주제분류 기반의 개인화 검색시스템에 관한 연구*

A Study on Personalized Search System Based on Subject Classification

김 광 영(Kwang-Young Kim)**
곽 승 진(Seung-Jin Kwak)***

목 차

1. 서 론
2. 이론적 배경
3. 연구의 방법
4. 시스템 설계 및 개발
5. 연구 결과 분석 및 평가
6. 결론 및 제언

초 록

본 연구의 목적은 이용자의 자신의 성향 정보에 맞는 정확한 검색결과를 제공하기위한 개인화 검색시스템을 설계, 구현, 평가하는 것이다. 이를 위해서 주제 분류와 하이브리드기반의 이용자 프로파일을 구성하였다. 제안한 개인화 검색시스템의 성능 평가를 위해서 전문가들이 직접 국내과학기술 DB의 논문의 적합성을 판별하기위해서 MRR, MAP과 사용성 평가를 측정하였다. 그 결과 전문가가 직접 "컴퓨터공학"과 "문헌정보학" 분야에서 평가한 국내과학기술논문 결과에서도 제안한 개인화 검색시스템이 일반 검색시스템보다 정확도가 더 높았다. 특히 중의성을 가진 키워드 부분에서 더 높은 정확도를 보였다. 질적 평가인 심층면담을 통한 평가에서도 개인화 검색시스템이 이용자가 선호하는 문서들을 우선적으로 제공함으로써 정보검색 시간 및 정보 획득에 용이하다고 응답하였다. 또한 다양한 추천 시스템을 제공함으로써 새로운 정보 탐색에 도움이 된다하였으며, 개인화 검색시스템이 일반 검색시스템보다 이용자의 만족도가 높음을 알 수 있었다. 따라서 본 연구에서 전문가 직접 평가 결과 모두 제안한 개인화 검색시스템이 일반 검색시스템보다 더 높은 효율성을 제공하는 것을 입증하였다.

ABSTRACT

The purpose of this study is to design, implement and evaluate a personalized search system using gathered information on users to provide more accurate search results. For this purpose, a hybrid-based user profile is constructed by using subject classification. In order to evaluate the performance of the proposed system, experts directly measured and evaluated MRR, MAP and usability by using the Korean journal articles of science and technology DB. Its performance was better than the general search system in the area of "Computer Science" and "Library and Information Science". Especially better results were shown when tested on ambiguous keywords. Evaluation through in-depth interviews proved that the proposed personalized search system was more efficient in looking up and obtaining information. In addition, the proposed personalized search system provided a variety of recommendation systems which proved helpful in navigating for new information. High user satisfaction ratings on the proposed personalized search system were another proof of its usefulness. In this study, we were able to prove through expert evaluation that the proposed personalized search system was more efficient in information retrieval.

키워드: 주제분류, 개인화 검색시스템, 이용자 프로파일, 개인화
Subject Classification, Personalized Search System, User Profile, Personalization

---

* 본 논문은 충남대학교 박사학위 논문의 일부를 요약한 것임.

** 한국과학기술정보연구원 정보기술연구실 선임연구원(kykim@kisti.re.kr)
*** 충남대학교 사회과학대학 문헌정보학과 부교수(sjkwak@cnu.ac.kr)

논문접수일자: 2011년 10월 1일 최초심사일자: 2011년 10월 17일 게재확정일자: 2011년 11월 17일
한국문헌정보학회지, 45(4): 77-102, 2011. [http://dx.doi.org/10.4275/KSLIS.2011.45.4.077]

1. 서 론

1.1 연구의 필요성 및 목적

오늘날 인터넷의 확산에 따른 정보의 과잉 속에서 정보이용자들은 자신이 원하는 정보를 찾기 위해서 많은 시간과 노력을 필요로 한다. 이러한 많은 정보들 중에서 이용자들이 선호하는 정보를 선별적으로 제공하기 위한 접근 방법 중의 하나로 개인화 검색시스템이 등장하게 되었다.

개인이 선호하는 정보를 분석하고 찾기 위해서는 다양한 접근 방법들이 요구된다. 개인화 검색시스템은 개인 이용자별로 맞춤형 정보를 만들고, 관리하고 표현하기 위한 많은 요소들을 해결을 해야 하며, 개별 맞춤형 정보는 비적합한 정보와 이용자가 관심 있는 정보로 구별되어야 한다.

개인화 관련 연구는 정보검색과 인공 지능, 데이터 마이닝 등의 다양한 분야에서 연구되고 있다. 초기 개인화 검색시스템 연구는 개인화 e-mail, 전자 신문, 유즈넷 그룹과 웹 문서의 필터링과 등급 시스템들에 관심이 집중되었다. 그 후 개인화 검색시스템의 연구는 브라우징 보조 기능 제공과 웹 사이트의 내비게이션 기능에 관심을 가졌다(Chaffee and Gauch 2000). 오늘날 웹 검색은 가장 일반적인 활동 중의 하나임으로 개인화 검색시스템에 관한 연구도 웹 검색에 중점을 두고 있다(Liu and Yu 2004; Speratta and Gauch 2005; Jiang and Tan 2009; Matthijs and Radlinski 2011).

