적합한 질의어를 생성하고, 분석된 연관 키워드들도 고려할 수 있도록 해야 한다. 이것의 구성은 온톨로지 관리자(Ontology Manager), 쿼리 관리자(Query Manager), 로케이션 관리자(Location Manager), 스테이징 관리자(Staging Manager)로 구성된다.

- 온톨로지 관리자(Ontology Manager) : 시스템에서 사용되는 온톨로지는 메타, 인스턴스, 로케이션 그리고 릴레이션 온톨로지가 있다. 이 온톨로지를 관리하고 온톨로지의 내용을 검색하여 제공하는 역할을 수행한다.
- 쿼리 관리자(Query Manager) : 분산된 시스템의 데이터를 통합하기 위한 방안으로 글로벌 질의를 사용하는 데 이를 로컬 질의로 분해하고, 로컬에서 수집된 결과를 저장하기 위해 스테이징 저장소에 수집용 테이블을 생성하는 역할을 수행한다.
- 로케이션 관리자(Location Manager) : 시스템에 참여하는 로컬 시스템의 접근에 필요한 정보를 관리하고, 글로벌 쿼리를 로컬 쿼리로 변환할 때 로컬 시스템의 접근 방법을 제공하는 역할을 담당한다.
- 스테이징 관리자(Staging Manager) : 쿼리 관리자가 요구하는 임시테이블 생성하고, 로컬 시스템에서 수집된 결과를 해당 임시테이블에 축적하고, 축적된 결과를 서비스 중재자에게 제공하는 역할을 담당한다.

서비스 계층은 전반적으로 사용자가 입력한 키워드와 연관된 키워드들을 논문 검색을 통해서 수집하여 분석하는 역할을 수행하는 것으로 다음과 같다.

- 서비스 중재자(Service Arbitrator) : 전반적인 시스템의 운용의 관리 및 중재 역할을 담당한다.
- 인터페이스 관리자(Interface Manager) : 인터페이스를 생성하고 인터페이스에서 입력된 데이터를 서비스 중재자에게 넘겨주고, 서비스 중재자가 제공하는 결과를 인터페이스에 표시하는 역할을 담당한다.
- 연관관계 분석기(Associations Analyzer) : 사용자가 입력한 키워드에 해당하는 논문을 통하여 사용자가 입력한 키워드와 키워드를 통해 수집된 논문들의 키워드간의 연관관계를 분석하여 효율적인 검색이 될 수 있도록 지원한다.

인터페이스 계층은 사용자와 시스템을 연결해주는 응용 어플리케이션 계층이다.

### 3.1.2. 멀티 온톨로지 영역(Multi Ontology Area)

제한하는 시스템을 운영하기 위한 에이전트들은 멀티 온톨로지 위에서 수행된다. 멀티 온톨로지라 함은 인스턴스 온톨로지, 메타 온톨로지, 로케이션 온톨로지를 기본적으로 이용하고, 연관관계 온톨로지를 추가한 것이다.

- 연관관계 온톨로지(Associations Ontology) : 사용자가 입력한 키워드를 가지는 논문들을 수집하여, 수집된 논문들에서 가지고 있는 키워드를 추출·분석을 통해 연관관계를 분석하여 저장한 키워드 연관관계 지식 저장소이다.
- 인스턴스 온톨로지(Instance Ontology) : 데이터가 가지는 값들 사이의 이질적 문제를 해결하기 위한 지식 저장소이다.
- 메타 온톨로지(Meta Ontology) : 각 시스템에서 가지고 있는 메타 데이터에서 발생하는 의미적 이질성과 구조적 이질성을 해결하기 위한 지식 저장소이다.
- 로케이션 온톨로지(Location Ontology) : 분산된 시스템에 접근하기 위한 통신 정보, 접근 경로, 보안 정보 및 권한 정보를 관리하는 저장소로 다른 온톨로지와 연관되어 사용되기 때문에 단순한 목록이라 볼 수 있지만, 다른 온톨로지들과 같이 맞물려 사용되므로 온톨로지라 한다.

