

문서 범주화를 이용한 지식관리시스템에서의 전문가 분류 자동화*

양 근 우**, 허 순 영***

Automation of Expert Classification in Knowledge Management Systems Using Text Categorization Technique

Kun-Woo Yang, Soon-Young Huh

This paper proposes how to build an expert profile database in KMS, which provides the information of expertise that each expert possesses in the organization. To manage tacit knowledge in a knowledge management system, recent researches in this field have shown that it is more applicable in many ways to provide expert search mechanisms in KMS to pinpoint experts in the organizations with searched expertise so that users can contact them for help. In this paper, we develop a framework to automate expert classification using a text categorization technique called Vector Space Model, through which an expert database composed of all the compiled profile information is built. This approach minimizes the maintenance cost of manual expert profiling while eliminating the possibility of incorrectness and obsolescence resulted from subjective manual processing. Also, we define the structure of expertise so that we can implement the expert classification framework to build an expert database in KMS. The developed prototype system, "Knowledge Portal for Researchers in Science and Technology," is introduced to show the applicability of the proposed framework.

Keywords : expert classification knowledge management system vector space model; portal

* 본 연구는 대학 IT연구센터 육성·지원사업의 연구결과로 수행되었음.

** 한국과학기술원 테크노경영대학원 박사과정

*** 한국과학기술원 테크노경영대학원 부교수

I. 서론

정보기술의 발달에 따라 개개인이 일상에서 접하고 처리해야 할 정보의 양은 크게 증가하고 있다. 특히, 획득 가능한 정보를 바탕으로 중요한 경영 의사 결정을 해야 하는 기업 환경 하에서 정보를 획득, 저장, 공유 및 검색하는 일련의 과정은 대부분의 조직이 직면하고 있는 핵심적인 과제 중 하나로 인식되고 있다. 지식관리시스템(Knowledge Management System)은 조직 내에 존재하는 이러한 가치 있는 지식을 효과적으로 관리하고자 하는 요구에 의해 탄생하여 발전하고 있다[Alavi and Leidner, 2001].

지식관리시스템에서 형식지(explicit knowledge)의 경우 데이터베이스와 같은 정보 기술을 이용하여 관리하기 용이한 반면 특정 상황에 대한 전문적인 지식, 기술, 기능 등을 포함하는 암묵지(tacit knowledge)의 경우 그 특징에 따라 비구조적인 경우가 많아 효과적으로 관리하기가 용이하지 않다[Alavi and Leidner, 2001; Nonaka and Takeuchi, 1995; Nonaka, 1994]. 암묵지를 관리하기 위한 방법은 크게 두 가지로 나누어 질 수 있는데 첫 번째 방법은 암묵지를 모델링하고 코드화할 수 있는 구조적인 방법을 개발하여 형식지를 다루는 방법과 유사하게 코드화된 암묵지를 저장, 전송, 활용할 수 있도록 하는 것이다. 예를 들어, Design Rationale 시스템[Hu et al., 2000]은 특정 사물을 디자인하는데 있어서 디자이너의 의도, 동기, 의사결정, 방법간의 장단점 등 그 배경 지식을 각 시스템이 제공하는 방식으로 표현하고 저장할 수 있는 기능을 제공하는데 이러한 시스템은 암묵지 관리를 위한 첫 번째 방식의 예로 볼 수 있다. 두 번째 방식은 암묵지를 그 소유자로부터 강제로 분리해 내는 방법이 아닌 필요한 지식을 소유하고 있는 사람에 대한 검색 방법을 제공하는 것이다.

전문가의 머리 속에 존재하는 암묵지를 완벽

하게 모델링하는 방법을 개발하는 것은 사실상 불가능하며 지식관리시스템의 개발과 관련한 문제들을 고려할 때 두 번째 방법이 보다 적용 가능하고 복잡도를 줄일 수 있는 방법으로 볼 수 있다. 따라서, 지식관리시스템 연구자들은 지식관리시스템에서 효과적으로 전문가를 검색할 수 있는 방법을 개발하기 위해 노력해 왔으며[Alavi and Leidner, 2001 Rus and Lindvall, 2002], 이러한 연구에서 제안하는 전문가 검색 방법들은 실제로 몇몇 상업용 지식관리시스템 엔진에 도입되어 활용되고 있다[Handysoft Microsoft Verity].

전문가 검색이 가능하도록 하려면 우선 조직 내의 어느 전문가가 어떠한 전문 지식을 보유하고 있는지를 알 수 있도록 전문가를 분류하는 과정이 선행되어야 한다. 이러한 전문가 분류 과정은 각 전문가가 보유하고 있는 전문 지식에 대한 전문가 프로파일(Expert Profile) 정보의 수집과정으로도 볼 수 있는데 이렇게 수집된 전문가 프로파일 정보로 구성된 전문가 데이터베이스는 조직의 문제 해결을 위해 필요한 전문가를 검색하는 대상이 된다. 이러한 전문가 분류 즉, 프로파일 정보 수집은 첫째, 수동으로 시스템 관리자나 각 사용자가 수행하거나 둘째, 미리 정의된 알고리즘을 이용하여 시스템에 의해 자동적으로 수행될 수도 있다. 현재 전문가 검색기능을 제공하는 대부분의 상업용 지식관리시스템 엔진에서는 수동 프로파일 정보 입력 방법을 도입하여 활용하고 있는데 이 방식의 경우 구현의 용이함을 제공하는 대신 프로파일 정보의 갱신을 위한 지속적인 관리 비용을 유발하는 단점을 가지고 있다. 수동 전문가 분류 방법의 경우 프로파일 정보 갱신에 필요한 관리 비용 이외에도 다음과 같은 단점을 가지고 있다. 첫째, 시스템 관리자나 전문가 자신이 프로파일 정보를 입력할 경우 해당 입력 정보의 객관성을 유지하기 어렵다. 각 전문가가 자신의 전문 분야나 전문 지식의 수준을 스스로 입력

하게 되면 그 기준이 모호하고 서로 다른 전문 분야에 대해서도 다른 기준이 적용될 가능성이 있다. 둘째, 사용되는 전문 분야 분류 체계가 수 많은 전문 분야로 구성되어 있고 한 사람의 시스템 관리자가 이렇게 다양한 분야에 대한 전문가를 할당하는 경우 모든 분야에 대한 분류를 수행할 수 있는 지식을 갖고 있기 힘들다는 한계가 있다. 셋째, 전문 분야별로 계속해서 새로운 개념이나 연구 경향이 나타나는 등의 변화를 가질 수 있는데 이러한 분야 자체의 변동을 반영하기가 쉽지 않다.

본 연구에서는 이러한 수동 전문가 분류 방법의 단점을 개선하고자 지식관리시스템에 등록된 전문가의 전문 분야 정보를 자동으로 수집하고 특정 주제 분야에 대한 전문 지식의 수준을 측정할 수 있는 전문가 분류 및 관리 자동화 방법론을 제안하고자 한다. 전문가 프로파일 정보 수집과 갱신 자동화를 통해 전문가를 분류하고 전문가 데이터베이스를 구축하는 것은 전문 분야 할당을 위한 프로파일 정보의 구조를 정의하고 시스템적으로 전문가의 전문 지식을 측정하는 방법론 개발의 어려움으로 인해 지금까지의 연구에서 성공적으로 해결되지 못했다. 본 연구에서 제안하는 전문가 분류 및 관리 방법론은 전문가의 전문 분야 및 전문 지식의 정도를 측정 또는 계산하는데 활용할 수 있는 전문가 프로파일 구조의 정의를 포함하고 있는데 이를 통해 제안된 전문가 분류 자동화 방법론을 지식관리시스템에 실제로 적용할 수 있게 된다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 지식관리시스템에서의 전문가 관리 방안과 관련한 기존 연구를 정리하고 3장에서는 본 연구에서 정의하여 활용하는 전문가 프로파일의 구조를 소개하고 자동 문서 범주화 기법을 응용한 전문가 자동 분류기의 작동 원리를 설명한다. 4장에서는 3장에서 소개한 전문가 자동 분류기를 이용한 전문가 분류 자동화 방법론을

설명하고 제안된 방법론을 적용하여 개발된 '과학기술 연구를 위한 지식포털' 시스템을 소개한다. 마지막으로 5장에서 향후 연구 방향과 함께 결론을 맺는다.

