

# 지능형 전문가관리 프레임워크를 위한 주제 분야 계층 자동 생성

양근우<sup>a</sup>, 이상로<sup>b</sup>

<sup>a</sup> 계명대학교 경제통상학부 전자무역학전공  
704-701, 대구광역시 달서구 달구벌대로 2800  
Tel: +82-53-580-6959, Fax: +82-53-580-5313, E-mail: kwyang@kmu.ac.kr

<sup>b</sup> Information Systems, Rawls College of Business Administration, Texas Tech University  
Lubbock, TX 79409, USA  
Tel: +1-806-742-3919, Fax: +1-806-742-3193, E-mail: sangno.lee@ttu.edu

## Abstract

In this paper, we introduce the methodology for the automatic generation of the subject field hierarchy for Intelligent Expert Management Framework using WordNet. Intelligent Expert Management Framework, which is proposed as an appropriate method to manage valuable tacit knowledge within the organization, defines the expert profile structure and proposes the efficient method to automate the process to collect and update the expert profile information based on the profile structure defined. To increase the satisfaction level of users, additional intelligent search features are defined and users can be given the list of experts in related or similar expert fields when they perform expert searches based on the expert database being built. To enable automatic profiling of the organizational experts as well as intelligent expert searches, the subject field hierarchy, upon which the expert profiles are classified and expert searches for similar fields are performed, should be predefined. In this paper, we propose the WordNet library method that first eliminates the ambiguity of the senses of nominal data values, constructs the subject field hierarchy by overlapping the hypernym of the remaining senses, and lastly adjusts the derived hierarchy to the preference of users. Based on the proposed methodology, we expect to avoid the prohibitive costs in building large subject field hierarchies when manually done as well as maintain the objectivity of the hierarchies.

## Keywords:

Knowledge Management System; Expert Management; WordNet; Taxonomy

## 서론

조직 내 중요한 지식 자산의 효과적인 관리, 공유, 확산을 위한 지식관리 시스템 (KMS: Knowledge Management System)은 오늘날의 치열한 경쟁 환경 하에서 기업의 경쟁력을 유지하기 위해 꾸준히 연구되어 오고 있다. 지금까지 지식관리 시스템과 관련한 연구의 대부분은 조직 내 전문가가 보유하고

있는 형식지, 암묵지 등으로 분류되는 지식을 코드화하여 지식저장소 (Knowledge Repository)에 효과적으로 저장하고 이를 사용자가 쉽게 검색하여 활용할 수 있도록 하는데 그 초점이 맞추어져 왔다. 그러나 이러한 시도가 지식의 동적이고 상황 종속적인 특징으로 인해 조직이 직면한 다양한 형태의 문제에 대한 효과적인 해답을 제공하기에는 많은 문제가 있다는 점이 밝혀지고 있고 따라서 많은 지식관리 시스템 프로젝트가 실패하고 있는 실정이다. 이에 지식관리 연구의 다른 한 방향은 조직 내의 전문 지식을 보유한 전문가의 프로파일을 효과적으로 관리하고 필요 시 전문가의 검색을 도와 조직의 문제 해결에 도움을 주는 방법이 더 효과적이라고 제시하고 있다 [1, 4, 5, 6].

지능형 전문가 관리 프레임워크 (IEMF: Intelligent Expert Management Framework) [9]는 지식관리 시스템에서의 두 가지 전문가 관리 기능인 전문가 식별과정 (Expert Identification)과 효과적인 전문가 검색 (Expert Search)을 모두 지원하는 효과적인 방법론으로 제안되었다. 수동적인 전문가 식별 과정의 문제점을 해결하기 위해 개별 전문가에 의해 등록된 지식물을 분석하고 분류하여 자동으로 전문가 프로파일 정보를 생성, 지속적으로 갱신함으로써 이러한 과정을 자동화하였다. 또한, 전문가 검색을 위해서는 추가적인 검색 연산자를 정의하고 제안하여 사용자의 검색 만족도를 높일 수 있는 방안을 제공하고 있다.

