

# 개인화 검색시스템에 관한 연구

- 과학기술학회마을을 중심으로 -

## A Study of Personalized Retrieval System through Society of Korean Journal Articles of Science and Technology

김 광 영(Kwang-Young Kim)\*

곽 승 진(Seung-Jin Kwak)\*\*

### < 목 차 >

- |                        |                         |
|------------------------|-------------------------|
| I. 서론                  | 3. 저자 키워드 기반 주제어 추천 시스템 |
| II. 관련연구               | 4 유사한 사용자 추천 시스템        |
| III. 과학기술학회마을 논문 검색서비스 | V. 평가 및 결과              |
| IV. 개인화 검색시스템          | VI. 결론 및 제언             |
| 1. 질의어 기반의 개인 성향 분석    |                         |
| 2. 공동 저자 군집화 내비게이션     |                         |

### 초 록

이 연구에서는 기존의 과학기술학회마을에서 제공하고 있는 일반 검색서비스에 대해서 분석하고 이를 기반으로 과학기술학회마을 논문 서비스에 적합한 개인화 검색서비스들을 개발하였다. 즉, 질의어 기반의 개인화 검색시스템, 논문의 공동 저자 기반의 내비게이션 시스템, 논문의 저자 키워드 기반 주제어 자동 추천 시스템과 유사한 사용자 자동 추천 시스템들이 있다. 이 연구에서는 설문 조사를 통해서 KISTI 과학기술학회마을 논문 서비스에 적합한 개인화 서비스를 모색하였다.

키워드: 개인화, 검색시스템, 키워드 주제 정보, 저자 군집화 시스템

### ABSTRACT

In this research, we analyze about the general service provided by Society of Korean journal articles of science and technology. Personalized retrieval services which are suitable to the articles service were developed based on this. That is, there are personalized retrieval system based on user's keyword, authors navigation system, automatic topic recommendation system based on author's keyword, and similar user automatic recommendation system. In this research, personalized service methods being suitable to the articles service of Society tries to be considered through the user survey.

Keywords: Personalized Retrieval System, Keyword Topic Information, Author Clustering System

\* 한국과학기술정보연구원 정보기술연구실 선임연구원(kykim@kisti.re.kr) (제1저자)

\*\* 충남대학교 사회과학대학 문헌정보학과 조교수(sjkwak@cnu.ac.kr) (교신저자)

• 접수일: 2010년 2월 26일 • 최초심사일: 2010년 3월 8일 • 최종심사일: 2010년 3월 26일

## I. 서론

대부분의 검색시스템은 사용자의 검색 의도를 반영하지 않고 문서 검색을 하고 있다. 사용자에게 적합한 다양한 서비스들을 제공하기 위해서는 우선 사용자의 성향을 분석할 필요가 있다. 그렇지만 사용자의 주요 관심 정보를 파악하는 것은 쉽지 않다.

사용자의 성향 정보를 분석하기 위해서 일반적으로 사용자 프로파일을 만들어 사용한다. 사용자 프로파일을 생성하는 방법에는 사용자가 직접적으로 개인 정보나 관심 정보를 이용하는 명시적(explicit)인 방법과 사용자의 행동을 통해 암시적(implicit)으로 개인 성향 정보를 추론하는 방법으로 나눌 수가 있다. 사용자의 명시적인 표현을 통해서 프로파일을 이용할 경우에는 빠르고 정확하게 사용자의 성향을 분석할 수 있지만 대부분의 사용자들은 이것을 번거롭게 생각하며 동적인 취향 변화를 반영하기 어렵다는 단점이 있다. 그래서 현재 몇몇의 개인화 검색시스템들은 암시적인 방법을 이용하여 사용자의 관심 정보를 자동으로 분석하여 제공하려고 시도하고 있다.

현재 일반적인 검색 엔진에서는 서로 다른 사람이 똑 같은 질의어를 보낼 경우에 똑 같은 결과를 제시한다. 이것은 정보 요구가 서로 다른 사람들에게는 적합하지 않다. 예를 들면 “바이러스”라고 입력을 했을 때, 어떤 사람은 생물/의학 개념에서 바이러스에 대한 정보를 찾을 것이고 다른 사람은 컴퓨터에 관련된 바이러스에 대한 정보를 찾기를 원할 수도 있다.

위와 같은 문제는 언어의 중의성 문제에서 발생하는 것들이다. 이러한 언어의 중의성 문제도 개인화 검색시스템에서 개인의 성향 정보를 분석하여 개인에게 적합한 문서를 제공함으로써 해결 할 수가 있었다.<sup>1)</sup>

또한 개인화 검색시스템은 전자상거래 분야에서 상거래를 활성화시키고 정보 제공자(웹 상점)에 대한 만족도와 충성도를 높이기 위해, 사용자의 취향을 기반으로 한 정보(상품) 추천은 필수적으로 요구되고 있다. 전통적인 추천 기법으로, 사용자의 관심과 콘텐츠의 유사도를 비교하는 콘텐츠 기반 추천(Content-based Recommendation)과 유사한 관심을 가진 사용자들 간에 관심 콘텐츠를 상호 추천하는 협업추천(Collaborative Recommendation)이 있다. 이러한 추천 시스템들은 모두 사용자의 취향을 기술한 프로파일을 기초로 한다.<sup>2)</sup>

현재의 Google, Yahoo, Naver 및 전자 상거래의 사이트들은 다양한 개인화 서비스를 제공하고 있다. Google의 iGoogle은 웹에서 찾은 뉴스, 사진, 날씨 등의 콘텐츠를 자신의 페이지에 추가할 수 있다. 즉 사용자 마음대로 꾸밀 수 있는 구글 가젯 서비스를 제공한다. 네이버는 N디스크, 메일, 포토앨범, 가계부, 통합계좌 등의 PIMS(Personal Information Management System)부터 개인

1) 김광영, 심강섭, 박승진, “분류와 사용자 질의어 정보에 기반한 개인화 검색시스템,” 한국문헌정보학회지, 제43권, 제3호(2009. 9), pp.163-180.

2) 이승화, 최형기, 이은석, “사용자 기기에서 이용한 웹 데이터 분석을 통한 사용자 취향 분석 방법,” 정보과학회논문지, 제15권, 제3호(2009. 3), pp.189-199.

화 서비스를 제공하고 있다.

