

지식포탈을 위한 개인화 지식 제공 방안

이홍주* · 김종우** · 김광래*** · 안형준**** · 권철현***** · 박성주*****

Personalized Information Delivery Methods for Knowledge Portals

Hong Joo Lee* · Jong Woo Kim** · Gwang Rae Kim*** · Hyung Jun Ahn****
Chul Hyun Kwon***** · Sung Joo Park*****

Abstract

In order to provide personalized knowledge recommendation services, most web portals for organizational knowledge management use category or keyword information that portal users explicitly express interests in. However, it is usually difficult to collect correct preference data for all users with this approach, and, moreover, users' preferences may easily change over time, which results in outdated user profiles and impaired recommendation quality. In order to address this problem, this paper suggests knowledge recommendation methods for portals using user profiles that are automatically constructed from users' activities such as posting or uploading of articles and documents. The result of our experiment shows that the proposed method can provide equivalent performance with the manual category or keyword selection method.

Keywords : Personalized Information Delivery, Knowledge Portal, User Profile, Information Filtering

논문접수일 : 2004년 3월 2일

논문게재확정일 : 2005년 11월 2일

※ 본 연구는 2004년도 중소기업청 산학협동과제로 국방품질관리소와 공동으로 수행된 연구 결과물이며, 연구수행에 참여하여 추천도
델 구현 및 실험 환경 구현에 도움을 준 충남대학교의 백승기, 임남구 군 및 한국전산원의 변현진 양에게 감사드립니다.

* 주저자, 한국과학기술원 테크노경영대학원 박사과정, (130-722)서울시 동대문구 청량리동 207-43번지,

Tel : 02-958-3647, Fax : 02-958-3604, e-mail : hjlee@kgsm.kaist.ac.kr

** 교신저자, 한양대학교 경영대학 경영학부 부교수, (133-791)서울시 성동구 행당동 17번지,

Tel : 02-2220-1067, Fax : 02-2220-1169, e-mail : kjw@hanyang.ac.kr

*** 충남대학교 통계학과 박사과정

**** Lecturer, Management Systems, Waikato Management School, University of Waikato

***** 삼성 SDS Biz 컨설팅센터 선임컨설턴트

***** 한국과학기술원 테크노경영대학원 교수

1. 서 론

많은 양의 정보들을 다양한 원천들로부터 수집하여 제공하며, 분산된 사용자들 간의 지식을 공유하고 의견을 나눌 수 있는 통합된 접근점으로서의 지식포털이 활용되고 있다[박성주 외 4인, 2003; Davydov, 2001; Stab & Maedche, 2001]. 지식포털(Knowledge Portal) 혹은 정보포털(Information Portal)은 사용자들에게 관련 정보 검색과 정보 제공을 통해 의사결정에 도움을 주기 위한 게이트웨이로서[Davydov, 2001], 지식의 통합 접근을 제공하며 지식의 상호교환이 가능하도록 하는 것을 주목적으로 하고 있다[Stab & Maedche, 2001]. 지식포털은 제공되는 내용(contents)과 대상 사용자에게 따라 구별할 수 있으며, 크게 공용(public), 기업, 커뮤니티, 개인, 모바일, 보이스 포털로 나누어 볼 수 있다[Turban et al., 2004]. 본 연구에서는 한 기업 내의 지식공유를 목적으로 하는 기업포털 보다는 다양한 원천으로부터 흡수된 많은 양의 정보와 지식을 제공하고, 사용자들 간의 의사소통 가능케 하는 공용과 커뮤니티 포털을 대상으로 하고 있다. 공용이나 커뮤니티 포털에서는 다양한 원천들에서 생성된 많은 정보와 지식이 모이게 되므로, 정보의 과부하로 인해 사용자 개인이 필요한 정보를 찾거나 새로운 지식 중 관련 있는 지식을 찾는 것이 매우 어렵다[박성주 외 4인, 2003; Aas, 1997; Davydov, 2001; Deltor, 2000; Lee et al., 1997; Stab & Maedche, 2001]. 이와 같은 정보의 과부하를 경감시키기 위해 사용자의 프로파일을 기반으로 사용자에게 적합한 지식을 개인화하여 제공하려는 노력과 연구들이 많이 진행되었다[Aas, 1997; Foltz & Dumais, 1992; Hijikata, 2004; Lee et al., 1997; Mostafa, 1997; Pazzani & Billsus, 1997; Resnick et al., 1994; Yan & Garia-Molina, 1995; Zampa &

Lemaire, 2002].

지식포털에서 사용되고 있는 개인화된 지식제공 방안은 크게 다음의 두 가지로 나눌 수 있다. 첫 번째는 사용자들에게 미리 정의된 지식맵에서 선호하는 분야들을 선택하도록 한 후에, 사용자가 선택한 분야에 속한 신규 지식들을 추천하여 주는 분류정보에 기반을 둔 방식이다. 두 번째는 사용자가 선호하는 키워드들을 등록하도록 한 후에, 등록된 키워드나 이들의 조합에 기반을 둔 검색 식에 해당하는 신규 지식들을 사용자에게 추천하는 키워드 기반 방식을 사용하고 있다. 분류정보에 기반을 둔 추천 방식은 세밀한 분류체계 생성의 어려움과 변경관리가 어렵기 때문에, 분류체계 구성이 정적이며 광범위하게 구성된다. 이로 인해 하나의 분야에 속한 지식의 양이 많아지게 되므로 지식 과부하를 줄이기가 어려우며, 지식 분야 분류체계의 변경이나 사용자 관심분야의 변경 시에 시스템이 자동적으로 적응하기 어렵다[김재경 외 2인, 2003]. 관심키워드나 이들의 조합으로 만든 검색 식에 기반을 둔 추천방식은 지식 내의 다양한 배경지식에 대한 정보는 고려하지 않고 단순히 특정 키워드가 존재하는지 만으로 지식을 평가하기 때문에, 적합한 지식의 선정 범위가 협소해지며 유사 단어 및 다른 의미의 단어들로 인해 적합하지 않은 지식이 추천될 수도 있다.