### 3.2. 연관관계 분석 기반의 키워드 추천

본 장에서는 입력된 키워드에 대한 논문의 키워드를 추출하여 키워드와 키워드 간의 연관관계를 분석하는 방안에 대해서 기술한다. 연관관계 분석은 사용자가 입력한 키워드를 기준으로 해당 키워드를 포함하는 논문을 검색하여 검색된 결과에 대한 키워드를 추출하여 키워드간의 연관관계를 통계적 방법으로 분석한다. 해당 키워드와 통계적 연관성이 가장 높은 키워드를 이용하여 검색을 수행하면 검색 결과의 정확도를 향상시킬 수 있게 된다. 이를 위해 통계적 가중치 방법으로 데이터 마이닝 기법에서 많이 사용하는 Tf-idf[11] 가중치 방식을 이용하여 키워드간의 연관관계를 분석한 가중치 모델을 다음과 같이 정리한다.

사용자 키워드를  $uK$ 라 하고,  $uK$ 를 가지는 논문을  $p$ 라 하고, 각 논문  $p(j=1, 2, 3, \dots, n)$ 가 가지는 키워드 수는  $n(pKey)$ 라고 한다.  $uK$ 를 가지는 논문의 모든 키워드의 개수는 각 논문에 있는 키워드 개수의 합이 된다. 이를 정리하면 수식(1)과 같다.

$$\sum_j^n n(pKey)_j \tag{1}$$

그리고  $uK$ 와 같이 포함된 키워드들을 연관키워드이다. 이 키워드를  $aK$ 라 하고,  $uK$ 와 연관관계를 가지는  $i$  번째 키워드의 개수를  $n(aK)_i$ 로 정의한다. 이를 통해 연관 키워드의 빈도수( $KF_i$  :  $i$ 'th Key Frequency)는 수식(2)로 정리된다.

$$KF_i = \frac{n(aK)_i}{\sum_j^n n(pKey)_j} \tag{2}$$

$i$  번째 연관 키워드를 포함하는 논문의 숫자를  $n(PaK_i)$ 라 하고,  $uK$ 를 가지는 논문의 편수를  $nP$ 라 할 때,  $i$  번째 연관 키워드가 포함된 논문의 역 빈도수( $IPF_i$  :  $i$ 'th Inverse Paper Frequency)는 수식(3)과 같다.

$$IPF_i = \log \frac{nP}{n(PaK_i)} \tag{3}$$

이를 정리하여  $i$  번째 연관 키워드의 연관성 가중치

( $Aw_i$ )는 수식(4)로 정리된다.

$$Aw_i = KF_i \times IPF_i \quad (4)$$

이렇게 정리된 수식을 통해  $uK$ 를 포함하는 논문들이 가지는 키워드들에 대한 가중치를 분석하여,  $uK$ 와 연관성 가중치가 높은 우선순위에 따라  $aK$ 의 순위를 부여할 수 있고, 이를 사용자에게 제공하여 검색의 정확도를 향상시킬 수 있다.

이러한 가중치 평가방법을 적용하기 위해 디비피아(www.dbpia.com)에서 “온톨로지”를  $uK$ 로 하여 검색한 결과를 위에서 제시한 방식으로 연관관계 분석하였다. 이를 분석한 가중치는 표 1과 같다. 이는 수식 (1), (2), (3), (4)에 따른 연산 결과이며, “온톨로지”에 대한 연관 키워드 리스트를 구하고, 검색된 논문에서 각 연관키워드의 포함횟수( $n(aK)_i$ ), 키워드 빈도수( $KF_i$ ), 키워드가 포함된 논문의 역 빈도수( $IPF_i$ ) 그리고 연관성 가중치( $Aw_i$ )를 구한 것이다. 이를 오름차순에 따라 정리하여 왼쪽의 키워드부터 오른쪽 끝의 키워드까지 연관 정도에 따라 나열되어 있다. 이 연관관계 분석 정보는 연관관계 온톨로지에 저장하고, 새로운 검색에 이용할 수 있다.