본 논문에서는 이러한 지능형 전문가 관리 프레임워크 구현에 필수적인 주제 분야 계층을 자동으로 생성하는 방법론을 고안하여 제안하고자 한다. 이를 위해 WordNet [3, 7]이라는 온라인 어휘 검색시스템을 도입하여 활용하였으며 WordNet 내부에 정의된 어의간의 상위어, 하위어 관계를 통해 자동으로 초기 주제 분야 계층을 생성하는 과정을 제시하고자 한다. 이를 통해 효과적인 전문가 식별과 검색의 기반이 되는 주제 분야 계층의 구축과 관리 비용을 절감하고 주제 분야 분류의 객관성을 확보할 수 있을 것으로 기대한다.

## 지능형 전문가관리 프레임워크

### 전문가 프로파일링 자동화

전문가 프로파일이란 한 전문가가 보유하고 있는 전문 지식의 분야와 각 분야별 전문 지식의 정도를 나타내는 정보의 모음을 의미한다. 특정 전문가는 하나 이상의 주제 분야에 대한 전문 지식을 보유할 수 있으며 각 전문가의 전문가 프로파일은 해당 전문가의 관심이 변하거나 또는 해당 전문 분야 자체가 시간의 흐름에 따라 변화하는 등의 이유로 인해 수시로 변경될 가능성이 있다. 따라서, 수집된 전문가 프로파일 정보를 항상 최신의 것으로 유지하고 그 유용성을 보존하기 위해서는 지속적이고 시스템화된 전문가 프로파일 생성 방안이 필요하다. 이러한 관점에서 전문가 프로파일 관리 기법의 자동화는 수동 프로파일 관리 방법의 단점을 극복할 수 있는 방안으로 볼 수 있다.

전문가 프로파일 데이터베이스를 구축하기 위해서는 전문가가 보유하고 있는 전문 지식의 수준 역시 측정되어야 한다. 전문 지식은 (1) 활동성 (activeness), (2) 관련성 (relevance), (3) 유용성 (usefulness)의 세 가지 요소를 기준으로 평가될 수 있다. 지능형 전문가 관리 프레임워크에서 활동성은 전문가가 파일을 업로드하거나 게시판에 글을 등록하고 이메일이나 쪽지를 주고 받는 등의 행위를 통해 시스템에 지식물을 등록하는 지식 생성 활동을 얼마나 빈번히 수행하는지를 의미하며 관련성은 등록된 지식물이 얼마나 특정 주제 분야에 적합한가를 의미한다. 한편, 실제로 매우 가치 있는 지식이 시스템에 등록되었다 하더라도 시스템을

활용하는 다른 사용자가 해당 지식물을 이용하고 그로부터 해당 지식의 가치를 인정하기 전에는 그 지식이 유용하거나 의미 있는 것이라고 보기 힘들다. 따라서 등록된 지식에 대한 다른 사용자의 평가(유용성)를 전문 지식을 측정하는 과정에 포함시키는 것이 필요하다.

그림 1은 다양한 관련 연구에 채택되어 좋은 문서 범주화 성능을 보인 벡터 공간 모델 (Vector Space Model) [2]을 이용하여 전문가 자동 분류기를 학습시키고 학습된 분류기를 이용하여 전문가 프로파일 정보를 생성하는 과정을 보여준다. 전문가 프로파일 추출 과정은 특정 전문가에 의해 등록된 지식물과 등록된 주제 분야간의 유사도를 계산하는 단계를 포함하는데 이러한 계산을 위해 이전 학습 단계에서 도출된 분야별 문서 벡터를 활용하게 된다.

### 지능형 전문가 검색

KMS에 등록된 전문가를 통해 그들이 보유하고 있는 암묵지를 효과적으로 활용하기 위해서는 각 전문가가 전문 지식을 보유하고 있는 주제 분야를 정확히 구분하는 것뿐만 아니라 사용자의 검색 조건에 맞는 특정 전문 지식을 보유하고 있는 전문가를 검색하기 위한 효과적인 검색 방법을 제공하는 것도 매우 중요하다. 전문가를 검색함에 있어서 시스템은 등록된 전문가의 부족, 특정 분야 전문가의 부재 등 다양한 원인에 의해 사용자의 검색 요구에 대해 만족스러운 결과를 제공할 수 없는 경우가 흔히 발생한다. 이처럼 사용자의 요구에 부합하는 전문가의 검색이 불가능할 경우