지식포털에서는 명시적으로 사용자 선호를 파악하거나 입력 받을 수 있을 뿐만 아니라, 사용자의 지식포털 내의 지식생성, 지식활용과 같은 지식활동으로부터 사용자가 어떤 문서를 생성하였는지 혹은 어떤 문서를 보았는지에 대한 정보를 암묵적으로 획득할 수 있다. 이를 통해 사용자의 관심 분야나 지식의 속성을 파악할 수 있으며, 파악된 속성을 통해 사용자 프로파일을 생성하여 사용자 프로파일과 가장 유사한 지식을 추천하는데 활용할 수 있다.

본 연구에서는 지식포탈 내에서 활용 가능한 개인화된 지식제공 방안들을 제시하고 이들의 성과를 비교하고자 한다. 지식포탈 내에서의 사용자의 지식활동을 통해 자동적으로 사용자의 프로파일을 생성할 수 있는 자동 분류기반 방식과 벡터 공간 모델에 기반을 둔 추천 방식을 제시하고, 각 방식의 사용자 프로파일 생성을 위한 방안들을 제시한다. 본 논문의 구성은 다음과 같다. 2장에서는 지식포탈과 개인화된 지식제공 방안에 관한 관련 연구에 대해 기술하며, 3장에서는 개인화된 지식제공 모델을 구성하는 정보제공 방식과 사용자 프로파일 구성 방안들을 제시한다. 4장에서는 인터넷을 통한 추천 성과 비교 실험내용 및 실험결과에 대해 기술하고, 5장에서 연구의 의의와 향후 연구 방향을 제시한다.

2. 개인화된 지식제공 방안과 지식포탈

개인화 지식제공 기술은 일종의 정보 필터링(Information Filtering) 기술이다. 정보 필터링은 정보를 필요로 하는 사람에게 적합한 정보를 제공하는 여러 과정 전체를 이르는 말이다[박상규 외 6인, 2003 ; Belkin & Croft, 1992]. 정보 필터링의 수행 과정은 <표 1>에 기술된 것과 같이 문서를 색인하여 표현하는 과정, 문서간의 유사도 혹은 사용자 프로파일과 문서간의 유사도 계산하는 과정, 사용자의 프로파일을 생성하는 과정으로 구분할 수 있다[Aas, 1997].

문서를 색인하여 표현하는 과정에는 색인에 불

필요한 단어 및 조사 등을 제거하는 것과 TF-IDF(Term Frequency-Inverse Document Frequency) 방안을 통해 추출된 단어의 중요도를 계산하는 것이 포함된다[Aas, 1997 ; Foltz & Dumais, 1992]. 유사도 계산에 활용되는 Boolean 모델은 특정 단어가 문서에 포함되어 있으면 1, 없으면 0 점을 부여하는 방안이며[Pazzani & Billsus, 1997], 벡터공간 모델은 문서에 포함된 단어와 단어의 비중을 벡터로 표현하여 벡터간의 유사도를 내적으로 계산하는 방안이다. 또한 문서 분류기법에 활용되는 확률정보 기반방식을 통해 유사한 문서와 유사하지 않은 문서를 구분하는 방안도 활용된다. 사용자 프로파일을 생성하는 것은 사용자의 선호도를 파악하기 위한 것으로 크게 사용자로부터 명시적으로 선호도를 얻는 방법과 사용자의 행위를 바탕으로 암묵적으로 파악하는 방안으로 나누어 볼 수 있다[Hijikata, 2004].

지식포탈은 다양한 지식원천들과 지식관리 기능들에 대한 통합 접근점으로 활용되며, 기능은 크게 의사소통, 협동작업, 내용관리, 업무조정, 개인화, 커뮤니티, 전문가연결 기능으로 구분할 수 있다[이홍주 외 3인, 2003 ; Dias, 2001 ; Deltor, 2000]. 국내외 지식포탈 솔루션들이 제공하는 개인화 기능에는 크게 사용자 인터페이스를 설정하는 기능과 제공되는 내용을 자신에게 맞게 설정하는 기능으로 이루어진다. 국내 연구기관 및 기업에서 활용하고 있는 지식포탈과 국내외의 대표적인 지식포탈 솔루션들의 개인화된 지식제공현황에 대해 정리한 것이 <표 2>이다.

<표 1> 정보 필터링 단계

단 계	설 명
문서색인 및 표현	문서에서 색인에 불필요한 단어 및 조사 제거 추출된 단어의 중요도 계산 (예. TF-IDF)
유사도 계산	Boolean 모델, 벡터공간 모델, 확률정보 기반 방식
사용자 프로파일 생성	사용자 선호도 파악 명시적 - 선호 카테고리, 키워드 등 암묵적 - 방문시간, 마우스 클릭 여부 등

〈표 2〉 지식포탈들의 개인화된 지식제공 현황

지식포탈/지식포탈 솔루션	개발회사/운영기관	지식제공 매체	지식 개인화방안
Hummingbird Portal [Hummingbird]	Hummingbird	My page, 이메일	분류기반, 키워드기반, 원천기반
Plumtree Corporate Portal [Plumtree]	Plumtree	Personalized Page	분류기반, 키워드기반
ACUBE [ACUBE]	삼성 SDS	My Map, Keyword Map	분류기반, 키워드기반
BizFlow EIP [BizFlow]	핸디소프트	My Map	분류기반, 키워드기반
KISTI (www.kisti.re.kr)	한국과학기술정보연구원	My page, 이메일	분류기반, 키워드기반, 검색식
ITFIND (www.itfind.or.kr)	정보통신연구진흥원	My Page (My Map, Query 검색) 이메일, Webzine	키워드기반, 검색식, 관심 Webzine 선택
SERI (www.seri.org)	삼성경제연구소	이메일	분류기반
국가지식 정보통합 검색시스템 (www.knowledge.go.kr)	한국전산원	이메일	분류기반, 키워드기반