야후에서는 마이야후라는 개인화 서비스를 제공하고 있으며 개인 사용자에게 다양한 콘텐츠를 제공하고 플리커나 야후 캘린더 등의 다른 야후 서비스와 통합 및 다양한 레이아웃 옵션 등이 추가되었다. 방대한 정보에서 사용자의 요구에 맞는 정보 필터링과 개인화 서비스가 매우 중요해지고 있다.

본 연구의 목적은 기존의 과학기술학회마을<sup>3)</sup>에 적합한 개인화 검색시스템 개발과 서비스 방안을 연구하는 것이다. 이를 위해서 과학기술학회마을의 검색 서비스를 분석하고 이것을 이용하여 논문 검색 서비스에 적합한 개인화 서비스를 도출하고 실제 시스템을 개발한다. 그리고 실제 개발된 개인화 검색시스템과 다양한 서비스를 내부 전문가를 중심으로 평가를 하고 평가된 결과를 바탕으로 과학기술학회마을에 적합한 서비스와 향후 연구되어야 할 방안들을 모색하고자 한다.

## II. 관련연구

일반적으로 개인화(Personalization)라는 용어는 이용자 정보요구에 부합되는 콘텐츠를 제공하는 의미로 광범위하게 사용된다.<sup>4)</sup> 개인화 검색의 유형은 개인이 적극적으로 프로파일의 사항을 입력하면, 이를 이용하여 기본 프로파일을 작성하는 방법과 서비스 이용형태를 기반으로 하는 방식으로 클릭한 문서, 클릭 수 등의 정보를 이용하여 프로파일을 작성하는 방법과 이용자의 사회화 프로파일(Social Profile)을 기반으로 하는 방식이 있다.<sup>5)</sup>

대부분의 개인화 검색시스템에 관한 주요한 기능들은 개인의 클릭 정보나 개인의 검색 히스토리 정보를 이용하여 개인화 검색에 많이 활용을 하고 있다.

개인의 어떤 내용을 개인화하여 검색 결과에 반영하는지에 따라 링크정보를 이용하거나, 질의어를 확장하거나, 결과를 재순위화하거나, 메타검색, 혹은 도메인별 검색 등이 있다. Jeh와 Widom (2003)은 이용자가 즐겨 찾는 페이지에서 링크되거나, 해당 페이지가 링크한 페이지에 더 많은 가중치를 두어 검색랭킹에 반영하는 형태를 연구하였다.<sup>6)</sup> 또한 개인화된 메타검색 엔진들도 개발되었는데, Inquirus2는 이용자가 원하는 주제 분야를 선정하고, 검색엔진을 선택할 때 해당 주제의 내용을 이용하였다.<sup>7)</sup>

3) 과학기술학회마을 홈페이지, <<http://society.kisti.re.kr>> [인용 2010. 2. 17].

4) Shahabi, Cyrus and Yi-Shin Chen, "Web Information Personalization : Challenges and Approaches," *Proceedings of 3rd Workshop on Databases in Networked Information System*, (2003), pp.5-15.

5) 이소영, 정영미, "웹 포털 이용자 로그 데이터에 기반한 개인화 검색서비스 모형의 설계 및 평가," *정보관리학회지*, 제23권, 제4호(2006. 12), pp.179-196.

6) Jeh, G. and J. Widom, "Scaling Personalized Web Search," *Proceedings of the 12th International World Wide Web Conference*, (2003), pp.271-279.

그 흐름을 보면 가장 간단한 유형으로 이용자의 속성 정보에 근거한 개인화이다.<sup>8)</sup> 이용자가 입력한 프로파일에 기반을 두어 맞춤형 검색을 제공하였고 그 다음으로는 이용자의 행동 정보에 근거하는 개인화로서 검색 이력(history) 위주의 다양한 기능을 제공하는 개인화 검색시스템이다. 2004년 10월 야후가 'My Search'를 시범 서비스로 시작하였으며 검색 결과를 저장하고 편집할 수 있는 기능 위주로 만들어진 시스템이다. 그 다음으로는 이용자의 자산 데이터를 대상으로 하는 개인화이다. 단순하게 개인 컴퓨터의 데이터를 효율적으로 찾아주는 기존의 소규모 솔루션의 데스크 탑 검색시스템이다. 그리고 2003년부터 폭발적인 인기를 얻고 있는 사회 연결망 개념을 도입한 서비스이다. 기본적인 'My Search'의 개념 모형에 자신의 정보를 다른 사람과 공유하는 것이다. 이 모델의 문제점은 활발한 공유가 가능한 네트워크를 구축하기가 너무 어렵다는 것이다.<sup>9)</sup>

개인화의 중심점을 사용자의 질의어 패턴 분석을 통하여 개인의 성향 정보를 분석한 시스템들도 있다. 즉 시스템은 질의어 랭킹 정보(Query Ranking Information)를 참조하여 웹 사용자의 주요 관심사를 파악한다.<sup>10)</sup>

대형 웹 포털이나 전문 검색시스템에서 이러한 개인화 검색서비스를 제공하기에 현실적으로 많은 어려운 점들을 가지고 있다. 현재 대형 웹 포털 사이트들도 이러한 문제점들을 극복하기 위해서 개인화된 다양한 서비스들을 제공하고 있다.

이와 같이 사용자의 관심 정보를 분석하기 위한 다양한 방법들이 연구되고 있다. 가장 보편화된 방법은 사용자가 사이트 방문 초기에 명시적으로 표현한 개인 정보나 관심 정보를 이용하는 것이다.<sup>11)</sup>

국외에서도 사용자의 클릭 히스토리를 이용하여 개인화 검색시스템에 반영하여 단어의 중의성 문제 해결을 시도하고 있다. 또한 질의어 - 질의어의 재조합을 통하여 사용자의 관심 분야를 검색하는 방법도 제공되고 있다. 즉 사용자가 "Windows"라는 질의어를 입력할 때 "Windows XP"와 "House windows" 등으로 질의어 - 질의어를 재구성하여 검색 결과를 다양하게 처리하는 방법도 제공되고 있다.<sup>12)</sup> 2007년 Ahu Sirg 등은 온톨로지 기반의 사용자 프로파일 정보와 사용자 행위 정보를 이용하여 사용자에게 적합한 문서를 재순위화 시키는 방법을 이용한 개인화 검색서비스를 제안 하였다.<sup>13)14)</sup>