표 1. “온톨로지”에 대한 연관관계 키워드 분석  
Table. 1 The Analysis of association keyword about keyword "Ontology"

$aK_i$	$n(aK)_i$	$KF_i$	$IPF_i$	$Aw_i$
시맨틱	377	0.083	1.32	0.109
owl	173	0.038	2.443	0.093
rdf	94	0.021	3.323	0.068
에이전트	89	0.02	3.402	0.066
메타데이터	84	0.018	3.486	0.064
xml	73	0.016	3.688	0.059
키워드	67	0.015	3.812	0.056
시소러스	32	0.007	4.878	0.034
토픽맵	29	0.006	5.02	0.032

### 3.3. 멀티 온톨로지 기반의 전문가 검색

본 장에서는 멀티 온톨로지를 이용한 전문가 검색 방안에 대해서 기술한다. 전문가를 찾기 위해 사용자가 입력한 키워드 정보를 통한 전문가의 논문 정보 검색은 입

력한 키워드로 표준 질의로 생성하고, 이는 해당 로컬 시스템에 적합한 로컬 질의로 변환되는 것으로 1차 질의를 수행하는 기본 과정이다. 1차 질의의 수행 결과는 키워드 연관관계 분석을 위하여 수집된 논문의 키워드만 추출하여 사용자가 입력한 키워드와 검색된 결과 논문들이 가지고 있는 키워드들을 분석하여 입력한 키워드와 수집된 키워드들 사이의 연관관계를 3.2에서 기술한 기법으로 수행한다. 수행한 결과를 통해 2차 질의를 수행하여 사용자에게 입력한 키워드와 연관관계가 높은 키워드를 포함한 질의결과를 제공한다. 이를 통하여 사용자가 원하는 정보에 대한 정확도를 향상시킬 수 있도록 한다. 이에 3차 질의는 사용자가 찾은 전문가를 확인하는 과정으로, 사용자가 선택된 전문가의 논문을 확인할 수 있도록 한다. 이 과정에 대한 전반적인 과정은 그림 2와 같다.

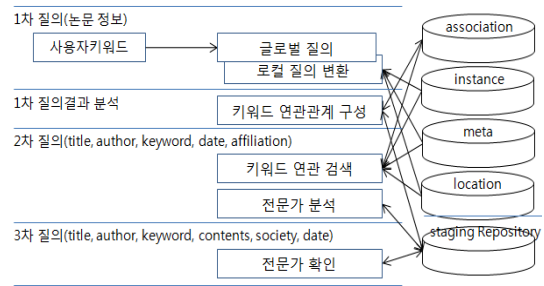


그림 2. 질의 처리를 위한 멀티 온톨로지 적용  
Fig. 2 Apply of ontology for query process

이 과정에서 사용되는 멀티 온톨로지가 적용되는 단계는 표 2와 같다. 메타 온톨로지는 표준 항목을 기준으로 로컬 스키마들과 연관관계를 온톨로지 구성한 것이다.

첫 번째 단계는 글로벌 항목으로 표현된 질의문으로 키워드 연관관계를 포함한 질의이다. 글로벌 항목은 메타 온톨로지의 표준 항목으로 질의를 구성한다. 이렇게 함으로써 모든 시스템에서 공통으로 이용할 수 있는 질의가 된다.

이때 질의는 일반적 질의문에서 가지는 FROM문을 가지지 않는데 이는 글로벌 질의로 로컬의 테이블 정보를 포함하지 않았기 때문이다. 두 번째 단계는 글로벌 항목으로 구성된 질의에서 토큰을 분리한 것으로 이때

생성되는 토큰은 데이터 항목에 해당하는 영역은 메타 온톨로지를 참조하고, 조건과 같은 값에 해당하는 영역은 인스턴스 온톨로지를 참조하기 때문에 토큰은 두 개 영역으로 분리된다. 세 번째 단계는 분리된 토큰 중 메타 온톨로지에 대한 로컬 시스템에 따른 분리를 위한 매핑과정이다.