사용자에게 개인화하여 지식을 제공하는 기능은 대부분 나의 페이지에서 자신의 지식맵을 구성하고, 구성된 지식맵에 해당하는 지식을 나의 페이지 내에 보여주는 방식으로 이루어진다. 자신의 지식맵을 구성하는 방안은 지식포탈 내에 구성된 지식 분류체계에서 선호하는 지식 분류를 선택하여 지정하는 분류기반 방식과, 관심 키워드를 등록하면 해당 키워드가 들어 있는 지식을 추천하는 방식으로 이루어지는 키워드기반 방안이 있다. 또한 사용자가 설정한 지식 원천에 새로 등록된 지식을 보여주는 방식으로 이루어지는 원천기반 방식이 있다. 앞의 방식들은 모두 사용자가 시스템 내에서 명시적으로 선호분야나 선호 키워드를 등록하여야만 추천이 가능하다. 지식포탈 솔루션들이 제공하는 지식추천방안도 위와 유사하게 솔루션들이 제공하는 지식분류에서 선호 분야를 선택하거나 관심 있는 키워드를 입력하는 방안을 통해 사용자의 프로파일을 생성하고 개인화된 지식을 제공한다.

현재 운영 중인 지식포탈들이나 지식포탈 솔루션들은 명시적으로 사용자 선호를 입력 받고 있다. 하지만, 사용자의 지식포탈 내의 지식생성, 지식활용과 같은 지식활동으로부터 사용자가 어떤 문서를 생성하였는지 혹은 어떤 문서를 보았는지에 대한 정보로부터 선호도를 암묵적으로 획득할 수 있다. 다음 장에서는 지식포탈 내의 사용자 활동을 통해 자동적으로 생성되는 사용자 프로파일을 작성하고 이에 기반을 두어 사용자에게 적합한 지식을 제공하는 개인화 방안을 제시하도록 한다.

3. 지식포탈에서의 개인화된 지식제공 방안

지식포탈의 사용을 통해 사용자가 관심 있는 분야나 키워드들을 추출하여, 적합한 지식을 제공하는 방식에는 크게 분류정보를 활용하여 선호하는 분야를 자동으로 파악하는 자동분류기반

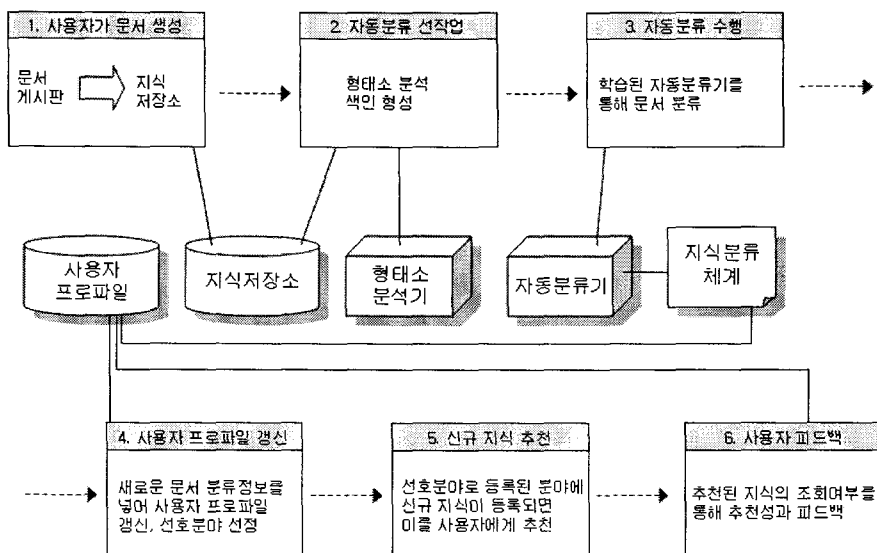
방식과 지식 내에 들어있는 키워드들을 이용하여 사용자 프로파일을 생성하고, 사용자 프로파일과 지식을 비교하여 가장 유사한 지식을 제공하는 벡터 공간모델 방안이 있을 수 있다.

3.1 자동 분류기반 방식

자동 분류기반 방식은 사용자가 생성하였거나 조회한 문서가 이미 가지고 있는 분류정보를 통해 사용자가 선호하는 분야를 자동으로 파악하여, 해당 분야에 속하는 지식을 추천하는 방안이다. 기존의 분류정보를 이용하는 방안은 사용자에게 미리 자신의 선호분야를 선택하도록 하고 해당분야에 분류된 문서를 추천하는 방식이었다. 이 경우, 사용자가 자신의 관심 분야를 입력하여야 하며 관심분야를 수정하지 않는 한 같은 분야의 문서가 추천된다. 그러나 사용자의 관심 분야가 고정되어 있지 않고 시간의 흐름에 따라 변화할 때는 사용자의 관심분야 변화를 반영하지 못할 수 있다. 자동 분류기반 방식은 이와 같은 단점을 해결하고자 사용자가 직접 관심분야

를 입력하지 않고 사용자가 관심을 보인 문서들이 가장 많이 속한 분야를 사용자의 선호분야로 파악한다. 선호분야가 파악된 후에는 기존의 분류정보를 이용하는 방안과 동일하게 해당 분야에 등록된 최신 지식을 사용자에게 추천한다.