II. 관련 연구

지난 수년간 지식의 암묵적인(tacit) 면에 대한 관심이 학계는 물론 산업계에서도 크게 증가하였다[Alavi and Leidner, 2001; Augier and Vendelo, 1999; Kreiner, 2002; Zack, 1999]. 암묵지와 관련한 과거의 연구들은 이러한 종류의 지식을 획득, 저장, 검색하여 활용하는 것과 관련한 문제를 효과적으로 해결하지 못한 이유로 인해 널리 활용되지 못하였다. 이러한 노력의 예로 Design Rationale 시스템[Buckingham-Shum and Hammond, 1994; Conklin and Yakemovic, 1991; Hu et al., 2000]을 들 수 있는데 이 분야의 몇몇 진취적인 연구 결과가 특정 분야에 활용된 경우는 있었으나 그 제한된 적용 가능성으로 인해 크게 확산되지는 못하였다.

Kreiner[2002]는 그의 연구에서 암묵지는 의도적으로 형식화 할 경우 그 가치가 낮아질 수 있으며 암묵지는 암묵적인 방식으로 관리되어야 함을 강조하였다. 지식관리와 관련된 다른 연구들에서는 전문가의 지식을 억지로 코드화하기 보다는 조직내의 전문가를 효과적으로 검색할 수 있는 방법을 제공하는 것을 암묵지를 관리할 수 있는 효과적인 방법의 하나로 제안하였다[Alavi and Leidner, 2001; Desouza, 2003]. 이에 따라 몇몇 상업용 지식관리시스템 엔진의 경우 실제로 전문가 검색 기능을 탑재하기 시작하였다[Handysoft Microsoft Verity].

전문가 검색 기능을 활용하기 위해서는 우선 조직 내의 전문가에 대한 정보, 즉 각 전문가의 전문 분야와 전문 지식의 수준 등이 파악되어야 하는데 이와 같은 전문가 프로파일을 작성하기 위해서는 크게 두 가지 방법이 있다. 그 하나는

시스템 관리자나 전문가 자신이 시스템에 각 전문가의 전문 분야와 전문 지식의 정도를 수동으로 등록하는 방식이다. 또 다른 방법은 각 전문가에 의한 지식관리시스템에서의 지식 활동을 기반으로 프로파일 정보 수집을 자동화하는 것이다. 전문가 검색 기능을 제공하는 대부분의 상업용 지식관리시스템 패키지에서는 전자의 방식을 채택하고 있는데 이 방식이 구현의 용이함을 제공하는 반면 취합된 전문가 정보를 최신의 것으로 유지하기 위한 지속적인 노력을 요구하는 단점을 가진다. 이 외에도 1장에서 설명한 바와 같이 많은 단점을 가지는 수동 전문가 프로파일 구축 방식을 보완하기 위해 객관적인 기준에 의해 프로파일 정보를 자동으로 수집하고 전문가 데이터베이스를 구축하는 효과적인 방법이 필요하다 할 수 있다.

지식관리시스템 사용자에게 의한 지식 생성 활동에는 지식베이스에 문서 파일을 업로드하거나, 전자게시판에 글을 게시하는 것, 전자우편이나 메모를 주고 받는 행위 등이 포함된다. 지식 생성 활동의 결과물은 지식관리시스템 사용자의 전문 지식을 나타내는 것으로 볼 수 있는데 시스템의 관점에서 볼 때 사용자의 전문 지식을 분석할 수 있는 유일한 방법은 사용자와 시스템과의 상호 작용에 의한 결과물을 분석하는 것이다. 이러한 관점에서 전문가 프로파일 정보 추출을 자동화하는 것은 시스템에 등록되는 지식물의 내용을 이해하고 이를 미리 정의된 주제 영역에 대해 분류해 내는 과정이라고 볼 수 있다.

지식관리시스템 사용자에게 의한 지식 생성 활동의 결과로 생성되는 지식물의 내용은 대부분 단순 텍스트 정보이거나 문서와 같이 텍스트 형태로 변경 가능한 것이다. 따라서 자동 문서 범주화 기법은 수동적인 관리자의 개입 없이도 시스템에 등록된 대부분의 지식물에 대한 분류가 가능하도록 하는 가장 이상적인 방법이다. 컴퓨터에 의해 텍스트로 이루어진 문서의 내용을 파

악하여 미리 정의된 범주를 할당하는 자동 문서 범주화 기법에 관한 연구는 문서 범주화 기법의 효율을 높이기 위한 알고리즘 관련 연구와 이를 다양한 분야에서 활용하기 위한 연구 등을 포함하여 활발히 진행되고 있다[Baeza-Yates and Riberio-Neto, 1999 Lam, Ruiz, and Srinivasan, 1999 Lee, Chuang, and Seamons, 1997 Zack, 1999]. 본 연구에서도 관리자의 관리 노력을 최소화하면서 등록된 지식물을 자동으로 분류하여 해당 지식물을 등록한 시스템 사용자의 전문 분야를 구분해 내는데 자동 문서 범주화 기법을 활용하고자 한다.

Ⅲ. 전문가 프로파일과 전문가 자동 분류기

이 장에서는 본연구에서 제안하는 전문가 프로파일의 구조와 정의된 전문가 프로파일 구조에 따라 전문가의 전문 분야 및 전문 지식의 정도를 측정하기 위해 설계된 전문가 자동 분류기를 Vector Space Model[Baeza-Yates and Riberio-Neto, 1999 Salton and Lesk, 1968]이라는 문서 범주화 기법을 이용하여 학습시키는 과정을 소개한다.

3.1 기본 개념정의

본 연구에서 제안하는 전문가 자동 분류의 과정을 자세히 설명하기에 앞서 몇 가지 개념에 대한 정의가 필요한데 그 정의와 설명은 다음과 같다. 첫째, *전문가 프로파일(Expert Profile)*이란 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 프로필 정보를 항상 최신의 것으로 유지하고 그 유용성을 보존하기 위해서는 지속적이고 시스템화된 전문가 프로필 갱신 방안이 필요하다. 이러한 관점에서 전문가 프로필 관리 기법의 자동화는 수동 프로필 관리 방법의 단점을 극복할 수 있는 방안으로 볼 수 있다.

한편, 특정 사용자의 전문 지식을 그 사람이 시스템에 등록한 지식물의 내용을 기반으로 결정하기 위해서는 다음과 같은 가정이 필요하다. "지식을 등록한 사용자는 해당 지식에 대한 내용을 완벽히 이해하고 있으며 등록한 지식에 대한 최초 등록자이다." 이 가정이 의미하는 바는 등록된 지식을 단순히 다른 출처로부터 복사하여 시스템에 등록한 것이 아닌 해당 지식의 내용을 제공할 수 있는 정도의 해당 분야에 대한 전문 지식을 지식 등록자가 가지고 있다는 것이다. 이러한 가정이 성립한다면, 우리는 특정 전문 분야에 대한 전문 지식을 보유하고 있는 전문가를 검색하는데 사용되는 전문가 프로필 데이터베이스를 구축할 수 있게 된다. 시스템의 관점에서 보면 전문가의 전문 분야를 구분한다는 것은 시스템과 해당 전문가의 상호작용의 결과를 분석하는 것이 유일한 방법으로 볼 수 있는데 이러한 시스템과의 상호 작용에는 파일을 업로드 혹은 다운로드하거나 게시판에 글을 등록하고 전자우편이나 메모를 주고 받는 등의 행위가 포함된다. 이러한 지식 활동의 결과물은 행위의 대상이 되는 등록된 파일, 게시된 글, 전자우편 내용 등이 된다. 따라서, 이러한 상호 작용의 결과물을 분석하여 해당 전문가 혹은 사용자의 전문 분야를 분석하는 것이 전문가 프로필 수집 및 갱신을 자동화할 수 있는 방법이라 할 수 있다.