표 2. 멀티 온톨로지를 이용한 질의 변환 사례  
Table. 2 Example of query conversin using multi-ontology

분석영역	분석내용
글로벌 항목	SELECT title, author, keyword, contents WHERE keyword = "온톨로지" AND keyword = "시맨틱";
분리된 토큰	meta inf : title, author, keyword, contents instance inf : 온톨로지, 시맨틱
메타 온톨로지 영역	title{ptitle, mainTitle}, author{name, authors, {lastname, firstname}}, keyword{k, keys}, contents{paper}
인스턴스 온톨로지 영역	온톨로지{ontology}, 시맨틱{semantic, 시맨틱}
로케이션 온톨로지 영역	ip, port, id, pass, path
연관 온톨로지 영역	온톨로지{+시맨틱, +OWL, +RDF, +에이전트, +메타데이터, +xml}

표 2에서 표현된 'title{ptitle, mainTitle}'가 의미하는 것은 글로벌 항목 'title'에 해당하는 로컬 시스템의 항목이 'ptitle'과 'mainTitle'이 있음을 예시로 나타낸 것이다. 네 번째 단계는 분리된 토큰의 인스턴스 온톨로지에 대한 매핑과정이다. 표현적 이질성이나 의미적 이질성을 이 단계에서 해결한다. 언어적 표현 차이도 포함한다. 다섯 번째 단계는 로컬 시스템에 접근하기 위한 위치정보와 접근 권한 정보를 포함하는 것이다. 이 단계에서 로컬 시스템에 접근하기 위한 질의가 완성된다. 마지막으로 여섯 번째 단계는 이 논문에서 제시하는 키워드 연관관계를 분석한 결과를 질의에 포함시키는 단계이다.

이상은 멀티 온톨로지가 전문가 망을 사용자에게 실시간으로 제공하는 과정이다. 이를 이용해서 사용자는 원하는 분야에 대한 학술 전문가를 더 효율적으로 찾을 수 있다.

#### IV. 시스템 적용 및 비교 성능 평가

제안되는 시스템의 전반적인 수행과정은 크게 2단계로 나누어지며, 그림 3과 같다. 1단계는 사용자의 키워드 입력에서부터 키워드 연관관계 분석을 통한 전문가 정보를 사용자에게 제공하는 단계로 그림 3에서 1번부터 15번까지이다.

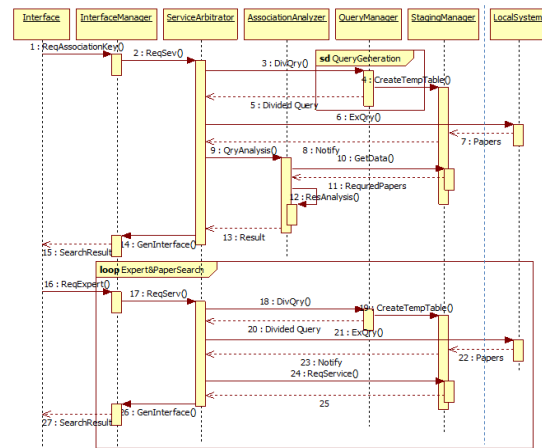


그림 3. 제안 시스템의 전반적인 시퀀스 다이어그램  
Fig. 3 The overall sequence diagram of proposed system

이를 통해 수행된 결과가 그림 4에서 제시되는 웹 인터페이스이다. 과정에서 1번-8번 과정이 1차 질의에 의해서 수집된 사용자 키워드를 통한 논문 정보를 수집하는 과정이다. 9번-13번 과정이 수집된 논문에서 키워드를 추출하여 연관관계를 분석하는 과정이다. 이를 통해 키워드와 연관키워드에 해당하는 전문가를 사용자에게 제공한다. 2단계는 사용자가 전문가를 확인할 수 있도록 저술 정보를 제공하는 단계이며, 그림 3에서 16번-27번까지이다. 이 단계는 사용자가 전문가를 선택할 때마다 해당 전문가의 저술 정보를 검색해야 하므로 반복 수행하게 된다.

그림 3과 같은 과정으로 진행되는 시퀀스 다이어그램에서 1단계에 해당하는 연관관계 분석 결과에 따른 전문가 검색결과이다. 사용자가 입력한 키워드 '온톨로지'와 연관관계를 가지는 키워드 들을 연관관계 가중치 우선순위에 따라 조합한 리스트를 화면상의 왼쪽에 제공된다.