본 연구에서 제시하는 지식포탈에서의 자동분류기반 방식을 이용한 지식추천 방식은 <그림 1>과 같다. 사용자가 게시물이나 문서 같은 지식을 생성하면 지식 저장과 함께 자동분류를 위한 사전 작업을 수행한다. 사전 작업은 형태소 분석기를 통해 단어의 어근과 어미를 분리하고, 접속어, 조사 등을 제외한 후에 추출된 단어와 빈도수를 가지고 색인을 생성하는 것이다. 사전 작업을 통해 변환된 문서를 가지고 자동 분류기를 통해 이 문서가 어떠한 분야에 얼마의 확률로 속하게 되는지 파악하게 된다. 자동 분류기는 분류기를 생성하기 위하여 미리 학습되어 있어야 한다. 사용자가 생성한 문서에 할당된 분류정보를 가지고 사용자 프로파일을 갱신하여야 하는데, 이는 사용자가 기존에 생성한 문서들과 새로 생성한 문서를 가지고 새로운 선호분야를



<그림 1> 지식포탈에서의 자동분류기반 지식추천 방식

선정하는 과정이다. 사용자 프로파일 갱신을 통해 사용자가 선호하는 분야를 선정한 후에, 선호 분야에 새로운 지식이 등록되면 등록된 신규지식을 사용자에게 추천하는 것으로 지식추천이 이루어진다.

자동분류기반 방식은 사용자가 지식포털에서 활용한 문서에 관한 정보의 일부인인 분류정보만을 사용하여 지식을 추천한다는 측면에서 문서와 사용자에 관련된 많은 유용한 정보를 간과하고 있다는 단점이 있다. 또한 분류체계의 구성에 따라 추천의 정확도와 시스템 성능에 큰 영향을 미칠 수 있다. 예를 들어 분류체계가 광범위하게 구성된 경우에는 사용자가 명시적으로 선호분야를 표시하는 방식과 동일하게 지식 과부하를 경감시키기 어렵다.

3.2 벡터 공간 모델 방안

3.2.1 벡터 공간 모델 개요

벡터공간(vector space) 모델은 사용자가 생성하였거나 조회한 지식에 속한 키워드들을 가지고 벡터를 생성한다. 생성된 벡터를 사용자의 프로파일로 활용하며, 사용자 프로파일과 지식을 비교하여 가장 유사한 지식을 추천한다[Baeza-Yates & Ribeiro-Neto, 1999 ; Hijikata, 2004 ; McCallum & Andrew, 1996]. 벡터공간모델에서는 문서를 벡터공간 내의 하나의 벡터로 취급한다. 즉 문서 i 는 $D_i = (w_{i1}, \dots, w_{in})$ 로 표현할 수 있다. w_{ni} 은 문서 i 에서 n 번째 키워드의 출현 빈도비율이다. 벡터공간 모델에서 두 문서간의 유사도는 식 (1)과 같이, 두 벡터 사이 각의 코사인 값을 통해 계산된다. 즉 두 벡터가 일치할 경우 벡터가 이루는 각은 0이 되어 코사인 값은 1이 되고, 두 벡터가 수직으로 서로 완전히 무관할 경우 코사인 값이 0이 된다.

$$\text{sim}(D_i, D_j) = \frac{\vec{d}_i \cdot \vec{d}_j}{|\vec{d}_i| \times |\vec{d}_j|} = \frac{\sum_{n=1}^m w_{nd_i} \times w_{nd_j}}{\sqrt{\sum_{n=1}^m w_{nd_i}^2} \times \sqrt{\sum_{n=1}^m w_{nd_j}^2}} \quad (1)$$

3.2.2 사용자 프로파일 구성

벡터공간모델에서 사용자 프로파일의 작성과정은 사용자의 흥미와 관심분야를 시스템이 활용할 수 있는 형태로 전환하는 과정이다. 본 연구에서는 벡터 공간 모델에 기반한 평균 문서 벡터, 로키오 알고리즘을 활용하여 사용자가 열람한 문서에서 추출된 벡터를 통해 사용자 프로파일을 구성한다.

(1) 평균 문서 벡터(average document vector)

평균 문서 벡터 방안은 사용자가 관심 있어 한 문서에 포함된 각 단어의 출현빈도를 합한 후 관심 있는 문서 전체의 수로 나눈 평균치를 사용자 프로파일로 사용한다. 사용자 프로파일 P 는 개별 단어들의 평균 출현빈도의 집합으로 이루어지며, 개별 단어들의 평균 출현빈도 P_i 는 다음과 같이 계산된다.

$$P_i = \sum_{j=1}^m \frac{f_{ij}}{m} \quad (2)$$

f_{ij} 는 i 번째 단어가 j 문서에서 나타나는 빈도의 비율이며, m 은 관심 있는 문서 전체 수를 나타낸다.

(2) 로키오 알고리즘(Rocchio algorithm)

로키오 알고리즘은 사용자가 표현한 선호 여부를 모두 고려하여 사용자 프로파일을 구축하는 방식이다[Aas, 1997]. 로키오 알고리즘에 근거한 사용자 프로파일 P 는 다음과 같이 계산된다.

$$P = \alpha\mu_1 - \beta\mu_2 \quad (3)$$

여기서 μ_1 은 선호하는 문서들의 평균벡터, μ_2 는 선호하지 않는 문서의 평균 벡터를 나타낸다. α , β 는 선호 문서와 비선호 문서에 대한 가중치이다. 위와 같은 방안으로 초기에 사용자 프로파일을 구축한 후, 사용자 프로파일 갱신 시에는 기존의 선호도 정보와 새로운 문서의 선호여부를 활용하여 다음과 같이 계산된다.