---

7) Glover, E. J. et al., "Web Search - Your Way," *Communications of the ACM*, Vol.44, No.12(2000), pp.97-102.  
8) Bonnet, Monica. "Personalization of Web Services : Opportunities and Challenges," *Ariadne*, Issue 28, <<http://www.ariadne.ac.uk/issue28/personalization>> [cited 2010. 02. 17].  
9) 이소영, 정영미, 전개논문, pp.179-196.  
10) 박건우, 이상훈, "질의어 패턴 자동분석을 통한 커뮤니티 기반 개인화 검색," *정보과학회논문지*, 제36권, 제4호 (2009. 8), pp.321-326.  
11) G. Linden, B. Smith and J. York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, (January/February 2003), pp.76-80.  
12) Filip Radlinski, Susan Dumais, "Improving Personalized Web Search using Result Diversification," *Proceedings of the 29th annual international ACM SIGIR*, (2006), pp.691-692.  
13) Ahu Sieg, Bamshad Mobasher and Robin Burke, "Web Search Personalization with Ontological User Profiles,"

ODP(Open Directory Project)<sup>15)</sup>는 텍소노미를 이용하여 다양한 분야 및 웹기반 개인화 검색 시스템에 적용한 것도 있다.<sup>16)17)</sup> 구글의 PageRank 알고리즘을 확장한 Personalized Page Rank 기술을 적용한 웹기반 개인화 검색시스템도 있다.<sup>18)</sup> 이와 같이 개인화 검색시스템에 대한 관심도가 점점 높아지고 있고 활발한 연구들이 진행되고 있다. 본 연구에서도 과학기술학회마을의 기존 논문 검색서비스를 개인화 검색시스템에 적용시켜보고자 한다.

### Ⅲ. 과학기술학회마을 논문 검색서비스

본 연구에서는 현재 KISTI에서 서비스하는 과학기술학회마을 논문 검색서비스를 분석하고 이를 기반으로 개인화 검색시스템에 적용 가능한 서비스들을 도출하였다.

과학기술학회마을에서는 제공되고 있는 논문 검색서비스의 기능들은 기본 검색, 고급 검색, 통합 검색을 지원하고 학술지 검색 부분에서는 학술지별, 학회별, 주제별 검색 기능을 제공하고 있다. <표 1>은 과학기술학회마을 검색서비스에 대한 상세 기능들을 설명한 것이다. 일반적으로 대부분의 사용자들은 기본 검색 기능을 사용하고 있다. 즉 제목, 초록, 키워드, 저자 등을 사용자가 선택할 수 있도록 검색 기능을 제공하고 기사 명을 클릭하여 원문 정보를 제공 받을 수가 있다. 또한 전문가를 위해서 고급검색 기능을 제공하고 있다. 고급 검색 기능은 전문 사용자가 예약어 및 연산자를 사용하여 검색 Query를 만들어 검색하는 전문가용 검색서비스를 지원한다. 이와 같이 일반 사용자와 전문가 사용자로 두 그룹으로 나누어 검색서비스를 제공하고 있다. 또한 학술지명, 학회명 검색 지원과 DDC분류법을 이용하여 학술지 분류 방법을 이용한 주제별 학술지 검색을 제공하고 있다.

본 연구에서는 논문의 서지 정보인 주제별 분류 DCC, KDC 등의 분류된 정보를 이용하여 사용자의 성향을 분석하는 시스템을 도출하였다. 즉, 사용자 검색 질의어와 검색된 문서들의 주제별 분류 정보를 이용하여 사용자의 성향 정보를 판단하는 개인화 검색시스템을 개발하였다. 이 개인화

*Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, (2007), pp.525-534.

14) J. Trajkova and S. Gauch, "Improving ontology-based user profiles," *Proceedings of the Recherched' Information Assistear Ordinateur, RIAO*, (2004), pp.380-389.

15) Open Directory Project Home page, <<http://www.dmoz.org>> [cited 2010. 02. 09].

16) P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschutter, "Using ODP metadata to personalize search," *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR*, (2005), pp.178-185.

17) C. Ziegler, K. Simon, and G. Lausen, "Automatic computation of semantic proximity using taxonomic knowledge," *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM*, (2006), pp.465-474.

18) G. Jeh and J. Widom, "Scaling personalized web Search," *Proceedings of the 12th international conference on World Wide Web*, (2003), pp.271-279.

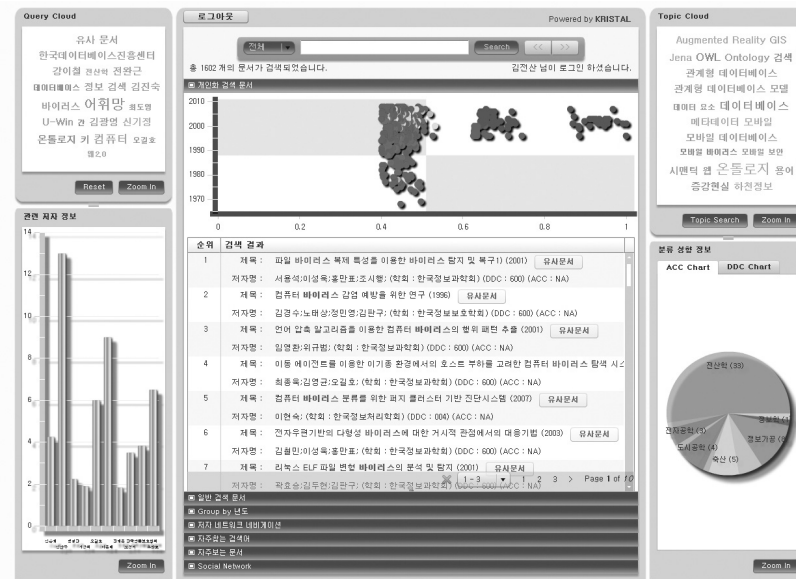
검색시스템을 기반으로 사용자에게 필요한 다양한 서비스를 도출하였다. 즉, 논문의 저자 키워드를 중심으로 사용자의 성향 정보에 맞는 주제 키워드 자동 추천 시스템을 개발하였고 개인 성향에 적합한 문서들을 중심으로 공동 저자 군집화 시스템을 적용하여 관련된 연구자 중심으로 빠르게 탐색할 수 있는 내비게이션 기능을 개발 할 수 있었다. 이와 같이 논문 데이터의 기본 서지 항목 중에서 공동저자 정보, 저자 키워드 등의 구조적인 정보들을 이용하여 개인화 서비스를 위한 다양한 기능들을 도출할 수가 있다. 또한 주제별 분류 기반으로 분석된 사용자의 성향 정보를 이용하여 유사한 사용자 자동 추천 시스템도 개발하였다.