전문가 프로필 데이터베이스를 구축하기 위해서는 전문가가 보유하고 있는 전문 지식의 수준 역시 측정되어야 한다. 전문 지식은 (1) 활

동성(activeness), (2) 우수성(excellence), (3) 평가(assessment)의 세가지 요소를 기준으로 측정될 수 있다. 본 연구에서 활동성은 전문가가 시스템에 지식물을 등록하는 지식 생성 활동을 얼마나 빈번히 수행하는지를 의미한다. 활동성은 또한 빈도(frequency), 최신도(recency), 분량(volume)의 요소로 구성되어 있다고 볼 수 있다. 빈도란 특정 전문가에 의해 시스템에 등록된 지식물의 숫자를 의미한다. 더 많은 지식을 등록한 전문가가 해당 분야에 대한 전문 지식을 상대적으로 더 많이 보유하고 있다고 볼 수 있는 것이다. 지식은 시간이 지남에 따라 그 가치가 변하게 되는데 최신도는 이러한 지식의 특성을 지식물의 평가에 반영하도록 한다. 따라서, 지식물은 그 가치를 평가함에 있어 시간 요소에 의해 조정되어야 할 필요가 있는 것이다. 마지막으로 분량 요소는 양이 방대한 지식을 등록한 전문가에 대한 고려를 위해 도입된다. 일반적으로 말해서, 분량이 큰 지식물이 더욱 많은 정보를 전달할 수 있다고 볼 수 있는데 따라서 시스템에서는 이러한 분량이 큰 지식물을 등록하는데 드는 시간과 노력을 인정해 줄 필요가 있는 것이다. 아래의 식 (1)은 앞에서 설명한 전문 지식의 활동성을 수식으로 표현하고 있다

$$\sum_{i=1}^n w_j(T_i) * w_j(V_i) \quad (1)$$

식 (1)에서 T_i 와 V_i 는 지식물 i 가 등록된 시간과 그 분량 요소를 각각 의미한다. 또한, w_j 는 주제 분야 j 에서의 최신도에 대한 가중치를 의미하며 마찬가지로 w_j 는 분야 j 에서의 분량 가중치를 의미한다. 각 주제 분야는 지식베이스에 등록되어 있는 전체 지식물의 내용이 서로 다르게 되는데 이에 따라 새롭게 등록되는 지식물에 대한 최신도와 분량 요소를 결정하기 위해 사용되는 기준이 서로 다르게 적용 된다. 따라서, 이러한 주제 분야간의 차이를 고려한 최

신도와 분량에 대한 가중치를 적용하여 전문 분야 지식 측정치를 수정해 줄 필요가 있다. 예를 들어, T_i 의 경우 특정 기준 시점으로부터 해당 지식물이 등록된 시점까지의 일수 혹은 경과 시간 등을 이용할 수 있으며 V_i 는 해당 지식물의 단어 혹은 글자수 등을 활용할 수 있다. 또한 이들에 대한 가중치인 w_{rj} 와 w_{vj} 는 해당 분야에 따라 달라지게 되는데 이는 지식물이 등록되는 빈도와 평균적인 지식물의 분량 등이 분야마다 다른 양상을 보이기 때문이다.

한편, 빈도 요소의 경우 다른 요소의 취합을 통해 자동적으로 반영되는 것으로 볼 수 있는데 이는 모든 지식물이 이러한 활동성 계산에 의해 영향을 받기 때문이다. 즉, 전문가가 더 많은 지식물을 등록할수록 해당 전문 분야에 대한 전문 지식을 해당 전문가가 더 많이 보유하고 있다고 생각할 수 있다. 특정 전문가에 대해서 식 (1)을 이용해 계산된 모든 값의 합은 특정 전문 분야에 대한 해당 전문가의 활동성의 정도를 나타낸다.

전문 지식을 구성하는 두번째 요소인 우수성은 등록된 지식물이 얼마나 특정 주제 분야에 적합한가를 의미한다. 등록된 지식물이 속하는 주제 분야와 해당 분야에의 적합도를 계산하는 것이 가능하다면 우리는 이러한 요소를 이용하여 해당 지식물이 특정 주제 분야를 나타내는데 얼마나 우수인가를 결정할 수 있게 된다. 우수성은 다음과 같은 식을 이용하여 계산될 수 있다

$$\sum_{i=1}^n S_j(D_i) \quad (2)$$

식 (2)에서 D_i 는 문서 혹은 지식물 i 를 S_j 는 해당 지식물의 주제 분야 j 에 대한 우수성 점수를 각각 의미한다. 우수성 점수 S_j 는 학습용 문서를 이용하여 미리 학습 과정을 거친 전문가 자동 분류기에 의해 측정되는데 전문가 자동 분류기는 해당 지식물에 대해서 주제 분야 계

계(subject field hierarchy)에 등록되어 있는 모든 분야와의 적합도를 각각 계산하여 제공한다. 즉, 모든 분야와의 적합도 중 가장 높은 점수를 보이는 분야가 해당 지식물이 속하는 분야이자 해당 적합도 점수가 그 분야에 대한 우수성 점수가 된다. 계산의 복잡도를 고려하여 본 연구에서는 최고 점수 하나만을 활용한다. 그러나 만약 최근 연구의 경향인 복합 분야 지식으로의 확장을 고려한다면 적합도 계산 목록에서 상위 두 개 혹은 세 개의 점수를 활용할 수 있겠다.

평가 역시 등록된 지식물의 유용성을 평가하는데 중요한 요소 중 하나로 볼 수 있다. 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 평가를 전문 지식을 측정하는 과정에 포함시키는 것이 필요하다. 평가 요소는 다음과 같이 표현될 수 있다.

$$\sum_{i=1}^n \sum_{k=1}^m A_k(D_i) \quad (3)$$

식 (3)에서 A_k 는 사용자 k 에 의해 평가된 지식물 i 에 대한 평가 점수를 나타낸다. 하나의 등록된 지식물은 한 명 이상의 사용자에게 의해 활용되고 평가될 수 있으므로 지식물 $i(D_i)$ 의 사용자로부터 받은 모든 평가 점수를 합산하여 해당 지식물의 평가 점수를 도출하게 된다. 또한 특정 전문가가 등록한 모든 지식물에 대한 평가 점수를 합산하면 해당 전문가의 전체 평가 점수가 되는데 식 (3)은 이것을 표현하고 있다.