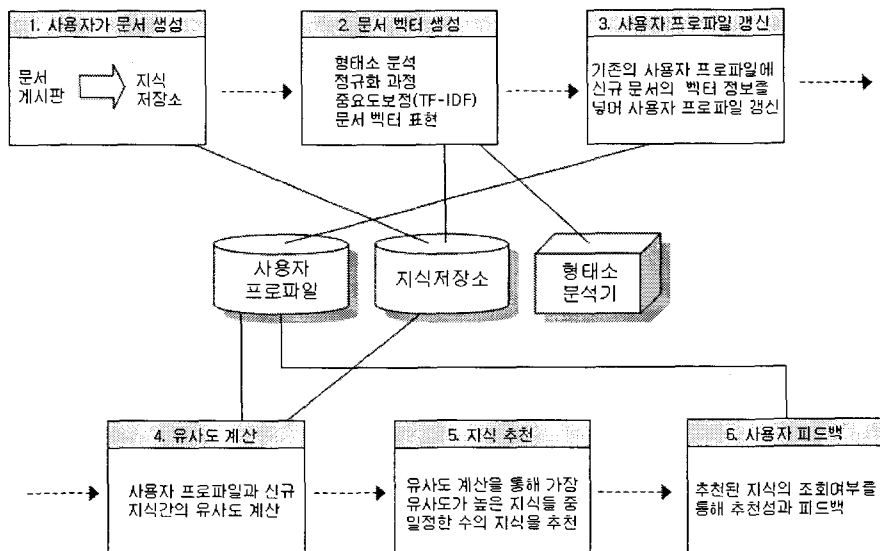
$$P^{new} = yP^{old} + \alpha\mu_1 - \beta\mu_2 \quad (4)$$

여기서 P^{new} 는 새로운 사용자 프로파일이며, P^{old} 는 갱신되기 이전의 사용자 프로파일이며, y 는 기존 프로파일에 대한 가중치이다.

3.2.3 벡터 공간 모델을 이용한 추천방식

본 연구에서 제시하는 벡터공간모델을 이용한 지식추천방식은 <그림 2>와 같다. 지식을 추천 시스템이 활용할 수 있는 형태로 전환하는 처음 과정은, 각 문서에서 단어와 그 빈도수를

추출하여 이를 벡터화하는 것이다. 형태소 분석기를 통해 단어의 어근과 어미를 분리하고, 접속어, 조사 등을 제외하는 선작업 과정을 거친 후에, 문서의 길이가 추출된 단어의 출현 빈도수에 미치는 영향을 없애기 위해 문서의 길이에 따라 빈도수를 조정하는 정규화 과정을 거친다. 정규화 과정을 마친 문서는 단어의 상대적인 중요도를 가리는 과정을 거치게 되는데, 즉 하나의 문서에 단어가 얼마나 자주 나오는지(TF)와 전체 문서 집합에서 그 단어가 출현하는 정도에 대한 역(IDF)을 곱하여 계산한다[Aas, 1997; Lee et al., 1997]. 이 과정에서 빈도가 일정한 임계치에 미치지 못하는 단어는 전체 추천에 미치는 정도가 미미하다고 간주하여 제외한다. 수식 (1)을 사용하여 생성된 문서들의 벡터와 사용자 프로파일과의 유사도 계산을 실시하며, 유사도가 높은 순서로 문서를 정렬하여 상위에 있는 문서들을 추천한다. 추천된 문서에 대해 사용자는 자신의 선호 여부를 표현하여 사용자 프로파일을 갱신하는 피드백 과정을 거쳐 다음 추천에 활용한다.



<그림 2> 지식포탈에서의 벡터공간모델 기반 지식추천 방식

4. 개인화 지식제공 방안 비교 실험

3장에서 제시된 개인화 지식제공 방안들의 비교 실험을 통해 각 방안들의 추천 성과에 대해 분석하고자 한다. 비교 실험은 각 방식을 통해 신문기사를 추천하며, 추천된 신문기사에 대해 사용자들이 선호도를 표시하는 것으로 추천 성과를 파악하고자 한다. 비교 실험에 사용될 개인화 지식제공 모델과 사용자 프로파일 생성 방안을 정리하면 <표 3>과 같다.

분류기반 방식에는 사용자가 명시적으로 선호 분야를 표시한 경우와 시스템이 자동적으로 선호분야를 파악한 경우의 두 가지 추천방식을 활용하였다. 벡터공간모델을 활용한 경우에도 사용자 프로파일 생성방안에 따라 평균 문서 벡터방안과 로키오 알고리즘으로 나뉘었으며, 지식추천의 성과를 비교하기 위하여 무작위 추천을 행하는 경우에 대해서도 비교 실험을 수행하였다.

4.1 실험 환경 및 과정

실험은 웹 기반으로 실시되었으며, 사용자의 선호도를 표시할 수 있는 인터페이스 및 환경은 JSP(Java Server Page)에 의해 작성되었고 각각의 추천방식의 알고리즘은 Java로 구현되었다.

(1) 자료의 수집

실험 자료는 특정 신문 사이트의 기사 중 정치, 경제, 사회, 국제, 문화생활, 정보과학, 스포츠 분야에 속하는, 실험실시 일주일전의 기사를 가져와서 기사 ID, 제목, 내용, 분야에 따라 DB에 저장하였다. 전체 기사의 수는 각 분야별 60개씩 총 420개의 기사를 수집하였다. 기사의 활용은 아래의 표와 같이 이루어졌다.

(2) 실험 과정

실험은 다음과 같은 순서로 이루어졌다.