〈표 1〉 과학기술학회마을 검색서비스

	업무	상세 정의
1. 논문검색	1.1 기본검색	<ul style="list-style-type: none"> <li>• 제목, 초록, 키워드, 저자 등을 사용자가 선택할 수 있도록 검색 기능 제공</li> <li>- 기사명 클릭을 통해 상세정보 및 원문정보를 확인 할 수 있는 기능 제공</li> </ul>
	1.2 고급검색	<ul style="list-style-type: none"> <li>• 사용자가 예약어 및 연산자를 사용하여 검색 Query를 만들어 검색하는 전문가 적인 검색 서비스 지원</li> <li>- 결과분류의 Analyze View 기능을 통해, 검색 결과의 Refine Result 기능 제공</li> <li>- 해당 기사가 포함된 저널 정보 제공</li> <li>- 해당 기사의 제목, 저자, 초록, 키워드, 참고문헌 정보 확인 기능 제공</li> <li>- 해당 학술지 내에서의 주요 키워드 및 권호정보 검색 기능 제공</li> </ul>
	1.3 통합검색	<ul style="list-style-type: none"> <li>• 과학기술학회마을이 가지고 있는 학회정보를 통합적으로 검색 기능 제공</li> <li>- 통합 검색 결과로 논문, 학술지, 학회 검색 결과를 보여 줌</li> <li>- 더보기 기능을 통해 논문 목록이 보이며, 논문 제목을 선택하여 해당 논문의 상세 정보 확인 기능 제공</li> </ul>
2. 학술지 검색	2.1 학술지별	<ul style="list-style-type: none"> <li>- 검색결과로 학술지명, ISSN, 수록범위, SCI 등재여부 View 기능 제공</li> <li>- 학술지명 클릭 시 해당 학술지의 권/호 목록 페이지 이동</li> <li>- 권/호 페이지에서 상세정보 연계</li> <li>- 상세 페이지에서 논문 내용 확인 및 원문이미지 뷰 기능 제공</li> <li>- 논문 기사가 보유 한 DOI 값 클릭 시 KoreaScience의 해당 영문화 페이지 연결 기능 제공</li> <li>- 상세페이지에서 학회지명/저자/키워드/해당 권호 재검색 및 이동 기능 제공</li> </ul>
	2.2 학회별	<ul style="list-style-type: none"> <li>- 검색결과로 학회에서 발행하고, 구축되어진 학술지 목록 View</li> <li>- 학술지 명 선택 시 권/호 목록 페이지 뷰</li> <li>- 권/호 페이지에서 상세정보 연계</li> <li>- 상세 페이지에서 논문 내용 확인 및 원문이미지 뷰 기능 제공</li> <li>- 논문 기사가 보유 한 DOI 값 클릭 시 KoreaScience의 해당 영문화 페이지 연결 기능 제공</li> <li>- 상세페이지에서 학회지명/저자/키워드/해당 권호 재검색 및 이동 기능 제공</li> </ul>
	2.3 주제별	<ul style="list-style-type: none"> <li>- 해당 주제 분류 선택 시 주제어에 해당하는 하위 학회지 명, ISSN, 수록범위, 언어, SCI 여부 View</li> <li>- 학회지명 선택 시 학회지가 보유한 권/호 정보 페이지 이동</li> <li>- 권/호 정보 선택 시 권/호 정보가 보유 한 기사 리스트 뷰 기능 제공</li> <li>- 상세 페이지에서 논문 내용 확인 및 원문이미지 뷰 기능 제공</li> <li>- 논문 기사가 보유 한 DOI 값 클릭 시 KoreaScience의 해당 영문화 페이지 연결 기능 제공</li> <li>- 상세페이지에서 학회지명/저자/키워드/해당 권호 재검색 및 이동 기능 제공</li> </ul>
	2.4 통합검색	<ul style="list-style-type: none"> <li>- 과학기술학회마을이 가지고 있는 학회정보를 통합적으로 검색 기능 제공</li> <li>- 통합 검색 결과로 논문, 학술지, 학회 검색 결과를 보여 줌</li> <li>- 더보기 기능을 통해 논문 목록이 보이며, 논문 제목을 선택하여 해당 논문의 상세 정보 확인 기능 제공</li> </ul>

### IV. 개인화 검색시스템

본 논문에서는 개인 성향 분석을 위해 질의어에 기반한 개인화 검색시스템을 사용하였다. 이 개인화 검색시스템을 기반으로 하여 다양한 개인화 서비스를 도출 및 개발하였다. 그 중에서도 대표적인 것들만을 소개하고자 한다. 첫째, 개인 성향 정보를 이용하여 개인에게 적합한 주제 정보를 추천하기 위해서 논문의 키워드 기반 자동 주제 추천 서비스를 개발하였다. 둘째, 개인 성향 정보를 이용하여 논문의 공동 저자를 군집화 하는 저자 내비게이션 기능을 개발하였다. 이것을 이용하여 찾고자 하는 저자를 중심으로 문서를 검색할 수 있도록 하였다. 마지막으로 유사한 사용자의 자주 추천 시스템을 이용하여 유사한 사용자가 자주 보는 관심 주제 정보와 키워드 정보를 제공하였다.



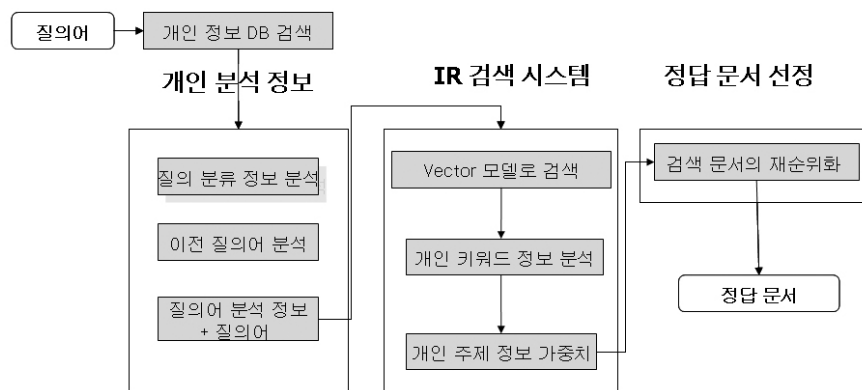
〈그림 1〉 구현된 개인화 검색시스템

〈그림 1〉은 개인화 검색시스템을 구현한 화면이다. 개인별로 로그인을 하여 개인 성향 정보를 기반으로 논문 키워드 주제 정보, 관련 저자 정보, 저자 네트워크 내비게이션, 자주 찾는 검색어, 자주 보는 문서, Social Network 등 다양한 정보를 제공하는 시스템을 개발하였다.