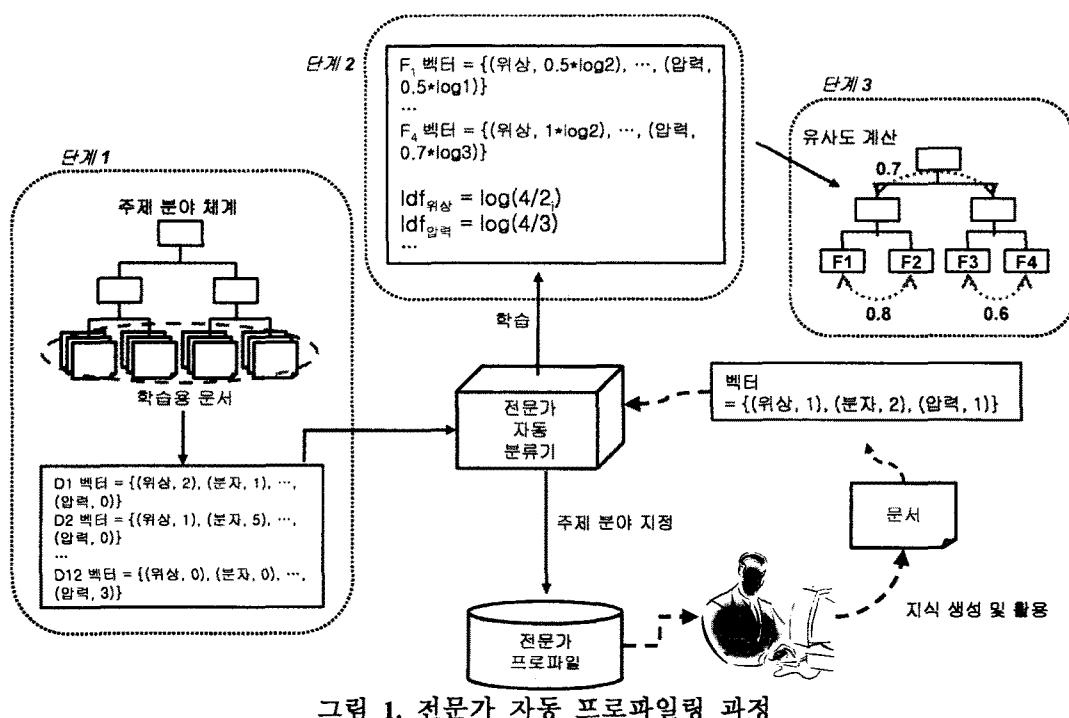


그림 1. 전문가 자동 프로파일링 과정

시스템이 이와 유사한 혹은 관련 있는 분야의 전문가를 대신 검색하여 제공한다면 사용자의 검색 만족도를 크게 향상시킬 수 있을 것이다. 이는 특정 분야의 전문가가 존재하지 않을 경우 이와 유사한 분야의 전문가가 활용 가능한 인력 중 사용자의 문제 해결을 위해 유용한 정보를 제공할 수 있는 가능성이 가장 높기 때문이다. 또한 다수의 특정 분야 전문 인력이 필요한 상황이라면 해당 분야 전문가의 부족 시 유사 분야의 전문가가 가장 최선의 대안이 될 수 있다.

지능형 전문가 관리 프레임워크에서는 지식 표현 방법론인 FAH (Fuzzy Abstraction Hierarchy) [8]를 도입하여 유사 분야 전문가 검색 결과의 일부로 검색 대상 분야와의 적합도 혹은 검색 결과내의 순위 등을 제공한다. 이를 통해 검색된 전문가가 얼마나 검색 조건을 만족하는지를 사용자에게 알려줄 수 있다면 이 또한 사용자에게 매우 유용한 정보가 될 수 있다. FAH와 문서 범주화 기법을 접목하면 전문가 검색 기능의 향상을 위한 두 가지 추가 정보 제공이 가능해 진다. 그 하나는 FAH의 퍼지 논리 계산을 이용한 주제 분야 체계에 정의된 주제 분야간의 유사도이며 이는 유사한 분야의 전문가를 검색하기 위해 활용된다. 또 다른 하나는 검색된 전문가가 특정 주제 분야에 대한 전문 지식을 어느 정도 보유하고 있는지를 판단하는데 활용할 수 있는 해당 분야와의 적합도 점수이다. 이와 같은 주제 분야간의 유사도와 각 전문가의 분야별 적합도 점수는 KMS 사용자가 필요로 하는 전문가를 보다 효과적으로 검색하는데 크게 도움이 된다.