<표 3> 비교 실험을 위한 지식추천 모델

추천방식	사용자 프로파일 구성방식	구분	설명
분류기반방식	사용자가 관심분야 직접선택	CU	최초에 사용자가 자신의 관심분야를 직접 입력하고, 입력된 분야의 기사를 추천
	자동으로 사용자 관심분야 선택	CA	사용자가 선호한다고 선택한 기사들의 분야 중 가장 많은 분야를 선호분야로 선택하여, 해당분야의 기사를 추천
벡터공간모델	평균문서벡터	VA	사용자 프로파일 생성방안으로 평균 문서 벡터 방안을 활용한 벡터공간모델 기반 지식추천 방식
	로키오 알고리즘	VR	사용자 프로파일 생성방안으로 로키오 알고리즘을 활용한 벡터공간모델 기반 지식추천 방식
무작위추천		RA	전체 기사 중 무작위로 추출하여 추천

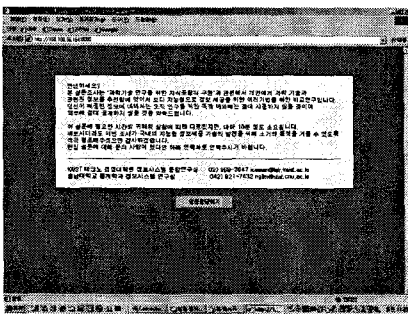
<표 4> 실험 자료의 구성

Group	비율(개수)	설명
초기 사용자 프로파일 생성 대상	1/3(140개)	최초에 주어지는 기사에 대해 사용자가 선호/비선호를 표시하는데 사용
추천 대상	2/3(280개)	추천을 위한 데이터 집합

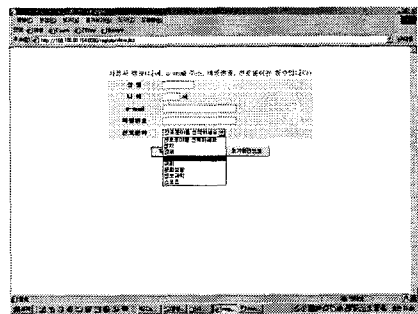
- ① 사용자가 실험사이트에 가입하여, 자신의 선호분야를 표시한다.
- ② 최초 사용자 프로파일 생성을 위해 <표 4>의 '초기 사용자 프로파일 생성대상' 그룹 140개의 기사에서 각 분야별로 3개씩 총 21개의 기사를 사용자에게 보여준다.
- ③ 사용자가 제시된 신문기사에 대해 자신의 선

호/비선호 여부를 체크하면, 그 결과에 의하여 3장에서 제시한 방법에 의해 사용자의 선호분야를 자동 선택하고, 또한 평균 문서 벡터, 로키오 알고리즘 방식의 2가지 사용자 프로파일이 생성된다.

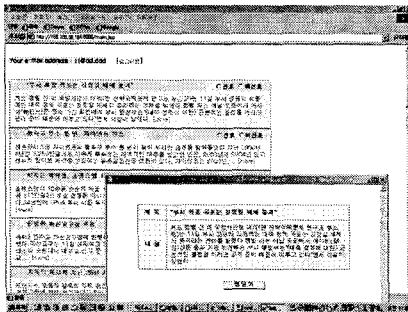
- ④ <표 4>의 추천 대상 280개 기사 중 각각의 사용자 프로파일과의 유사도 계산에 의해서



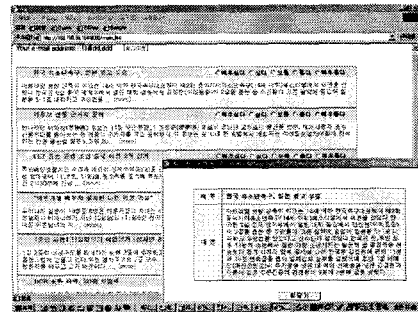
(1) 실험 사이트 첫 화면



(2) 개인 정보 및 선호분야 입력



(3) 초기 사용자 프로파일 생성 대상에 대한 선호여부 표시



(4) 추천 대상에 대한 선호도 표시



(5) 실험 사이트 종료 화면

<그림 3> 실험화면

상위 4개의 기사를 추천한다. 또한 CA, CU 방식에 의해 4개, RA에 의해 4개가 추천되어 총 20개의 기사가 추천된다.

- ⑤ 추천된 기사에 대해 사용자는 5점 척도(매우 선호, 선호, 보통, 비선호, 매우 비선호)에 의해 자신의 선호여부를 표시하게 된다.
- ⑥ 자신의 선호여부를 체크한 후에 설문은 종료된다.

실험의 샘플 수는 총 207명이었으나 누락 등 무성의하게 작성한 설문응답자를 제거하여 총 173명에 대해서만 분석하였다.

4.2 실험결과

(1) 설문 응답자의 특성

실험에 참여한 응답자들은 주로 대학생과 대학원생들이며, 이들이 설문사이트를 통해 선호한다고 명시적으로 표현한 관심분야는 <표 5>와 같다.

<표 5> 분야 별 선호도

선호분야	빈도수(명)	%
문화생활	47	27.2
경제	37	21.4
스포츠	36	20.8
정보과학	27	15.6
정치	12	6.9
사회	8	4.6
국제	6	3.5

(2) 각각의 개인화 방안에 대한 만족도 비교

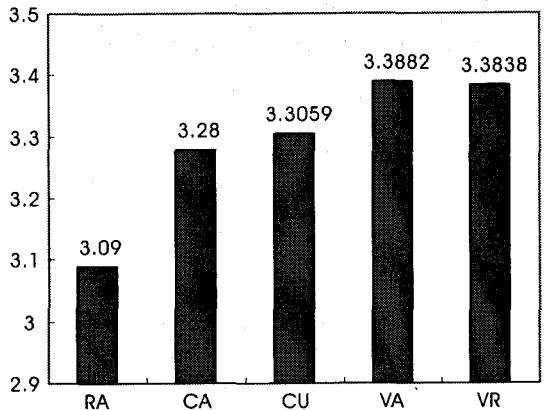
각 개인화 방안에 의해 추천된 기사에 대한 설문자의 평균 만족도를 <그림 4>에 표시하였다.