둘째, 지식 생성 활동(Knowledge Creation Activity)이란 다른 사용자가 새로운 지식을 습득하는데 도움이 되는 지식물을 지식관리시스템에 추가하는 사용자의 모든 행위를 의미한다. 많은 조직에서 활용하고 있는 일반적인 지

식관리시스템에서 본 연구에서 정의한 지식 생성 활동의 예를 들면 파일을 업로드하거나 게시판에 글을 게시하는 것과 같이 지식물을 시스템에 등록하는 것과 전자우편 또는 메모를 통해 다른 사용자와 연락을 취하는 등의 행위가 있다. 시스템의 관점에서 보면 이러한 활동의 결과로 발생하는 콘텐츠가 해당 행위의 주체가 되는 사용자의 관심이나 전문 지식을 평가하기 위해 활용될 수 있는 유일한 대상이라 할 수 있다. 다시 말해서 수동으로 등록하는 방식을 고려하지 않는다면 사용자와 시스템간의 상호작용을 추적하여 분석하는 것이 해당 사용자가 보유하고 있는 전문 지식을 판별하는 가장 합리적인 방법이라 할 수 있다. 다만 이를 위해 필요한 가정은 관련된 지식물이 실제로 해당 사용자가 이해하고 있는 것이어야 한다는 점이다. 본 연구에서는 사용자의 전문 지식을 분석하고 측정하는데 이러한 지식 생성 활동의 결과인 지식물을 이용하게 된다.

셋째, 어떤 전문가가 어떤 주제 분야에 대한 전문 지식을 보유하고 있는지를 찾아내고 각 주제 분야에 대한 전문 지식의 수준을 측정하는 일련의 과정을 *전문가 식별(Expert Identification)*로 정의한다. 전문가 프로필 정보 수집 과정을 통해 일단 전문가 프로필 데이터베이스가 구축되면 시스템 사용자는 구축된 데이터베이스를 이용하여 필요한 전문 지식을 보유한 전문가를 검색하기 위한 질의를 수행할 수 있게 된다. 따라서, 전문가 식별과정은 그 선행 조건의 일부로 전문가 프로필 정보 수집 단계를 포함하게 된다.

넷째, *전문가 점수(Expert Point)*란 한 전문가가 지식 생성 활동을 통해 특정 주제 분야에서 획득한 점수의 총합을 의미한다. 전문가 점수는 전문가 프로필 정보 수집 과정에서 전문가 자동 분류기에 의해 앞서 설명한 전문 지식 수준 측정 알고리즘을 이용하여 자동으로 도출된다. 즉, 전문가의 전문 지식을 측정하기 위한

기준으로 제시된 활동성, 우수성, 그리고 평가의 항목 점수를 집계하여 적합한 가중치를 적용한 후 이를 해당 전문가의 전문가 점수로 활용하게 된다.

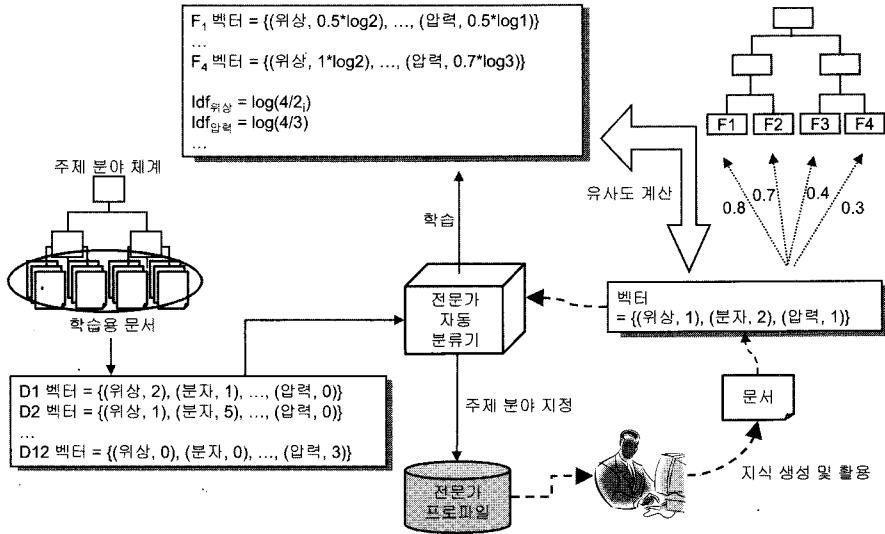
3.2 문서 범주화를 이용한 전문가 자동 분류기의 학습

새로 등록된 지식물을 적합한 주제 분야에 따라 분류하기 위해서는 전문가 자동 분류기를 학습용 문서를 이용하여 학습시키는 과정이 필요하다. 이 절에서는 벡터 공간 모형(Vector Space Model)이라는 문서 범주화 기법을 기반으로 전문가 자동 분류기를 학습시키는 과정을 설명한다. 또한 학습 과정과 함께 학습된 분류기를 이용하여 전문가 프로필 정보를 추출하는 과정도 자세히 다루고자 한다.

문서간의 유사 정도를 측정하고 분류하기 위한 문서 범주화 관련 연구가 지금까지 활발히 진행되어 왔는데 잠재 의미 색인(Latent Semantic Indexing), 베이저언 확률 모델(Bayesian Probabilistic Model), 벡터 공간 모형 등과 같은 다양한 기법들이 제안되었다. 많은 문서 범주화 기법 중, 본 연구에서는 다양한 관련 연구에 채택되어 좋은 문서 범주화 성능을 보인 벡터 공간 모형을 채택하여 전문가 분류기를 학습시키고 지식물을 분류하는데 이용한다[Baeza-Yates and Riberio-Neto, 1999]. 본 연구에서는 벡터 공간 모형을 이용한 전문가 자동 분류기의 학습 과정을 다음과 같은 두 단계로 보았다.

- 단계 1. 학습용 문서 준비
- 단계 2. 분야별 통합 문서 벡터 생성

<그림 1>은 위에 제시한 단계 1과 단계 2를 거치는 전문가 자동 분류기의 학습 과정과 학습된 분류기를 이용하여 전문가 프로필 정보를 생성하는 과정을 보여준다. 전문가 프로필



<그림 1> 전문가 자동 분류

생성 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. 이러한 전문가 프로필 자동 생성 과정은 IV장에서 자세히 설명하고 다음 절에서는 전문가 자동 분류기의 학습 과정에 대해 자세히 다루고자 한다.

3.2.1 학습용 문서 준비

전문가 자동 분류기를 학습시키기 위해서는 미리 그 주제 분야가 결정된 학습용 문서가 준비되어야 한다. 학습용 문서를 이용하여 전문가 자동 분류기를 효과적으로 학습시키기 위해서는 다음과 같은 두 가지 기본 가정이 필요하다. 첫째, 학습에 사용되는 각 문서에는 정확한 주제 분야가 할당되어 있다. 둘째, 특정 주제 분야에 속하는 전체 학습 문서 집합은 해당 문서 집합에서 사용된 어휘들을 통해 그 주제 분야를 완벽하게 표현한다. 이러한 사실이 성립된다고 가정하면, 학습용 문서들로부터 의미 있는

어휘들을 추출하여 각 주제 분야별 벡터 공간을 생성할 수 있는데 이렇게 생성된 벡터 공간은 새로 지식베이스에 등록된 지식물을 분류하는데 활용된다. 학습용 문서로부터 주제 분야에 대한 문서 벡터 공간을 생성하는 과정에서 필요한 개념들에 대한 정의는 다음과 같다.