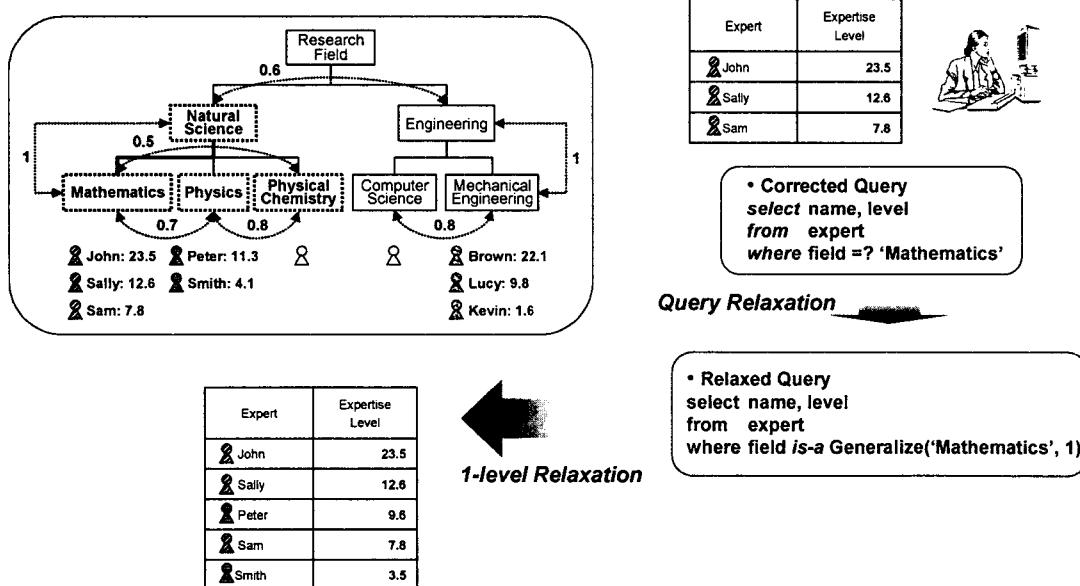
전문가 검색은 사용자의 검색어 입력으로부터 시작되는데 시스템은 정의된 주제 분야 분류 체계에서 사용자가 입력한 검색어와 일치하는 주제

분야명이 존재하는지를 검색하게 된다. 이러한 주제 분야명 검색은 주제 분야 체계에서의 가장 하위 수준으로부터 시작하며 만약 일치하는 분야명을 찾는데 실패하면 다음 상위 수준 분야들을 대상으로 검색이 수행된다. 최하위 수준에서 최상위 수준까지의 검색 과정에서 만약 일치하는 주제 분야명이 존재할 경우 전문가 프로파일 데이터베이스로부터 해당 분야의 전문가 목록이 구성되어 사용자에게 제공된다. 이 과정에서 사용자에게 추가적인 정보를 제공하기 위해 유사 분야 전문가 검색을 수행할 수 있다.

전문가 검색의 결과로 구성된 전문가 목록은 그 결과가 만족스럽지 못할 경우 사용자의 필요에 따라 질의 완화 혹은 질의 강화를 포함하는 FAH 연산을 수행하는데 활용된다. 만약 사용자의 검색 요구를 만족시키는 전문가의 수가 지나치게 많거나 혹은 적을 경우 각각 그 하위 혹은 상위 수준의 주제 분야에 대한 질의 처리를 통해 질의 결과를 줄이거나 늘릴 수 있다. 조금 더 구체적으로 살펴보면 만약 너무 많은 전문가가 검색될 경우 질의 강화를 적용하여 시스템에서 제공하는 더 구체적인 하위 주제 분야 목록으로부터 특정 주제 분야를 선택하여 검색 영역을 줄임으로써 검색 조건을 강화할 수 있다. 반대로 너무 적은 검색 결과가 제공될 경우 질의 완화를 통해 유사 분야 전문가를 포함하는 더 많은 전문가의 목록을 제공하는 것이 가능하다. 그림 2는 지능형 검색 기능 중 질의 완화를 통한 유사 분야 전문가 검색의 예를 보여주고 있다.