가장 높은 만족도를 보이는 방식은 벡터공간 모델 중에 평균문서벡터 방안을 활용한 추천방

식(VA)에 의한 것이며, 평균 만족도는 3.3882이다. 개별 추천 방식의 성과에 대한 통계적인 비교를 위해서 사용한 가설은 다음과 같다.

$$H_o : \mu_{CA} = \mu_{CU} = \mu_{VA} = \mu_{VR} = \mu_{RA} \quad (5)$$

이를 검정하기 위해서 반복측정 분산분석(repeated measures ANOVA) 방법을 사용하였다. 그 결과 신뢰수준 95%에서 p-value의 값이 0.001로써 귀무가설을 기각하게 되어 각각의 추천 방식 간에는 차이가 있다고 말할 수 있다. 개별 방식의 추천 성과차이를 검정해본 결과는 <표 6>에 표시되었다.



<표 6> 추천 방식의 만족도 비교 검정

귀무가설	p-value	검정결과
$H_o : \mu_{CA} = \mu_{CU} = \mu_{VA} = \mu_{VR} = \mu_{RA}$	0.001	H_o 기각
$H_o : \mu_{CA} = \mu_{CU} = \mu_{VA} = \mu_{VR}$	0.643	H_o 채택
$H_o : \mu_{CA} = \mu_{CU}$	0.757	H_o 채택
$H_o : \mu_{VA} = \mu_{VR}$	0.951	H_o 채택

분류기반방식이나 벡터공간모델의 추천 성과가 무작위 추천방안의 추천 성과보다 통계적으로 유의하게 좋은 것으로 볼 수 있다. 분류기반

방식에 의한 추천과 벡터공간모형을 활용한 추천방식 중에는 벡터공간모형을 활용하는 방안의 만족도가 좀 더 높았으나 통계적으로 유의한 수준은 아니었다. <표 6>의 세 번째 줄의 결과는 사용자가 명확하게 자신의 선호분야를 직접 밝히는 경우와 시스템에 의해 자동으로 선호분야를 파악하여 추천하는 경우에 대한 만족도가 통계적으로 동일한 것으로 볼 수 있다. 벡터공간모형을 활용하는 방안에서 평균문서벡터나 로키오 알고리즘을 활용하는 경우에 대한 만족도도 통계적으로 동일하게 파악되었다.

5. 토의 및 결론

본 연구에서는 지식포탈에서 활용 가능한 개인화 지식제공방안을 설계하고, 이들의 추천 성과를 실험을 통해 비교하였다. 지식포탈의 지식분류체계를 활용하는 분류기반 방식과 문서에 포함된 키워드를 활용하는 벡터공간모형을 제시하였다. 추천 성과비교 실험에서 무작위로 사용자에게 제공된 정보들보다 분류기반방식이나 벡터공간모형에 기반하여 선정된 정보들이 더 높은 추천 성과를 보였다. 지식포탈의 활용에 의해서 자동적으로 파악된 선호분야에 기반을 둔 추천방식이나 사용자가 선호하는 문서에 속한 단어들에 기반을 둔 방안의 추천 성과가 사용자가 명시적으로 선호분야를 표시한 경우와 유사한 것으로 파악되었다. 이러한 자동화된 추천 방법의 사용을 통해, 사용자가 선호분야를 직접 표시하고 변경하는 노력을 경감할 수 있으며, 지식포탈에서의 사용자 활동에 기반하여 자동으로 사용자 프로파일을 생성하여 지식추천을 행할 수 있다. 위와 같은 개인화 지식제공방안을 통해 지식포탈에서의 정보 과부하를 줄일 수 있으며, 자동분류기반 방식과 벡터공간모형 방식을 통해 분류체계 변경에 대한 시스템의 유연성 및 적용

성을 높일 수 있다.

추후 연구 과제로는 사용자로부터 추천 성과에 대한 피드백을 받아 지속적으로 사용자 프로파일을 갱신하여 교정해 나가기 위한 피드백 방안에 대한 연구가 필요하다. 또한 지식이 지속적으로 축적되면 비교하여야 하는 대상 문서의 수나 키워드 수가 증가하게 되므로, 추천시스템의 추천 성과를 떨어뜨리지 않으면서도 계산 복잡도를 줄일 수 있는 차원 축소(Dimension Reduction) 방안에 대한 연구도 필요하다.

참고 문헌

- [1] 김재경, 조윤호, 안도현, “고객 선호도 변화를 반영한 상품추천방법론”, *경영정보학회 춘계 학술대회 논문집*, 2003, pp. 620-625.
- [2] 박상규, 이재성, 신승은, 강유환, 오효정, 장명길, 서영훈, “문서필터링을 위한 질의어 확장과 가중치 부여기법”, *정보처리학회논문지B*, 제10권 제7호, 2003, pp. 743-751.
- [3] 박성주, 이홍주, 김종우, 김규중, 안형준, “과학기술 연구개발조직의 팀 연구 지원을 위한 지식포털 모델”, *Information Systems Review*, 제5권 제2호, 2003, pp. 151-168.
- [4] 이홍주, 유기현, 김종우, 박성주, “연구개발 조직의 지식경영시스템 기능에 대한 인지적 중요도에 관한 연구: 정부출연 연구소를 중심으로”, *경영정보학연구*, 제13권 제3호, 2003, pp. 243-259.
- [5] Aas, K., “A survey on personalized information filtering systems for the World wide web”, Norwegian Computing Center, 1997.
- [6] Baeza-Yates, R. and Ribeiro-Neto, B., *Modern information retrieval*, Addison Wesley, 1999.
- [7] Belkin, N.J. and Croft, W., “Information filter-