#### 1. 질의어 기반의 개인 성향 분석

아래의 〈그림 2〉와 같이 개인화 검색시스템은 사용자 성향을 분석하는 개인 분석 정보시스템과

분석된 정보를 이용하여 검색을 처리하는 정보 검색시스템으로 구성이 된다. 분석된 사용자 성향 정보를 이용하여 검색 결과에 반영하고 사용자 특성에 맞는 결과를 제공하는 시스템이다.<sup>19)</sup> 선행 논문의 시스템은 콘텐츠의 주제 분류인 DDC를 이용하여 사용자의 성향을 판단한 시스템이었다. 반면 본 연구에서는 KISTI의 기사 주제 분류를 기반으로 사용하였다. 즉 다양한 주제 분류에서도 시스템이 유연한지 여부를 실험적으로 테스트하기 위해서이다.



〈그림 2〉 질의어 기반의 개인화 검색시스템

본 연구에서는 〈그림 2〉와 같은 시스템을 기반으로 하여 다양한 개인화 서비스를 연구 및 개발 하였다. 본 연구에서는 개인의 질의어를 이용하여 검색된 문서들을 주제별 분류하기 위해서 kNN 분류기를 사용하여 분류를 하였다. 주제별 분류 대상은 DDC, KDC, BIST<sup>20)</sup> 등의 다양한 주제 분류 정보들이 학회마을 서지 DB에 있지만 본 연구에서는 실험적으로 BIST 주제 분류를 이용하여 개인화 검색시스템을 구축하였다. BIST 주제 분류 중에서도 앞 두 자리인 AA, AB, AC 등과 같이 대분류만을 사용하였다.

〈표 2〉 BIST 분류

BIST 분류	설명	비고
AA	건설공학	
AB	건축공학	
AC	토목공학	
...	...	
RB	자원공학	

19) 김광영 등, 전계논문, pp.163-180.

20) KISTI의 과학기술 문헌분류(BIST).



본 연구에서는 개인 성향 정보를 분류하기 위해서 대분류를 선택하였다. 세부적인 분류 항목으로 개인의 성향 정보를 설정하는 것 보다 포괄적인 개념인 대분류로 하는 것이 실험적 결과로 볼 때 용이한 것으로 나타났다. 즉, 소분류 항목을 이용하여 개인의 성향 정보를 분석할 경우에는 소분류 항목에 벗어난 범위에 대해서는 개인의 성향을 결정하기가 어려웠다. 그러므로 본 연구에서는 대분류 항목을 개인의 성향을 분석하는데 사용하였다.

예를 들면 컴퓨터 관련 분야에 관심이 많은 사용자들 대부분의 질의어들은 “컴퓨터/전산학” 관련될 것이다. 그러므로 관심 주제 분류를 이용하여 그 분야의 주제 분류를 중심으로 가중치를 더 많이 줄 수가 있다. 만약 “바이러스”라는 키워드로 검색을 할 경우에 성향 정보 분석을 위해서 검색된 문서들을 주제 분류 정보를 이용하여 분류기로 분류한다. 그 결과 대부분의 문서들은 BIST 주제 분류에서는 “축산”으로 분류된다. 이것은 “바이러스” 키워드가 포함된 문서들의 비중이 축산 분야에 많음을 나타낸다. 그렇지만 컴퓨터 관련 분야에 관심이 있는 사용자가 “바이러스” 키워드를 검색 하더라도 실제 분류는 “축산”이지만 전체적인 관심도가 “전산학”임으로 전산학에서 나타나는 “바이러스” 키워드를 중심으로 문서들을 상위로 랭킹 시킨다. 이렇게 함으로써 개인의 성향에 맞는 적합한 문서들을 제공할 수가 있었다.

## 2. 공동 저자 군집화 내비게이션

공동 저자 군집화 내비게이션은 사용자가 찾고자하는 논문 정보에 대해서 같은 연구자들을 중심으로 군집화를 처리하여 관련 저자들의 논문을 빠르게 탐색하고자 한다. 개인의 성향 정보에 맞게 검색된 문서들을 중심으로 분석을 한다.

논문에는 공동 저자들이 같은 연구 주제를 가지고 지속적으로 연구 또는 개발을 할 것이라는 가정에 기반을 두어 실험적으로 시스템에 적용한 것이다. 본 연구에서는 논문의 특징인 공동 저자들을 중심으로 그 결과를 군집화 하여 관련된 비슷한 연구자들을 빠르게 검색 할 수 있도록 시스템을 개발하였다.

본 연구에서는 논문의 서지 DB에 있는 관련 저자 정보 중에서 첫 번째와 교신 저자를 중심으로 군집화를 처리하였다. 일반적으로 논문에서는 첫 번째 저자와 교신 저자가 중요할 것이라는 가정에 기반하여 실험적인 값을 적용하였다.

$$\langle \text{식 1} \rangle \quad Author(k) = \sum_{i=1}^{i=n} (\alpha * F_i) + (\beta * L_i) + (\gamma) M_i$$

$\langle \text{식 1} \rangle$ 은 저자에 대한 가중치는  $i=1$  첫 번째 저자부터  $n$  마지막 저자까지 반복되며 첫 번째 저자는  $F$ 이며 중간 저자는  $M$ 이며 마지막 저자는  $L$ 이다. 실험적으로  $\alpha$  값은 0.3을  $\beta$  값은 0.6을

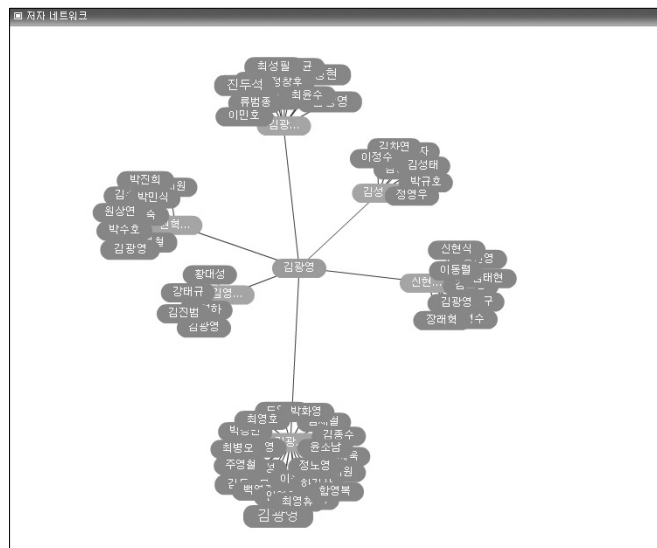
그리고 마지막으로  $\gamma$  값은 0.1로 설정하였다. 위의 상수 값들은 환경에 따라 다르게 설정할 수가 있다. 본 연구에서는 실험적으로 첫 번째와 마지막 저자를 중심으로 군집화를 처리하기 위해서 설정한 값들이다. 마지막 교신 저자에 대한 가중치를 더 많이 한 것은 보통 교신 저자의 경우에는 계속적으로 연구를 하지만 첫 번째의 저자의 경우에는 연구 분야가 많이 바뀌게 되는 것을 실험적으로 볼 수가 있었다.