## 주제 분야 계층 자동 생성

### WordNet과 적용 가능성



WordNet은 1990년대 프린스턴 대학의 연구자들에 의해 개발된 온라인 영어 어휘 검색 시스템이다. WordNet은 단순히 알파벳순으로 어휘가 나열된 온라인 사전으로 개발된 것이 아니라 어휘 또는 개념 간의 유사도, 포함관계 등 보다 개념적인 방식으로 어휘 간 의미론적 관계를 표현하기 위해 개발되었다. WordNet은 지속적으로 포함하는 어휘의 수를 늘려오고 있으며 특히, 동의어 (synonym), 반의어 (antonym), 하위어 (hyponym) 등을 포함하는 17가지 종류의 어휘간 관계를 포함한다. 지금까지 120,000개의 명사, 동사, 형용사, 부사 등의 형태로 표현된 약 90,000개 영단어의 의미를 포함하고 있는데 이는 물리적인 어조나 설명에 기반한 단어의 형태를 기준으로 하는 일반적인 사전과는 달리 해당 어휘의 사전적 의미를 포함하는 어의 또는 의미를 기반으로 구조화되어 있다는 점에서 차별화된다. 이러한 관점에서 WordNet은 전통적인 사전과 의미론적 네트워크 개념을 포함하는 동의어 사전의 기능을 결합한 것으로 볼 수 있다. 그렇지만 WordNet은 또한 동의어 사전과는 달리 어휘간의 다음과 같은 네가지 독특한 추가적인 의미론적 네트워크 구조를 제공한다는 점에서 차별화된다.

우선 하위어 (hyponym), 상위어 (hypernym) 관계는 subordination/superordination, subset/superset 또는 is-a 관계 등으로도 불리는데 예를 들어 “탁자 (table)”는 “가구 (furniture)”의 하위어이며 반대로 “가구”는 “탁자”的 상위어가 된다. 한편, 분의어 (meronym), 전체어 (holonym) 관계는 부분 (part-of/has-part), 소속 (member-of/has-member), 성분 (substance-of/has-substance) 등의 관계를 포함하는데 예를 들어 “탁상 (tabletop)”과 “다리 (leg)” 등은 “탁자 (table)”의 분의어이고 그 역으로 “탁자”는 두 어휘의 전체어가 된다. 이중에서 본 논문에서는 특히 각 개념의 하위어, 상위어 관계를 이용하여 주제 분야 계층을 자동으로 생성하는데 활용하고자 한다.

WordNet에서의 하위어, 상위어 관계는 경로 형태의 계층으로 표현되는데 따라서 이와 유사한 개념 간 상하 관계를 가지는 주제 분야 계층을 위한 토대를

제공한다고 볼 수 있다. 즉, 필요한 모든 개념들의 상위어를 도출하여 이를 결합하는 방식으로 모든 개념 집합을 포함하는 주제 분야 체계를 구축하는 것이 가능하게 된다. 다음 절에서는 이러한 주제 분야 계층 자동 생성의 과정을 자세히 설명하고자 한다.

### 계층 자동 생성 과정

주제 분야간의 관계를 표현하는 계층의 자동 생성 과정은 그림 3에서 제시하는 바와 같이 (1) 데이터베이스에서 선택된 테이블로부터 전체 개념 집합을 도출, (2) 유사도를 이용한 관련 없는 어의의 제거, (3) WordNet을 이용하여 개념간 관련 어의의 상위어를 결합하여 주제 분야 계층 생성, (4) 사용자의 선호를 반영한 결과 계층의 조정 등 총 네 단계를 포함한다.