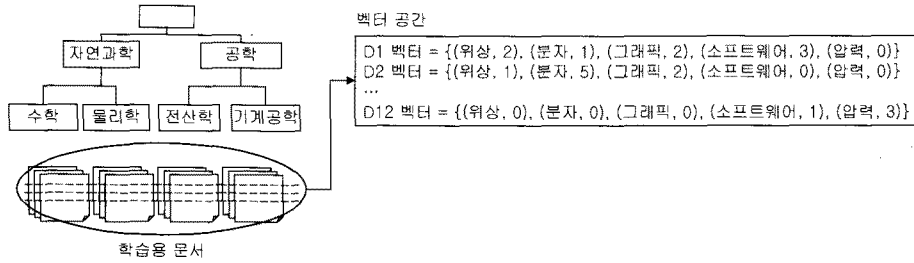
정의 1. 문서의 어휘 집합

TS(Term Set)는 식 (4)와 같이 표현될 수 있는데 관사나 접속사와 같은 문법 요소들을 제외한 해당 문서의 실제 내용을 대표하는 의미 있는 어휘들의 집합을 뜻한다.

$$TS = \{t_1, \dots, t_p\} \quad (4)$$

정의 2. 분야의 통합 문서

AD_j(Aggregated Document)는 주제 분야 j 에 속하는 문서 집합 $DS_j = \{D_{j1}, D_{j2}, \dots, D_{jd}\}$ 로부터 생성되는 통합 문서를 의미한다. 여기에서 $j = 1, \dots, N$ 이며 N 은 정의된 전체 주제 분야의 수 그리고 j 는 특정 주제 분야에 대한 색인을 의미한다. 또한 d 는 주제 분야 j 에 속하는 문서의 수를 나타낸다.



<그림 2> 벡터 공간 생성

정의 1은 주제 분야를 구분하고 주제 분야와 지식물의 유사도를 계산하는데 사용되는 모든 어휘들이 각 문서로부터 추출되어야 함을 의미한다. <그림 2>의 예와 같이 5개의 의미 있는 어휘들로 구성된 총 12개의 문서가 존재한다고 가정하자. 이 경우 전체 문서에 대한 어휘 집합 $TS = \{\text{위상, 분자, 그래픽, 소프트웨어, 압력}\}$ 이 되며 각 어휘와 개별 문서에서의 출현 빈도의 쌍으로 이루어진 어휘 벡터 공간을 생성하여 활용할 수 있게 된다

정의 2는 특정 주제 분야에 속하는 전체 문서에서 추출한 의미 있는 어휘와 그 출현 빈도의 쌍으로 구성된 통합 문서의 도출을 의미하는데 이렇게 구성된 통합 문서는 어휘와 그 출현 빈도를 기반으로 새로 등록된 지식물이 각 주제 분야에 얼마나 적합한지를 판단하는데 활용된다. <그림 2>의 예를 보면 수학, 물리학, 전산학, 기계공학의 네 분야가 존재하고 각 분야에는 세 개의 문서가 속하고 있다. 이러한 문서들이 각 분야별로 취합되어 분야 통합 문서를 생성하게 된다. 아래는 이러한 과정을 표현하고 있다.

$$DS_{\text{수학}} = \{D_{11}, D_{12}, D_{13}\} \rightarrow AD_{\text{수학}}$$

$$DS_{\text{물리학}} = \{D_{21}, D_{22}, D_{23}\} \rightarrow AD_{\text{물리학}}$$

3.2.2 분야별 통합 문서 벡터 생성

전문가 자동 분류기의 학습 과정 중 통합 문서 벡터 생성 단계에서는 각 주제 분야별로 취

합된 분야 통합 문서(AD_j)로부터 분야 벡터(FV_j)가 생성된다. 각 주제 분야에 대한 분야 벡터는 학습에 사용된 전체 문서 집합으로부터 추출된 분야 색인 어휘 집합(TS : term set)에 포함되어 있는 항목과 동일한 수의 요소를 포함하게 된다. 이 경우 분야에 따라 해당 어휘의 분야에 대한 관련성이 존재하지 않을 경우 출현 빈도가 0인 어휘들이 발생할 가능성이 있다. 특정 통합 문서 AD_j 에 대한 분야 벡터인 FV_j 는 전체 색인 어휘 집합에속하는 어휘들과 그 어휘들의 특정 주제 분야에 대한 관련성 혹은 설명력 수치로 구성된다. FV_j 에 대한 정의는 정의 3과 같다.

정의 3. FV_j : 분야 j 벡터

FV_j 는 전체 어휘 집합 TS 에 속하는 어휘와 각 어휘의 주제 분야 j 에 대한 설명력 혹은 관련성에 대한 가중치로 구성된다. FV_j 를 계산하기 위한 수식은 다음과 같다.

$$FV_j = \{(t_1, w_{j1}), (t_2, w_{j2}), \dots, (t_p, w_{jp})\} \quad (5)$$

여기에서 p 는 전체 색인 어휘의 수를 나타내며 어휘 tp 의 통합문서 AD_j 혹은 주제 분야 j 에 대한 설명력을 의미하는 w_{jp} 는 다음과 같이 계산된다.

$$w_{jp} = tf_{jp} \cdot id_{fp} \quad (6)$$

$$tf_{jp} = freq_{jp} / \max_1(freq_{j1}) \quad (7)$$

$$id_{fp} = \log(N/n_p) \quad (8)$$

통합 문서 AD_j 즉, 주제 분야 j 에서 어휘 t_p 의 상대적 중요도는 어휘 t_p 의 주제 분야 j 에 대한 어휘 빈도 tf_{jp} 를 이용하여 얻을 수 있으며 식 (8)의 n_p 는 어휘 t_p 가 출현하는 통합 문서의 수를 나타낸다. 어휘 빈도(term frequency)는 특정 문서에서 해당 어휘를 포함하는지의 여부를 고려하는 것으로 식 (7)의 $freq_{jp}$ 는 주제 분야 j 에서의 어휘 t_p 의 출현 횟수를 $\max_i(freq_{ji})$ 는 주제 분야 j 에서 가장 많이 출현하는 어휘의 출현 횟수를 의미한다. 또한, 역문서 빈도(idf: inverse document frequency)는 해당 어휘가 전체 문서 집합에서 출현하는지의 여부를 고려하는 것으로 전체 문서 집합에서 드물게 출현하는 어휘가 그 어휘를 포함하고 있는 문서를 전체 문서 집합의 다른 문서들과 구분하는데 더 중요하다는 것을 반영한다. 따라서, 위 식 (6)에서 idf_p 는 어휘 t_p 의 역문서 빈도를 나타내는데 식 (8)에서와 같이 그 중요도는 t_p 를 포함하고 있는 문서 수(n_p)에 역으로 비례한다.

<그림 2>의 예에서 네 개의 주제 분야에 대한 통합 문서 AD_j 가 가지는 각 색인 어휘의 빈도를 다음 <표 1>과 같이 가정하자.

<표 1>의 정보를 이용하여 정의 3에서 설명한 분야 벡터를 서로 다른 네 주제 분야에 대해 생성할 수 있다. 전체 어휘 집합 TS에 속하는 각 어휘의 설명력 즉, 해당 어휘가 특정 주제 분야를 얼마나 잘 설명하는지 혹은 해당 분

<표 1> 분야별 통합 문서에서의 단어 출현 빈도

	위상	분자	그래픽	소프트웨어	압력	최다 출현 단어의 출현 빈도수
수학	4	1	3	1	0	4
물리학	1	5	3	1	0	5
전산학	2	0	3	5	1	5
기계공학	0	0	0	1	3	3
단어 출현 분야수	3	2	3	4	2	

야에 얼마나 잘 맞는지를 계산해 낼 수 있는 것이다. 아래의 예는 '수학'과 '물리학'에 대한 어휘 설명력 (가중치) 계산과 두 주제 분야에 대한 분야 벡터 생성 과정을 보여주고 있다.