〈표 3〉 단계별 군집화 단계

1단계 : Last-Author 중심으로 공동 저자 군집화 처리
2단계 : 이름, 소속기관, eMail 별로 군집화
3단계 : 군집화별로 학회, 토픽 정보를 제공

〈표 3〉과 같이 1단계의 군집화 처리는 위의 〈식 1〉과 같이 처리하여 저자별로 가중치를 계산하여 마지막 저자를 중심으로 군집화를 처리하였다. 그리고 2단계에서는 서로 다른 문서에서 발생하는 같은 사람에 대해서는 첫 번째는 이름을 기준으로 그 다음 단계를 소속 기관을 중심으로 마지막에는 eMail를 중심으로 군집화를 처리하였다.

본 연구에서는 서지 DB에서 2단계에 있는 eMail 정보가 누락 된 경우에는 처리를 하지 않는다. 만약 eMail정보가 없을 경우에는 있는 사용자보다 정확도가 낮아서 작은 클러스터들이 많이 나타나게 된다.



〈그림 3〉 군집화 검색 결과

<그림 3>은 저자가 검색을 했을 때 저자 군집화를 한 결과를 보여 주고 있다. 위의 결과는 현재의 연구 그룹과 다른 동명이인의 그룹들로 나타나는 것을 볼 수가 있다.

본 연구에서는 단순히 군집화 결과만을 보여 주는 것이 아니라 위와 같은 군집화에서 특정 사용자를 클릭할 경우에는 그 주변의 연구자들을 중심으로 검색을 다시 수행하여 처리한다. 위의 군집화 처리는 내비게이션 기능을 제공하며 실제 개인 사용자는 특정 그룹을 클릭함으로써 관련된 저자들이 동시에 나타나는 문서들을 쉽게 찾을 수 있도록 하였다.

순위	검색 결과
1	제목 : 정보검색관리시스템 KRISTAL-2000 설계 및 구현 (2003) <b>【유사문서】</b> 저자명 : 이민호;진두석;김광영;주원균;서정현;류범종; (학회 : 한국정보과학회) (DDC : 600) (ACC : EJ0200)
2	제목 : 문서 단편화 기법을 이용한 XML 변환기의 설계 및 구현 (2004) <b>【유사문서】</b> 저자명 : 정창후;최운수;주원균;진두석;김광영;이민호;서정현; (학회 : 한국정보과학회) (DDC : 600) (ACC : EJ0200)
3	제목 : 문서 단편화 기법을 이용한 XML 저장 관리 시스템 개발 (2005) <b>【유사문서】</b> 저자명 : 정창후;최운수;서정현;진두석;김광영;주원균;이민호; (학회 : 한국정보과학회) (DDC : 600) (ACC : ND0303)
4	제목 : KRISTAL2002 상의 검색 성능 비교 연구 (2003) <b>【유사문서】</b> 저자명 : 김광영;진두석;최성필;서정현; (학회 : 한국정보처리학회) (DDC : 600) (ACC : ND0300)
5	제목 : 정보검색관리시스템 KRISTAL-2001 (2001) <b>【유사문서】</b> 저자명 : 진두석;서정현;최운수;최성필;김광영; (학회 : 한국과학기술정보연구원) (DDC : ) (ACC : )
6	제목 : 위치 정보를 이용한 확장 벡터 모델의 문서 길의 정규화에 관한 연구 (2003) <b>【유사문서】</b> 저자명 : 김광영;서정현;이민호;주원균;정창후;류범종; (학회 : 한국정보처리학회) (DDC : 600) (ACC : NA1500)
7	제목 : 이웃한 어절간의 위치 정보를 이용하여 KRISTAL2000 DBMS 검색 성능 향상 (2002) <b>【유사문서】</b> 저자명 : 김광영;서정현;최성필; (학회 : 한국정보과학회) (DDC : 600) (ACC : EJ0200)

<그림 4> 특정 군집화 클릭할 때 검색 결과

<그림 4>와 같이 군집화에서 특정 그룹을 클릭할 경우에 그 그룹과 관련된 저자들이 많이 나타나는 문서들을 중심으로 검색 결과를 보여 준다. 위와 같은 검색 결과를 제공함으로써 개인 사용자들은 자신이 찾고 싶은 연구자 중심의 문서를 보다 빠르고 정확하게 찾을 수가 있었다.

### 3. 저자 키워드 기반 주제어 추천 시스템

본 연구에서는 사용자의 질의어 히스토리 정보를 이용하여 사용자가 자주 찾는 토픽 정보를 자동으로 추천한다. 주제어 정보는 논문에서 키워드 정보를 중심으로 추출되며 사용자가 선정한 문서에 대해서 검색을 하여 상위 30개의 문서에 대해서 키워드 정보를 추출한다. 상위 30개의 문서를 선정한 이유는 검색 결과가 최소 상위 30%까지는 정확하다고 가정을 하였다. 물론 검색 결과 전체를 사용해도 가능하나 많은 문서를 선정할 때에는 시스템의 속도에 많은 영향을 준다. 본 논문에서는 실험적으로 상위 30개로 한정을 하여 실험을 하였다.

저자들의 키워드 정보의 가중치는 첫 번째, 두 번째, 세 번째 순으로 가중치를 계산한다.

$$\langle \text{식 2} \rangle \quad T(w_t) = \sum_{n,m,l} (n\alpha + m\beta + l\gamma)w_t \quad (\text{단, } \alpha > \beta > \gamma, \alpha + \beta + \gamma = 1.0)$$

〈식 2〉에서와 같이 키워드 중에서 첫 번째 것이 가장 높은 가중치를 가진다. n, m, l은 가중치 Wt에서 전체 첫 번째, 두 번째, 세 번째 키워드들에 대해서 일정한 전체 α, β, γ 값의 합한 값을 의미한다. 본 논문에서는 α, β, γ 값은 상수 값으로 조정 될 수가 있다. 본 논문에서는 α=0.6, β=0.3, γ=0.1로 실험적인 값을 사용하였다. 이것은 몇 번째 키워드에 대해서 가중치를 더 많이 줄 것인가를 결정할 수 있다. 본 연구에서는 저자들이 첫 번째 키워드가 가장 핵심 주제 정보를 그 다음 순위로 저자들이 키워드를 작성할 것이라고 가정을 하였다.