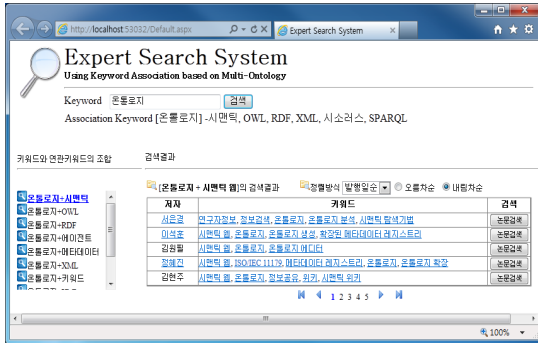


그림 4. 연관관계 분석을 통한 전문가 검색 결과 화면  
Fig. 4 The result of expert search Through association analysis

그림 4는 제공된 리스트에서 첫째 연관관계를 가지는 ‘시맨틱’과 같이 검색된 결과를 오른쪽 전문가 검색결과와 각 전문가의 키워드, 즉 관심분야를 제공하고 있다. 이 전문가와 관심분야를 통해 전문가를 확인하려면 해당 전문가의 논문검색 버튼을 클릭함으로써 확인할 수 있다.

성능 평가를 위해서 기존 키워드 검색과 제안된 방안을 비교 테스트 결과는 그림 5와 같다. 비교 테스트 방법은 기존의 키워드 검색과 제안된 방식 사이의 검색 건수와 검색에 소요된 시간을 비교하는 테스트를 수행하였다.

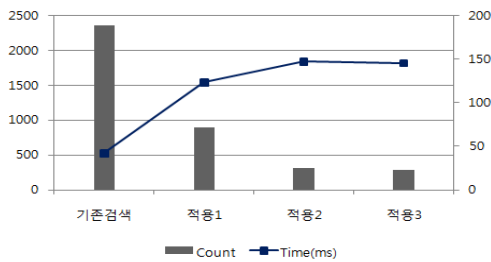


그림 5. 키워드검색과 제안된 방안의 성능 평가  
Fig. 5 The performance test between keyword search and suggested method

평가 결과 확인된 사항은 제안된 내용과 가깝게 검색에 대한 시간은 기존 검색보다 많이 산출되었다. 이는 검색을 수행하기 위한 사전 단계로 키워드 분석을 통한 연관성을 제시해야 하고, 분산된 여러 시스템에서 추출되어야 하기 때문에 초기 적용 단계인 적용1 단계와 시간

차이가 존재하지만, 적용된 이후에는 시간의 차이가 유사하거나 줄어든다. 그러나 검색의 연관성 정보가 추가됨에 따른 정확도 향상으로 검색의 결과 건수가 감소된다. 미미한 시간차이에 비해 정확도가 향상되므로 효율적이다.

## V. 결론

본 연구에서는 소셜 네트워크와 XMDR을 활용한 학술 분야별 중심 키워드 및 융합 분야 키워드 추천 기법을 제안하였다. 또한 제안하는 방법론의 실제 적용 가능성을 평가하기 위해 국내 우수 학술지에 게재된 논문의 131편의 키워드 600개에 대한 실험을 수행하였으며, 그 결과 제안한 방법에 따라 학문 분야별 중심 키워드 및 분야 융합을 위한 연결 키워드의 연관관계 분석을 확인할 수 있었다. 본 연구에서 제안하는 방안은 다음과 같은 장점을 제공한다. 첫째, 익숙하지 않은 분야의 논문을 검색하고자 하는 연구자에게 관심 분야의 키워드에 대한 초기 로드맵을 제공할 수 있다. 둘째, 특정 키워드에 대한 인접 키워드의 분포를 확인할 수 있다. 셋째, 이종 분야의 융합에 활용되는 연결 키워드를 발굴할 수 있다. 넷째, XMDR의 온톨로지를 활용하여 특정 키워드에 대한 연관관계 정의를 이용하여 동음이의어, 이음동이의어, 유사어 같은 연관관계 문제를 해결할 수 있다. 다섯째, 사용자가 요구하는 키워드를 기준으로 키워드간의 연관관계를 분석하는 방법을 제공함으로써 분석자의 주관 개입을 배제하고, 이를 객관화 방법을 제공하고 있다는 점에서 활용 가치가 높을 것으로 기대된다. 향후 본 논문은 데이터마이닝 기법의 도입으로 온톨로지를 이용한 연관관계 분석을 정련하고, 이를 통해 더욱 개관적인 분석이 되도록 할 예정이다.