$$\begin{aligned}
 W_{\text{수학 위상}} &= 4/4 \log(4/3) = 0.12 \\
 W_{\text{수학 분자}} &= 1/4 \log(4/2) = 0.08 \\
 W_{\text{수학 그래픽}} &= 3/4 \log(4/3) = 0.09 \\
 W_{\text{수학 소프트웨어}} &= 1/4 \log(4/4) = 0 \\
 W_{\text{수학 압력}} &= 0/4 \log(4/2) = 0
 \end{aligned}$$

$$\begin{aligned}
 W_{\text{물리학 위상}} &= 1/5 \log(4/3) = 0.02 \\
 W_{\text{물리학 분자}} &= 5/5 \log(4/2) = 0.3 \\
 W_{\text{물리학 그래픽}} &= 3/5 \log(4/3) = 0.07 \\
 W_{\text{물리학 소프트웨어}} &= 1/5 \log(4/4) = 0 \\
 W_{\text{물리학 압력}} &= 0/5 \log(4/2) = 0
 \end{aligned}$$

$$\begin{aligned}
 FV_{\text{수학}} &= \{(위상, 0.12), (분자, 0.08), (그래픽, 0.09), (소프트웨어, 0), (압력, 0)\} \\
 FV_{\text{물리학}} &= \{(위상, 0.02), (분자, 0.3), (그래픽, 0.07), (소프트웨어, 0), (압력, 0)\}
 \end{aligned}$$

분야 벡터 생성 과정의 결과로부터 '위상'이라는 어휘가 주제 분야 '수학'에 대해 가장 높은 설명력을 가지며 '소프트웨어'와 '압력'이라는 어휘는 '수학'이라는 주제 분야와 전혀 관련이 없는 것을 볼 수 있다. 동일한 계산 과정에 따라 '물리학', '전산학' 그리고 '기계공학'과 같은 다른 분야에 대해서도 분야별로 5개의 어휘를 가지는 어휘벡터 공간을 구성할 수 있다. <표 2>는 주제 분야별로 각 어휘가 가지는 상대 중요도 가중치를 보여주는 주제 분야 벡터 계산 결과를 요약하고 있다.

<표 2> 벡터 공간

	위상	분자	그래픽	소프트웨어	압력
수학	0.12	0.08	0.09	0	0
물리학	0.02	0.30	0.07	0	0
전산학	0.05	0	0.07	0	0.06
기계공학	0	0	0	0	0.30

IV. 전문가 자동 분류 과정

이 장에서는 학습 과정을 거친 전문가 자동 분류기를 통해 등록된 지식물을 자동으로 분류하여 전문가 프로파일 정보를 생성하는 방법을 소개하고 본 연구에서 제안하는 전문가 분류 자동화 방법론을 적용하여 개발된 원형시스템을 소개하고자 한다.

4.1 전문가 자동 분류기를 이용한 전문가 프로파일 생성

전문가에 의해 생성된 지식물은 해당 전문가의 전문 지식을 분류하는데 활용될 수 있는데 <그림 1>은 지식베이스에 저장된 지식물을 기준으로 자동으로 해당 전문가의 전문 주제 분야를 분류하고 전문 지식의 수준을 계산하는 과정을 통해 기본적인 전문가 프로파일 정보를 생성하는 과정을 보여주고 있다. 문서를 업로드하거나 게시판에 글을 게시하는 등의 지식 생성 활동은 미리 학습 과정을 통해 학습된 전문가 자동 분류기를 구동하게 된다. 전문가 자동 분류기를 통해 분류된 지식물은 전문가 프로파일을 생성하는데 활용되며 최신 전문가 프로파일 정보들로 구성된 전문가 데이터베이스는 시스템 사용자의 요구에 부합하는 전문가를 검색하는데 이용된다.

지식관리시스템에서 특정 분야의 전문가로 분류되기 위해서는 시스템으로부터 일정 이상의 전문가 점수를 획득해야만 하는데 지식물이 지식베이스에 등록되면 전문가 자동 분류기는 해당 지식물에 분야를 할당하고 지식물의 종류에 따라 3장에서 설명한 알고리즘에 의해 계산된 전문가 점수를 그 지식물을 등록한 전문가에게 부여하게 된다. 전문가가 특정 주제 분야에서 일정 이상의 전문가 점수를 획득하면 해당 전문가는 그 주제 분야에 대한 전문가로 분류되며 동시에 전문가 검색의 대상이 된다. 전

문가 점수는 또한 각 전문가의 특정 주제 분야에 대한 전문 지식의 정도를 나타내는 척도로도 이용된다.

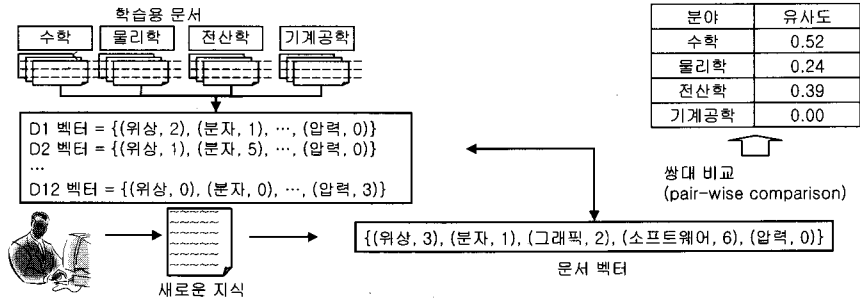
지식물에 대한 분류 과정은 전문가 자동 분류기의 학습 과정과 매우 유사한데 학습 과정에서는 미리 분류된 학습 문서가 사용되지만 실제 분류 과정에서는 등록된 지식물이 사용된다는 차이가 있다. 학습된 전문가 분류기가 전문가 프로파일을 생성하는 과정은 다음의 단계를 따르게 된다.

- 단계 1. 지식물 등록
- 단계 2. 문서벡터 생성
- 단계 3. 주제분야별 유사도 계산
- 단계 4. 전문가 프로파일 데이터베이스 갱신

<그림 1>에 나타난 것처럼 지식물이 지식베이스에 등록되면 (단계 1) 전문가 자동 분류기가 호출되어 새로 등록된 지식물을 분석하게 된다. 우선 분류기는 지식물로부터 문서 벡터를 생성하게 되는데 (단계 2) 이 단계까지는 전문가 분류기의 학습 과정과 매우 유사하다. 등록된 지식물로부터 생성된 문서 벡터는 문서 대 분야의 쌍대 비교를 통해 각 주제 분야와의 유사도를 측정하는데 이용된다 (단계 3). 이 과정에서 벡터 유사도 계산법을 도입하여 활용하는데 구체적인 계산 과정은 4.2절에서 소개하고자 한다. 이전 단계의 결과로 선택된 주제 분야와 해당 주제 분야에 대한 유사도는 전문가 프로파일 데이터베이스를 갱신하는데 이용된다 (단계 4).

4.2 전문 분야별 유사도 계산

일단 전문가 자동분류기가 학습용 문서를 이용한 학습 과정을 거치면 전문가로부터 등록된 지식물을 분류하는데 활용될 수 있다. 특정 지식물에 가장 적합한 혹은 관련 있는 주제 분야를 할당하기 위해서는 새로 등록된 지식물에서



<그림 3> 등록 지식의 분류

생성된 문서 벡터와 주제 분야별로 학습 과정에서 생성된 분야 벡터 간의 쌍대 비교 과정이 필요하다. 이는 같은 분야에 속하는 모든 학습 문서가 취합되어 하나의 통합 문서를 생성하고 새로 등록된 지식물과 이렇게 통합된 문서간의 유사도가 계산됨을 의미한다. 계산된 유사도 중 가장 높은 값이 자신이 속한 주제 분야에 대한 새로운 지식물의 적합도를 나타낸다.