〈그림 5〉 사용자 질의어와 저자 키워드 기반 자동 추천 주제 정보

〈그림 5〉는 사용자 질의어와 관련된 저자 키워드 토픽 주제 정보를 자동으로 추천하는 태그 그룹 시스템이다.

#### 4. 유사한 사용자 추천 시스템

본 연구에서는 사용자의 성향 정보를 이용하여 자신과 유사한 사용자를 찾아서 유사한 사용자가 자주 보는 문서와 자주 찾는 키워드 정보를 자동으로 추천하여 준다. 사용자의 성향 정보 프로파일은 자동으로 저장된다.

유사한 사용자 추천 시스템은 〈그림 6〉과 같이 사용자별 Profile 정보를 이용하여 사용자의 유사한 성향 정보 값과 유사한 키워드 정보 값을 이용하여 유사도 값을 측정한다. 성향 정보 값은 사용자별로 가장 높은 성향 정보들을 추출하여 같은 성향의 정보를 가진 사용자들을 찾는다. 또한 사용자별로 자주 찾는 키워드 정보를 이용하여 유사한 키워드들이 있을 경우 유사도를 계산하게 된다.

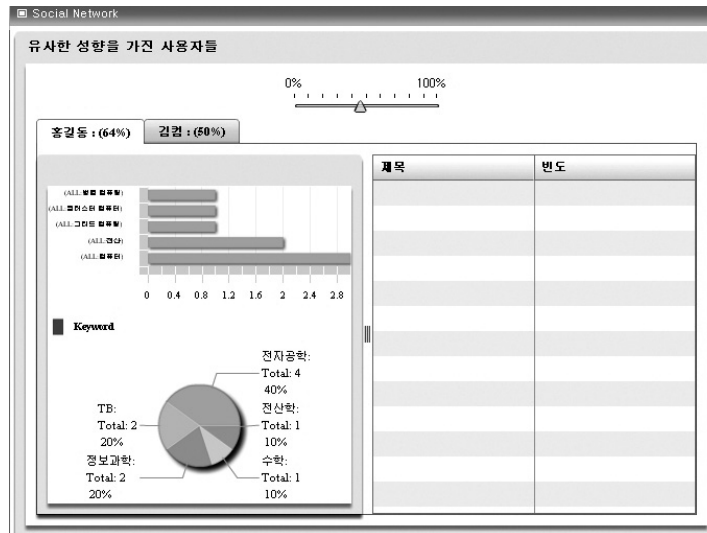
〈그림 7〉은 유사한 사용자를 자동으로 하는 것으로 유사도 값에 따라 사용자를 선별하게 된다. 유사도 값은 0.0에서 1.0까지로 계산한다.

사용자 Profile 정보				
질의어	질의어 빈도	분류정보	관련저자	...
질의어	질의어 빈도	분류정보	관련저자	...
질의어	질의어 빈도	분류정보	관련저자	...

사용자 문서 Profile 정보		
사용자ID	문서 번호	...
사용자ID	문서 번호	...
사용자ID	문서 번호	...

〈그림 6〉 사용자 Profile 정보



〈그림 7〉 유사한 사용자 자동 추천 시스템

## V. 평가 및 결과

본 연구의 실험을 위해서 사용한 DB는 KISTI가 운영하는 과학기술학회마을에서 서비스 중인 80만 건의 국내 논문의 서지 DB를 사용하였다. 논문은 제목, 초록, 키워드 정보가 포함 된 것들이 다. 검색시스템의 서버 사양은 리눅스 Redhat 4.1.2, 메모리 12G, 2CPU 인텔 Xeon 1.6GHz를 사용하였다. 과학기술학회마을 DB의 섹션 구성은 <표 4>와 같다.

〈표 4〉 과학기술학회마을 색션 구성

색션이름	설명
TIK	한글 제목
TIE	영어 제목
ABK	한글 초록
ABE	영문 초록
KW	키워드 정보
AUK	한글 논문 저자명
...	...

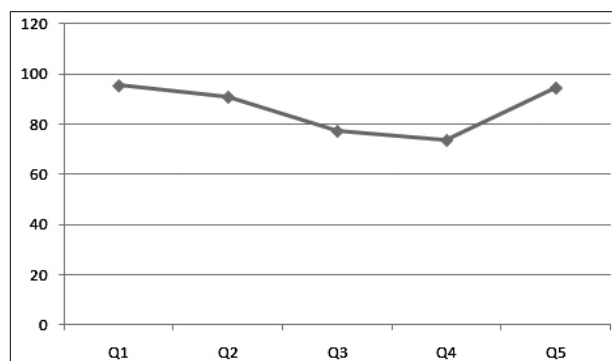
본 논문에서는 〈표 4〉와 같이 한글 제목, 영문 제목, 초록, 영문 초록, 키워드, 저자명 등을 각각 하나의 색션으로 나누고 이것들이 전체 하나의 문서를 이루는 구조로 적재하고 색인하여 사용하였다.

본 연구에서는 도출된 개인화 검색 시스템과 서비스들을 이용하여 베타 테스트를 하였다. 우선적으로 내부 전문가들 22명을 중심으로 설문 조사를 수행하였다. 이를 바탕으로 향후 시스템을 개선하여 실제 서비스에 반영하고자 한다. 내부 전문가들의 협조를 얻어 빠른 시간 내에 설문 조사를 할 수 있었고 기간은 2주정도가 소요되었다.

평가 결과로는 아래의 〈표 5〉와 같다. 〈표 5〉는 5점 만점을 100분율(%)로 환산하였다. 일반 검색보다 개인화 검색 결과에 대한 만족도가 가장 높고, 검색 결과를 2차원으로 보여 주는 것과 저자 내비게이션 기능에 대한 사용자 만족도가 가장 높았다. 저자 기반의 자동 토픽 정보와 소셜네트워크를 이용한 유사한 사용자 추천 서비스에 대한 만족도는 그렇게 높지 않은 것으로 나타났다.