첫째, 전체 개념 집합을 도출하기 위해서는 데이터베이스의 특정 테이블로부터 모든 필드 값을 추출해야 한다. 집합론에 의하면 이 과정에서 하위 집합을 구성하는 모든 최종 노드들이 겹친다. 구축되는 주제 분야 계층의 관점에서 보면 이 과정을 통해 계층의 최하위 레벨을 구성하는 분야들이 도출된다고 볼 수 있다. 이러한 과정은 사용자에 의해 해당 데이터베이스의 테이블을 선택하는 것으로 시작되며 이렇게 선택된 테이블의 필드 값이 모두 추출된다.

다음으로 두 번째 단계에서는 주제 분야를 구성하는 개념의 어의 중 관련 없는 것들을 제거함으로써 도출된 주제 분야의 모호함을 제거하게 된다. 만약 도출된 주제 분야가 문맥상 하나의 어의만을 가진다면 WordNet에 존재하는 해당 개념들의 상위어를 일치시키는 방법으로 간단하게 주제 분야 계층을 만들어 내는 것이 가능하다. 그렇지만 대부분의 경우 주제 분야를 의미하는 하나의 개념이 다수의 어의를 갖는 경우가 많다. 예를 들어, JAVA, PASCAL, BASIC, DBMS 등과 같은 IT관련 주제어의 경우 각기 여러 의미를 가질 수 있다. “JAVA”的 경우 WordNet에서 컴퓨터 프로그래밍 언어와

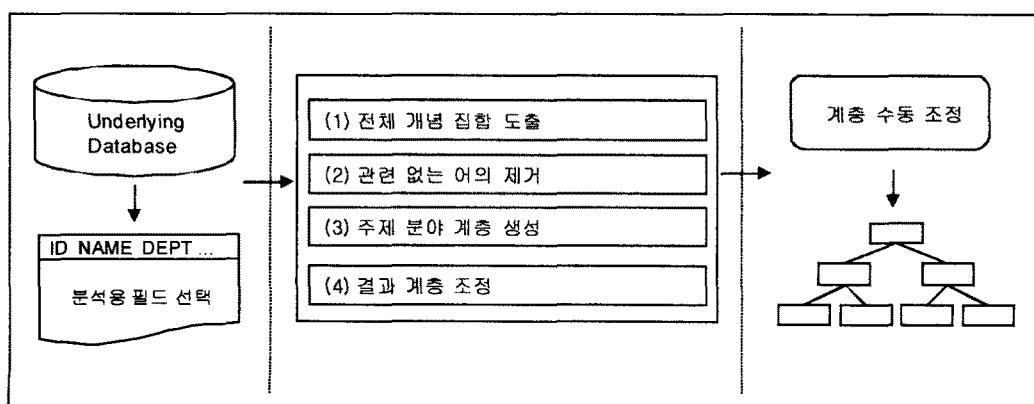


그림 3. 계층 자동 생성 과정

pascal, Pa – (a unit of pressure equal to one newton per square meter)  
 Pascal, Blaise Pascal – (French mathematician and philosopher and Jansenist;  
     invented an adding machine; contributed (with Fermat) to the theory of probability (1623-1662))  
 Pascal – (a programming language designed to teach programming  
     through a top-down modular approach)  
 BASIC – (a popular programming language that is relatively easy to learn;  
     an acronym for beginner's all-purpose symbolic instruction code; no longer in general use)  
 basic, staple – ((usually plural) a necessary commodity for which demand is constant)

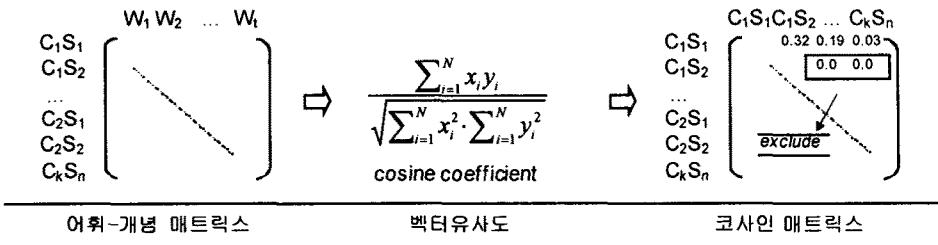


그림 4. 유사도를 이용한 어의 명료화 과정

관련한 어의 외에도 섬, 음료(커피) 등과 연관된 세 가지 어의로 설명되고 있다. 따라서, 프로그래밍 언어 분야와 관련된 주제 분야 계층을 생성한다면 컴퓨터 혹은 프로그래밍 언어와 관련된 각 개념 어의의 상위어를 선택하는 것이 필요하다. 그림 4는 “pascal”과 “basic”이라는 개념의 어의 명료화 과정을 도식화하고 있다.