전문가 자동 분류기의 학습 과정으로부터 도출된 분야 벡터는 각 어휘의 개별 주제 분야에 대한 설명력을 포함하고 있기 때문에 새로 등록된 지식물로부터 생성된 문서 벡터와의 유사도를 코사인 유사도 함수(Cosine Similarity Function)를 이용하여 요소 단위 비교를 통해 측정하는 것이 가능하다[Baeza-Yates and Ribiero-Neto, 1999]. 일반적으로 두 벡터 사이의 유사도는 두 벡터로부터 계산되는 cos 값으로 표현할 수 있는데 다음의 식 (9)는 이러한 코사인 유사도 함수의 수학적 표현을 나타낸다

$$Sim(V_1, V_2) = \cos \theta = \frac{V_1 \cdot V_2}{\|V_1\| \times \|V_2\|} \quad (9)$$

여기에서 $V_1 \cdot V_2$ 는 두 벡터 V_1 과 V_2 의 내적(inner product)을 나타내며 $\|V_1\|$ 은 벡터 V_1 의 절대값을 표시한다. 예를 들어, 어느 전문가에 의해 지식물이 지식베이스에 등록되었다고 가정하자. 계산의 복잡성을 고려하여 새로 등록된 지식물로부터 생성된 문서 벡터 DV가 <그림 3>에서 보는 것처럼 단지 4개의 의미 있는 단어로 구성되었다고 가정하면 생성된 문서 벡터는 {(위상, 3), (분자, 1), (그래픽, 2), (소프트웨어, 6), (압력, 0)}과 같다. 문서 벡터의 경우 벡터 내의 각 요소는 어휘와 해당어휘의 문서 내 출현 빈도의 쌍이 되며 분야 벡터는 출현 빈도 대신 각 어휘의 상대 중요도 가중치를 포함하게 된다. 이 경우 지식물을 분류하는데 있어서 한번에 하나의 문서만을 고려하게 되고 새로운 문서와 분야별 통합 문서 간의 상대적 유사도만을 고려하게 되므로 유사도 계산에서 문서 내 어휘의 출현 빈도를 상대적 중요도 가중치 값 대신 사용할 수 있다.

<그림 3>의 예에서 새로 등록된 지식물과 주제 분야 '수학'의 유사도를 계산한다면 다음과 같은 과정을 거치게 된다. '수학'의 분야 벡터 FV수학 과 문서 벡터 DV 는 각각(0.12, 0.08, 0.09, 0, 0) 과 (3, 1, 2, 6, 0) 이다. 따라서, 두 벡터 사이의 유사도는 다음과 같이 구할 수 있다

다음과 같은 과정을 거치게 된다. '수학'의 분야 벡터 FV수학 과 문서 벡터 DV 는 각각(0.12, 0.08, 0.09, 0, 0) 과 (3, 1, 2, 6, 0) 이다. 따라서, 두 벡터 사이의 유사도는 다음과 같이 구할 수 있다

$$\begin{aligned}
 Sim(FV_{\text{수학}}, DV) &= \frac{FV_{\text{수학}} \cdot DV}{\|FV_{\text{수학}}\| \times \|DV\|} \\
 &= \frac{(0.12 \times 3) + (0.08 \times 1) + (0.09 \times 2)}{\sqrt{0.12^3 + 0.08^2 + 0.09^2 + 0^2 + 0^2} \times \sqrt{(0 \times 6) + (0 \times 0)}} \\
 &= \frac{0.62}{0.17 \times 7.07} \\
 &= 0.52(\text{최종유사도})
 \end{aligned}$$

<표 3>은 새 지식물과 4개의 모든 주제 분야 간의 유사도 계산 결과를 보여 주고 있는데 <표 3>의 가장 마지막 열에 나타난 바와 같이 해당 지식물은 4개의 주제 분야와 0.52, 0.24, 0.39, 0.00의 유사도를 가진다. 따라서, 이 경우 지식물이 가장 높은 유사도 값을 가지는 '수학'이 해당 지식물이 속하는 주제 분야로 선택되며 유사도 0.52는 선택된 주제 분야 '수학'에 대한 지식물의 적합도를 의미한다. 일단 전문가 자동 분류기에 의해 주제 분야가 할당되면 지식물의 종류에 따라 계산된 전문가 점수가 해당 지식물을 등록한 전문가에게 부여된다. 이렇게 특정 전문가에게 부여된 전문가 점수와 해당 전문가에 의해 등록된 지식물의 유사도 값은 전문가가 보유하고 있는 특정 분야에 대한 전문 지식의 수준을 평가하는데 활용될 수 있다

<표 3> 유사도 계산

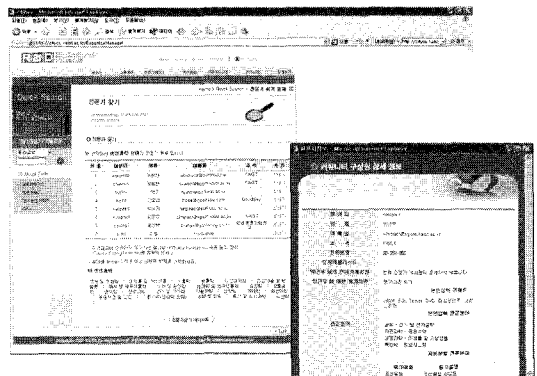
	FV · DV	FV	DV	Sim(FV, DV)
수 학	0.6375	0.1734	7.0711	0.52
물리학	0.5259	0.3112	7.0711	0.24
전산학	0.2999	0.1084	7.0711	0.39
기계공학	0.0000	0.3010	7.0711	0.00

4.3 지식 포탈에의 응용

본 연구에서 제안하는 전문가 자동 분류 프레임워크의 적용 가능성을 테스트하기 위해 전문가 분류 기능을 가진 지식관리시스템인 '지식 포탈'을 원형시스템으로 개발하였다. 정보 포탈(information portal) 혹은 기업 정보 포탈(EIP: enterprise information portal)은 다양한 형태의 정보시스템을 통합하고 이러한 정보시스템에 대한 효과적이고 효율적이며 동시에 개인화된 접근 경로를 제공하는데 적합한 시스템 아키텍처로 인식되고 있다[Deltor, 2000; Dias, 2001]. 이러한 포탈 시스템을 이용하면 사용자는 정보 원천의 종류에 상관없이 필요한 모든 정보에

친숙한 사용자 인터페이스인 웹 브라우저를 이용하여 접근하는 것이 가능하며 시스템은 개인의 사용자가 적시에 중요한 비즈니스 결정을 내리는데 도움을 줄 수 있는 개인별맞춤 정보를 제공하게 된다. 본 연구에서는 전문가 자동 분류 및 전문가 데이터베이스 구축을 통한 전문가 검색을 쉽고 사용자에게 친숙한 인터페이스를 통해 제공할 수 있는 웹 기반 지식관리시스템을 개발하는데 정보 포탈 아키텍처를 기본 구조로 선택하였다.

본 연구에서 제안하는 전문가 자동 분류 방법론은 현재 공개 서비스 중인 '과학기술 연구'를 위한 지식포탈(<http://www.z4you.net>)에 적용되었다. 전문가 자동 분류기는 자바 언어를 이용하여 개발되었으며 이를 위한 문서 범주화 엔진으로는 Carnegie Mellon University에서 개발된 RAINBOW[McCallum, 1996]를 이용하였다. <그림 4>는 이 장에서 설명한 전문가 자동 분류 기능을 지식 포탈에서 수행한 결과를 보여주는 화면 예이다.