- ing and Information Retrieval : Two sides of same coin?", *Communications of the ACM*, Vol. 35, No. 12, 1992, pp. 29-38.
- [9] Davydov, M., *Corporate Portals and e-Business Integration*, McGraw-Hill, 2001.
- [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] 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.
- [12] Foltz, P.W. and Dumais, S.T., "Personalized Information Delivery : An Analysis of Information Filtering Methods", *Communication of the ACM*, Vol. 35, No. 12, 1992, pp. 51-59.
- [13] Hijikata, Y., "Implicit User Profiling for On Demand Relevance Feedback", *Proceedings of the 9th international conference on Intelligent user interface*, 2004, pp. 198-205.
- [14] Lee, D.L., Huei Chuang, and Kent Seamons, "Document Ranking and the Vector-Space Model", *IEEE Software*, Vol. 14, No. 2, 1997, pp. 67-75.
- [15] Mostafa, J., Mukhopadhyay, S., Lam, W., and Palakal, M., "A Multilevel Approach to Intelligent Information Filtering : Model, System, and Evaluation", *ACM Transactions on Information Systems*, Vol. 15, No. 4, 1997, pp. 368-399.
- [16] Pazzani, M. and Billsus, D., "Learning and Revising User Profiles : The Identification of Interesting Web Sites", *Machine Learning*, Vol. 27, 1997, pp. 313-331.
- [17] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J., "GroupLens : An Open Architecture for Collaborative Filtering of Netnews", *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 175-186.
- [18] Stabb, S. and Maedche, A., "Knowledge Portals : Ontologies at Work", *American Association for Artificial Intelligence*, Summer, 2001, pp. 63-75.
- [19] Turban, E., King, D., Lee, J. K., and Viehland, D., *Electronic Commerce 2004 : A Managerial Perspective*, 3rd ed., Prentice Hall, 2004.
- [20] Yan, T.W. and Garia-Molina, H., "SIFT-A Tool for Wide-Area Information Dissemination", *Proceedings of 1995 USENIX Technical Conference*, 1995, pp. 177-186.
- [21] Zampa, V. and Lemaire, B., "Latent Semantic Analysis for User Modeling", *Journal of Intelligent Information Systems*, Vol. 18, 2002, pp. 15-30.
- [22] ACUBE, ACUBE Enterprise Knowledge Portal, <http://www.sds.samsung.co.kr/main.html>
- [23] BizFlow, BizFlow EIP, http://corona.handysoft.co.kr/product/pdt3_0102.asp?menu_type=product&m=2
- [24] Hummingbird, Hummingbird Portal, <http://www.hummingbird.com/products/portal/index.html>
- [25] McCallum and Andrew, K., "Bow : A Toolkit for Statistical Language Modeling, Text Retrieval, Classification and Clustering", <http://www.cs.cmu.edu/~mccallum/bow>, 1996.
- [26] Plumtree, Plumtree Corporate Portal, <http://www.plumtree.com/products/platform/>

■ 저자소개



이 홍 주

저자는 한국과학기술원(KAIST) 산업경영학과를 졸업(1997)하고 KAIST 테크노경영대학원 경영공학과정에서 석사(1999)학위를 취득하였다. 현재 KAIST 테크노경영대학원 경영공학과정에서 박사과정에 재학 중이며 주요 관심분야는 개인화, CRM, 모바일 커머스, 가상 협업시스템 등이다.



김 종 우

저자는 서울대 수학과에서 이학사(1989), 한국과학기술원 경영과학과에서 공학석사(1991)를 취득하고 한국과학기술원 산업경영학과에서 공학박사(1995)를 취득하였다. 현재 한양대학교 경영학부 부교수로 재직 중이며, 주요 관심분야는 경영정보시스템, 의사결정지원시스템, 전자상거래, 데이터 마이닝 응용, 지식경영시스템 등이다.



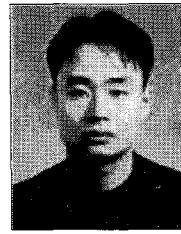
김 광 래

저자는 충남대학교 통계학과를 졸업(2003)하고 충남대학교 통계학과에서 석사(2005)학위를 취득하였다. 현재 충남대학교 통계학과에서 박사과정에 재학 중이며 주요 관심분야는 일반화 선형 모형, 통계 데이터베이스 시스템 구축 및 활용 등이다.



안 형 준

저자는 KAIST 테크노경영대학원 경영공학과정에서 박사(2004)학위를 취득하였다. 현재 University of Waikato에서 Waikato Management School에 재직중이며 주요 관심분야는 공급사슬관리, 지식경영시스템, 다중에이전트 기반 시스템 등이다.



권 철 현

저자는 서울대학교 재료공학부를 졸업(2000)하고, 과학기술원(KAIST)에서 경영공학석사학위를 취득(2002)하였으며, 동대학원에서 경영공학박사과정을 수료(2004)하였다. 현재 삼성SDS에 근무 중이며, 주요관심분야는 에이전트 시스템, 공급망관리, 공급망관계관리 등이다.



박 성 주

저자는 서울대학교 산업공학과를 졸업(1973)하고 한국과학기술원(KAIST)에서 산업공학석사(1975) 학위를 취득하였으며 Michigan State University에서 시스템공학박사(1978)학위를 취득하였다. 현재 한국과학기술원(KAIST) 테크노경영대학원 원장으로 재직 중이며 주요관심분야는 에이전트 시스템, 워크플로우 시스템, 가상 협업시스템 등이다.