〈표 5〉 설문 조사 결과 만족도에 대한 평균값

Q1	Q2	Q3	Q4	Q5
95.45	90.90	77.27	73.63	94.54



〈그림 8〉 설문 조사 결과 만족도에 대한 평균값

본 연구에서는 과학기술학회마을 논문 개인화 서비스를 위한 대표적인 실험적 다섯 개의 서비스들에 대한 설문조사를 했다. 본 연구에서는 평균 80%이상이면 사용자들이 선호하는 서비스 일 것이라고 가정하였다. 그 결과 <그림 8>과 같이 일반 검색 보다는 개인화 검색 결과의 만족도가 95.45% 가장 높았고, 검색 결과를 2차원으로 보여 주는 것은 94.54% 나타났고 저자 내비게이션 기능은 90.9%로 선정이 되었다. 이 결과와 같이 빠르게 적합한 문서를 찾을 수 있는 방법들을 선호하였다.

본 연구에서 개인화 검색서비스에 대한 만족도에 대해서 사용자들이 얼마나 일관성 있는 응답을 하였는가를 판단하기 위해서 신뢰도분석을 하였다.

<표 6> 항목 통계량

	평균	표준편차	N
Q1	4.77	.429	22
Q2	4.55	.510	22
Q3	3.86	.774	22
Q4	3.68	.646	22
Q5	4.73	.631	22

<표 7> 항목 총계 통계량

	항목이 삭제된 경우 척도 평균	항목이 삭제된 경우 척도 분산	수정된 항목 - 전체 상관관계	항목이 삭제된 경우 Cronbach 알파
Q1	16.82	3.108	.572	.615
Q2	17.05	2.903	.573	.599
Q3	17.73	2.113	.642	.538
Q4	17.91	3.134	.265	.720
Q5	16.86	3.076	.309	.700

본 연구에서는 신뢰도 계수  $\alpha$ 는 검사의 내적 일관성을 나타내는 값으로서 한 검사 결과는 알파 값은 0.691로 0.5이상으로 신뢰할 수 있다고 볼 수가 있다.

저자 기반의 자동 토픽 정보와 소셜네트워크를 이용한 유사한 사용자 추천 서비스에 대한 만족도는 그렇게 높게 나타나지 않은 원인에 대해서는 내부 의견 수렴 결과 다른 기능에 비해서 상대적으로 만족도가 낮은 것으로 나타났다. 하지만 소셜네트워크 기능에 대해서는 정확도가 낮은 것으로 분석 되었다. 향후 다양한 실험과 방법을 이용하여 소셜네트워크의 정확도를 높여 실제 서비스에 반영할 것이다.

## VI. 결론 및 제언

본 연구에서는 기존의 과학기술학회마을에서 제공하고 있는 일반 검색서비스에 대해서 분석하였다. 학회마을 서비스는 현재 논문 검색 부분에서는 기본 검색, 고급 검색, 통합 검색을 지원하고 학술지 검색 부분에서는 학술지별, 학회별, 주제별 검색 기능을 제공하고 있다. 일반 사용자와 전문가 사용자로 두 그룹으로 나누어 검색서비스를 제공하고 있다. 또한 학술지명, 학회명 검색지원과 DDC분류법을 이용하여 학술지를 분류 방법을 이용한 주제별 학술지 검색을 제공하고 있다. 이와 같이 현재의 과학기술학회마을 검색서비스는 일반적인 검색서비스만을 제공하고 있으며 개인의 취향이나 개인의 성향 정보를 반영을 하지 않고 있다.

본 연구에서는 과학기술학회마을 논문 검색서비스에 적합한 개인화 서비스들을 고려하였다. 과학기술학회마을 논문 데이터들은 공동저자 정보, 저자 키워드 등의 구조적인 정보들로 구성되어 있으므로 개인화 서비스를 위한 다양한 기능들을 도출할 수가 있다. 또한 주제별 분류를 위한 DCC, KDC 등의 분류된 정보들을 이용하여 개인화 검색시스템 개발하기가 용이하다. 따라서 본 연구에서는 과학기술학회마을에 서비스하고 있는 국내 논문들에 적합한 개인화 검색서비스들을 개발하였다. 즉, 질의어 기반의 개인화 검색시스템, 논문과 관련된 공동 저자 내비게이션 시스템, 저자 키워드기반 주제어 추천 시스템과 유사한 사용자 자동 추천 시스템을 개발 및 구현하였다. 이를 이용하여 과학기술학회마을의 대표적인 개인화 서비스를 선택하여 각 기능에 대한 사용자 만족도를 설문 조사를 하였다. 그 결과로 일반 검색 보다는 개인화 검색 결과의 만족도가 높았고, 검색 결과를 2차원으로 보여 주는 것과 저자 내비게이션 기능으로 선정이 되었다.

향후 과제로는 논문 검색뿐만 아니라 다양한 정보자원에 대한 적합한 개인화 검색시스템의 개발이 요구된다. 또한 저자 키워드 정보와 소셜네트워크 기능을 더 확대한 연구가 필요하고 사용자 태깅 정보를 개인화 검색시스템에 반영하여 사용자의 질의어뿐만 아니라 태그 정보도 함께 고려한 모델이 연구되어야 한다.

〈참고문헌은 각주로 대신함〉



〈부록〉 개인화 검색서비스 만족도 설문조사

질 문	매우 그렇다	그렇다	보통 이다	그렇지 않다	매우 그렇지 않다
	5	4	3	2	1
1. 학화마을의 일반 검색 결과보다 개인화 검색 결과 만족도는 얼마입니까?					
2. 학화마을 개인화 서비스 저자 내비게이션 기능의 만족도는 얼마입니까?					
3. 학화마을 개인화 서비스 주제어(Topic) 자동 추천 서비스의 만족도는?					
4. 학화마을 개인화 서비스 소셜네트워크 만족도는?					
5. 학화마을 개인화 서비스 검색 결과에 2차원 표시(년도/가중치)에 대한 만족도는?					
6. 학화마을 개인화 검색서비스 가장 추천하고 싶은 서비스는? (3가지 선택)	<input type="checkbox"/> 사용자 질의어 태그 클라우드 <input type="checkbox"/> 관련 토픽 태그 클라우드 <input type="checkbox"/> 개인화 검색 결과 2차원 표시 <input type="checkbox"/> 저자 네트워크 내비게이션 <input type="checkbox"/> 소셜 네트워크 <input type="checkbox"/> 키워드 기반의 개인화 검색 결과				
7. 학화마을 개인화 서비스 추가하고 싶은 서비스가 있다면? (주관식)					