<그림 4> 지식포탈에서의 전문가 자동 분류 화면 예

V. 요약 및 결론

본 연구에서는 지식관리시스템을 위한 전문가 자동 분류 방법론을 제안하였다. 조직에서 전문가를 식별하여 관리하는 것이 조직 내의

암묵지를 관리하는 효과적인 방법으로 인식되면서 적은 비용으로 전문가 프로파일 데이터베이스를 관리하는 방안의 필요성이 대두되었다. 본 연구에서 제안하는 전문가 자동 분류 방법을 도입하여 어떤 전문가가 어떠한 주제 분야에 대한 전문 지식을 보유하고 있는지 식별하고 그 전문 지식의 정도를 측정하여 관리하게 된다면 지식베이스에 저장된 전문가 프로파일 정보를 수동으로 관리하는데 드는 비용과 시간을 최소화하는 것이 가능하다. 동시에 프로파일 정보의 등록 오류나 갱신이 되지 않아 최신 정보를 반영하지 못하는 등의 문제도 해결할 수 있다.

본 연구에서는 또한 전문가가 보유하고 있는

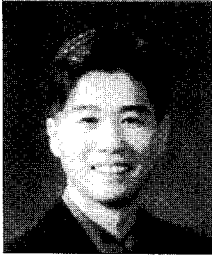
전문 지식에 대한 상세한 정보를 제공하기 위한 전문가 프로파일의 구조를 정의하였다. 자동으로 분석된 지식물은 그지식물을 등록한 전문가의 프로파일 정보를 갱신하는데 활용되며 지식관리시스템에 등록된 모든 전문가의 프로파일 정보를 취합하여 전문가 데이터베이스를 구축하게 된다. 제안된 전문가 자동 분류 방법론은 그 적용 가능성과 유용성을 테스트하기 위해 '과학기술 연구를 위한 지식포털'에 적용되었다. 최근 연구의 경향인 복합 분야(multi-disciplinary) 지식물의 채택이 가능하도록 전문가 자동 분류 방법론을 확장하는 것과 전문가 자동 분류기의 성능을 개선하기 위한 알고리즘 연구가 진행 중이다.

〈참 고 문 헌〉

- [1] Alavi, M. and Leidner, D.E., "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly*, Vol. 25, No. 1, March 2001, pp. 107-136.
- [2] Augier, M. and Vendelo, M.T., "Networks, Cognition and Management of Tacit Knowledge," *Journal of Knowledge Management*, Vol. 3, No. 4, 1999, pp. 252-261.
- [3] Baeza-Yates, R. and Riberio-Neto, B., *Modern Information Retrieval*, ACM Press, New York, 1999.
- [4] Bohn, R.E., "Measuring and Managing Technological Knowledge," *IEEE Engineering Management Review*, Winter 1997.
- [5] Buckingham-Shum, S.J. and Hammond, N., "Argumentation-based Design Rationale: What Use at What Cost?," *Human-Computer Studies*, Vol. 40, No. 4, April 1994, pp. 603-652.
- [6] Conklin, J.E. and Yakemovic, K.B., "A Process-oriented Approach to Design Rationale," *Human-Computer Interaction*, Vol. 6, No. 3-4, 1991, pp. 357-391.
- [7] Deltor, B. "The Corporate Portal as Information Infrastructure: Toward a Framework for Portal Design," *International Journal of Information Management*, Vol. 20, 2000, pp. 91-101.
- [8] Desouza, K.C., "Barriers to Effective Use of Knowledge Management Systems in Software Engineering," *Communications of the ACM*, Vol. 46, No. 1, January 2003, pp. 99-101.
- [9] Devedzic, V., "A Survey of Modern Knowledge Modeling Techniques," *Expert Systems with Applications*, Vol. 17, 1999, pp. 275-294.
- [10] Dias, C., "Corporate Portals: a Literature Review of a New Concept in Information Management," *International Journal of*

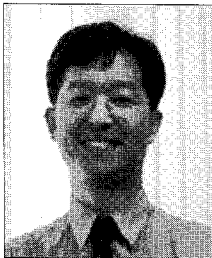
- Information Management, Vol. 21, 2001, pp. 269-287.
- [11] Gray, P.H., "A Problem-solving Perspective on Knowledge Management Practices," *Decision Support Systems*, Vol. 31, 2001, pp. 87-102.
- [12] Handysoft, BizFlow KMS, available on <http://corona.handysoft.co.kr/eng/>.
- [13] Hu, X. et al., "A Survey on Design Rationale: Representation, Capture and Retrieval," *Proceedings of 2000 ASME Design Engineering Technical Conferences*, September 10-13, 2000, Baltimore, Maryland.
- [14] Kreiner, K., "Tacit Knowledge Management: The Role of Artifacts," *Journal of Knowledge Management*, Vol. 6, No. 2, 2002, pp. 112-123.
- [15] Lam, W., Ruiz, M. and Srinivasan P., "Automatic Text Categorization and Its Application to Text Retrieval," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11, No. 6, 1999, pp. 865-879.
- [16] Lee, D.L., Chuang, H. and Seamons, K., "Document Ranking and the Vector-Space Model," *IEEE Software*, No. 2, 1997, pp. 67-75.
- [17] McCallum, A.K., Bow: A Toolkit for Statistical Language Modeling, Text Retrieval, Classification and Clustering, available on <http://www.cs.cmu.edu/~mccallum/bow>, 1996.
- [18] Microsoft, Microsoft SharePoint Products and Technologies, available on <http://www.microsoft.com/sharepoint/>.
- [19] Nonaka, I. and Takeuchi, H., *Knowledge Creating Company*, Oxford University Press, New York, 1995.
- [20] Nonaka, I., "A Dynamic Theory of Organizational Knowledge Creation," *Organization Science*, Vol. 5, No. 1, 1994, pp. 14-37.
- [21] Rus, I. and Lindvall, M., "Knowledge Management in Software Engineering," *IEEE Software*, No. 3, 2002, pp. 26-38.
- [22] Salton, G. and Lesk, M. E., "Computer Evaluation of Indexing and Text Processing," *Journal of the ACM*, Vol. 15, No. 1, 1968, pp. 8-36.
- [23] Verity, Verity K2 Architecture: Unprecedented Performance, Scalability and Fault Tolerance, *Verity White Paper*, available on http://www.verity.com/products/k2_enterprise/.
- [24] Zack, M., "Managing Codified Knowledge," *Sloan Management Review*, Summer 1999.

◆ 저자소개 ◆



양근우 (Yang, Kun-Woo)

현재 한국과학기술원 테크노경영대학원 경영공학전공 박사과정에 재학 중이다. 美 The Univ. of Maryland 에서 경영학 학사, 한국과학기술원 테크노경영대학원에서 경영정보전공 경영학 석사를 취득하였으며 주요관심분야는 지식관리 시스템, 전문가시스템, 지능형 에이전트 기술 등이다. 현재까지 경영과학회지, 지능정보시스템 학회논문지 등에 논문을 게재하였으며 DSI, PACIS, ICKIISS 등의 국제학술대회에서 논문을 발표하였다.



허순영 (Huh, Soon-Young)

현재 한국과학기술원 테크노경영대학원 부교수로 재직 중이다. 서울대학교 전자공학 학사, 한국과학기술원 경영과학 석사, 美 UCLA 경영정보학 박사 를 취득하였으며 MIS 관련 다양한 분야의 연구와 프로젝트를 수행하고 있다. 주요관심분야는 모델관리 시스템, 모바일 금융 시스템, 고객 관계 관리, 지식관리 시스템 등이다. 현재까지 Decision Sciences, Decision Support Systems, Omega, Journal of Database Management, Information and Software Technology 등에 논문을 게재하였으며 ICIS, DSI, PACIS, INFORMS 등의 국제학술대회에서 논문을 발표하였다. 또한 현재 한국경영과학회지와 한국데이터베이스 학회 등의 IT분야 편집위원이며 다양한 정부 부처 및 기업에서 자문 활동 중이다.

◆ 이 논문은 2003년 7월 22일 접수하여 1차 수정을 거쳐 2004년 2월 26일 게재확정되었습니다.