## 참고문헌

- [1] 소셜 네트워크와 데이터 마이닝 기법을 활용한 학문 분야 중심 및 융합 키워드 추천 서비스, 조인동, 김남규, 지능정보연구 제17권 제1호 2011. 3. pp.127-138.

- [ 2 ] Wennerberg, P. O., "Ontology Based Knowledge Discovery in Social Networks. Final Report," JRC(Joint Research Center) 2005.
- [ 3 ] 이승훈, 김지혁, 김홍남, 조근식, "웹 기반 소셜 네트워크에서 시맨틱 관계 추론 및 시각화," 지능정보연구, 제15권, 제1호, 2009. 3.
- [ 4 ] 황치곤, 정계동, 최영근, "XMDR 기반의 통합 검색을 위한 데이터 그리드 Wrapper 설계", 한국해양정보통신학회논문지 Vol.12 No.5 p.921-929, 2008.
- [ 5 ] D. Brickley and L. Miller, "FOAF Vocabulary Specification 0.91", Nov. 2007. <http://xmlns.com/foaf/spec/>와 XFN[ XFN(XHTML Friend Network) <http://gmpg.org/xfn/>
- [ 6 ] 한국한의학연구원 소셜 네트워크 온톨로지 구축", 김상균, 장현철, 예상준, 한정민, 김진현, 김철, 송미영, 한국콘텐츠학회논문지 '09 Vol. 9 No. 12.
- [ 7 ] P. Mika, "Ontologies are us: A unified model of social networks and semantics," Journal of Web Semantics, Vol.5, No.1, pp.5-15, 2007.
- [ 8 ] Funakoshi, K. and T. Ohguri, "A Content-based collaborative recommender system with detailed use of evaluations", Proceedings of the 4th International conference on knowledge-based Intelligent Engineering systems and Allied Technologies, 2007.
- [ 9 ] Huang, Z., H. Chen. and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering", ACM Transactions on Information Systems, Vol.22, No.1 (2004), 116~142.
- [10] Balabanovic, M. and Y. shohm, "Fab : Content based, Collaborative Recommendation", communications of the ACM, Vol.40, No.3(1997), 66~72.
- [11] G. Salton and M. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, 1983.

저자소개

**정계동(Kye-dong Jung)**



1985년 광운대학교 전자계산학 (이학사)  
1992년 광운대학교 산업정보학 (이학석사)

2000년 광운대학교 컴퓨터과학 (이학박사)  
1993년 ~ 2004년 광운대학교 정보과학원 교수  
2005년 ~ 현재 광운대학교 교양학부 교수  
※관심분야: XML 분산시스템, 분산 컴퓨팅기술, 이동에이전트

**황치곤(Chi-gon Hwang)**



1995년 창원대학교 경영학과 (학사)  
2004년 광운대학교 정보통신학과 (공학석사)

2008년~현재 광운대학교 컴퓨터 과학 박사과정  
2006년~현재 (주)인찬 연구원  
※관심분야: 웹서비스, XMDR, 이동에이전트, 상호 운용

**최영근(Young-keun Choi)**



1980년 서울대학교 수학교육과 (이학사)  
1982년 서울대학교 계산통계학과 (이학석사)

1989년 서울대학교 계산통계학과 (이학박사)  
1982년 ~ 현재 광운대학교 컴퓨터과학과 교수  
1992년 ~ 2000 광운대학교 전산정보원 원장  
2002년 ~ 2005 광운대학교 교무연구처장  
※관심분야: 객체지향 설계, 분산시스템, 이동에이전트, 상호운용