일단 관련 없는 어의를 제거하는 방법으로 주제 분야 어의의 모호함이 사라지면 다시 말해 동일한 분야에 해당하는 관련 있는 어의 집합이 도출되면 세 번째 단계에서는 이렇게 도출된 개별 주제 분야 어의의 상위어로부터 주제 분야 계층이 생성된다. 예를 들어, “JAVA”라는 주제 분야의 경우 컴퓨터 관련 어의로서 “programming language”라는 구체적인 개념으로부터 시작하여 “artificial language,” “language,” “communication,” “social relation,” “relation,” 그리고 최상위 개념으로 “abstraction”까지 일련의 상위어를 가지게 된다. 마찬가지로 또 다른 컴퓨터 관련 어의인 “DBMS”는 “software,” “code,” “coding system,” “writing,” “written communication,” “communication,” “social relation,” “relation,” “abstraction”의 상위어 관계로 이루어져 있다. 컴퓨터 관련 어의만을 고려한다면 “JAVA,” “PASCAL,” “BASIC”的 경우 “programming language”라는 분야에서 동일한 상위어 개념을 공유하게 되고 여기에 “DBMS”라는 개념까지 포함한다면 “communication” 단계에서 공유 개념을 가지게 된다. 따라서, 유효한 어의 집합에 속하는 모든 상위어가 도출된다면 이러한 상위어의 일치 과정을 통해 해당 필드의 주제 분야 계층이 자동으로 구성될 수 있다. 그림 5는 네 개의 주제 분야와 열 개의 어의를 가진 분야의 개념적 계층 관계를 표현하고 있다.

마지막으로 전 단계에서 상위어 일치를 통해 얻어진 계층은 여전히 완전한 주제 분야 계층으로 활용하기에는 다소 부족한 부분이 존재하는데 예를 들어 특정 개념의 경우 하나의 공통 상위 개념만을 공유하면서도 불필요하게 여러 단계를 거치면서

매우 긴 중간 노드를 포함할 가능성이 있다. 또는 경우에 따라서 상위어 일치 과정에서 어떠한 상위 개념도 공유하지 않아 모든 개념이 포함된 하나의 주제 계층 대신 여러 개로 나누어진 작은 계층의 집합이 도출될 가능성도 있다. 이것은 부분적으로 어의간에 전혀 공유되는 요소가 없을 경우에 자연스러운 결과로 나타나거나 상위어를 만들고 유지하는 과정에서 일관성이 유지되지 못하는 한계에서 발생하는 오류에 기인한다고 볼 수 있다. 어느 경우이건 최종 주제 분야 계층의 완성을 위해 불필요하게 추가된 노드를 제거하거나 수동으로 분리된 노드를 상위 개념으로 연결하는 수정과정이 필요하게 된다.

## 결론 및 향후 연구 방향

본 연구에서는 지능형 전문가 관리 프레임워크를 위한 주제 분야 자동 생성 방법론을 제안하였다. 조직 내 지식 자산의 효과적인 관리와 공유, 활용을 위해 제안된 지능형 전문가 관리 프레임워크는 전문가 자동 식별과 검색을 위해 주제 분야 간 유사도를 포함하는 주제 분야 계층을 요구한다. 이러한 주제 분야 계층은 미리 공신력 있는 기관에 의해 분류된 것을 활용하거나 특별한 원칙에 의해 구성하는 것으로 가정하였다.

본 논문에서는 이처럼 미리 정해진 주제 분야 체계를 기반으로 전문가 자동 식별과 유사 분야 전문가 검색을 수행하는 방식을 개선하고자 주제 분야로 구분될 수 있는 개념들과 개념의 어의간 관계를 이용하여 자동으로 이러한 주제 분야 체계를 도출하는 방법론을 제안하였다. 그리하여 전체적인 지능형 전문가 관리 프레임워크에서 주관적인 개입을 최소화하는 방안을 마련하였다. 다만 본 연구에서 제안한 주제 분야 계층 자동 구축 방법론 역시 주제 분야 계층 생성을 위한 전문가 개입을 완전히 배제할 수 있는 완벽한 방법을 제안하지는

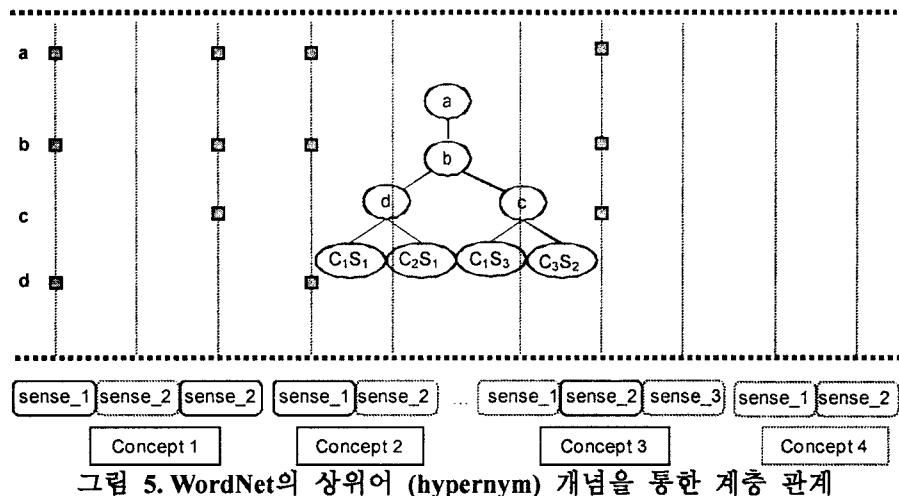


그림 5. WordNet의 상위어 (hypernym) 개념을 통한 계층 관계

못한 한계가 존재하며 다만 최대한 효율적인 주제 분야 계층 생성과 전문가 개입을 최소화할 수 있다는 점에 그 의의가 있다 하겠다.

향후 제안된 방법론을 실제 데이터를 활용하여 주제 분야 자동 생성 실험을 실시하고 그 결과를 전문가에 의해 생성된 주제 분야 체계와 비교, 검토하여 그 유효성을 검증하는 과정이 필요하며 제안된 방법론을 응용하여 지능형 전문가 관리 프레임워크를 구현한 시스템의 완성도를 높이기 위한 노력이 요구된다.

Knowledge Management Systems," *Journal Database Management*, Vol. 18, No. 3, pp. 47–68.

## References

- [1] Augier, M., and Vendelo, M.T. (1999). "Networks, Cognition and Management of Tacit Knowledge," *Journal of Knowledge Management*, Vol. 3, No. 4, pp. 252–261.
- [2] Baeza-Yates, R., and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York, NY: ACM Press.
- [3] Christiane, F. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- [4] Desouza, K.C. (2003). "Barriers to Effective Use of Knowledge Management Systems in Software Engineering," *Communications of the ACM*, Vol. 46, No. 1, pp. 99–101.
- [5] Hansen, M.T., Nohria, N., and Tierney, T. (1999). "What's Your Strategy for Managing Knowledge?" *Harvard Business Review*, Vol. 77, No. 2, pp. 106–116.
- [6] Liebowitz, J. (2001). "Knowledge Management and Its Link to Artificial Intelligence," *Expert Systems with Applications*, Vol. 20, No. 1, pp. 1–6.
- [7] Miller, G.A., Beckwith, R., Felbaum, C., Gross, D., and Miller, K. (1990). "Introduction to WordNet: an On-line Lexical Database," *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235–244.
- [8] Moon, K.H., and Huh, S.Y. (2005). "An Integrated Query Relaxation Approach Adopting Data Abstraction and Fuzzy Relation," *Submitted for publication to Information Systems Research*.
- [9] Yang, K.W., and Huh, S.Y. (2007). "Intelligent Search for Experts Using Fuzzy Abstraction Hierarchy in