웹 검색 업체들도 개인화 특징들을 채택하고 있다. 2005년도에 Yahoo의 개인화 웹 포털과 Google Lab 개인화 검색들이 대표적인 예이다. 웹 포털에서 개인화는 개인별 이용자들에 초점을 맞추어 서비스를 다양한 형태로 제공하고 있다. 즉 이용자가 특별하게 선택한 주제에 대한 관련 뉴스, 주식정보와 상업적 사이트의 광고 등을 제공하는 것이다.

현재 대부분의 개인화 검색시스템들은 이용자 프로파일을 이용하여 이용자 맞춤 검색을 수행하고 있다. 이용자 프로파일은 일반적으로 통계 정보를 포함한다. 예를 들면 이용자 그룹이나 개개인의 관심이나 선호도를 나타내는 이름, 나이, 성별, 국가, 소속기관, 전공 등과 같은 요소들을 이용한다.

일반 검색시스템에서는 서로 다른 사람이 같은 키워드로 검색할 경우에 똑 같은 결과를 제공 받는다. 이것은 정보 요구가 서로 다른 사람들에게는 적합하지 않다. 예를 들면 "바이러스"라고 입력을 했을 때, 어떤 사람은 생물/의학에 관련된 "바이러스"에 대한 정보를 찾을 것이고 다른 사람은 컴퓨터에 관련된 "바이러스"에 대한 정보를 찾기를 원할 수도 있다. 이와 같은 언어의 중의성 문제들은 현실 속에서 많이 발생한다. Shen(2005), Qui(2006) 및 김광영(2009) 등은 개인화 검색시스템에서 이용자 프로파일 정보를 이용하면 이와 같은 언어의 중의성 문제를 해결할 수 있다고 주장하였다. 이와 같이 현재 개인화 검색시스템에서는 개인 이용자에게 가장 적합한 검색 결과를 제공하기 위한 다양한 방법들이 연구되고 있다.

이에 본 연구는 개인화 검색시스템에 관한 다양한 접근과 방법들이 논의되고 있는 상황에서 개인화 검색시스템의 성능을 제고시키기 위한 방법의 하나로 주제분류 기반의 개인화 검색시

스템의 효과에 관심을 갖고 도서관 및 학술정보센터의 학술정보 이용자들에게 개인별 정보요구에 부합하는 검색결과를 제공하기 위하여 주제분류 기반의 개인화 검색시스템을 구축하고 그 성능을 평가하고자 한다.

2. 이론적 배경

2.1 개인화 시스템

일반적으로 개인화(personalization)라는 용어는 이용자의 정보요구에 부합되는 콘텐츠를 제공한다는 의미로 광범위하게 사용되며(Shahabi and Chen 2003), 이용자에게 부합되는 콘텐츠를 제공하기 위해서는 먼저 이용자의 정보요구를 수집해야 한다.

개인 정보는 주로 두 가지 방식으로 수집된다. 하나는 명시적 방법으로 관심 정보나 평가등급과 같이 피드백 정보를 요청하는 방법과 다른 하나는 암시적 방법으로 온라인 문서를 읽는 시간과 같은 이용자의 행동을 관찰하는 방법이다(Speretta and Gauch 2005).

개인화 검색시스템은 수집된 이용자의 정보를 프로파일에 저장하고 관리하며 이용자의 프로파일 정보를 이용하여 개인의 성향을 분석하고 추론한다. 이용자의 프로파일을 구성하는 방법은 일반적으로 키워드 프로파일, 시맨틱 네트워크 프로파일 및 개념/온톨로지 프로파일이 있다.

2.2 선행 연구

국내에서도 다양한 이용자 프로파일기반 개인화 검색시스템에 관한 연구들이 진행되고 있으며 특히 네트워크 및 개념기반의 프로파일을 중심으로 연구되고 있다.

김경중과 조성배(2001)는 이용자 프로파일을 퍼지 개념 네트워크기반으로 구성하고 구축된 퍼지 문서 추출 시스템을 이용하여 이용자의 성향에 반영하는 개인화 검색시스템을 제안하였다. 생성된 퍼지 개념 네트워크로 개인화된 문서를 추출하여 최종적으로 이용자에게 가장 적합한 웹 문서를 선택하게 하는 방식이다.

문현전(2007) 등은 도메인 온톨로지를 사용하여 개인화된 개념기반의 검색 기법을 제안하였다. 도메인 온톨로지를 이용하여 콘텐츠의 대표 개념 추출, 콘텐츠 가중치와 개념 가중치를 이용하여 이용자 프로파일을 구성한 개념기반 개인화 검색시스템이다. 콘텐츠의 대표 개념은 TScore[1] 기법을 이용하여 추출하였고, 이용자 프로파일은 개인 정보 수집 모듈을 통해 개념 가중치가 높은 개념을 대상으로 구성하였다. 개념기반의 검색을 위해서는 이용자 프로파일의 개념 집합과 콘텐츠의 대표 개념 집합 간에 유사도를 비교하여 이용자가 선호하는 개념을 우선순위로 콘텐츠를 검색하였다.

박건우와 이상훈(2008)은 웹 이용자들의 키워드 사용 빈도수에 대한 랭킹 정보를 통해 최근 주요 관심사를 파악하고, 주요 관심사 별로 형성된 커뮤니티를 이용한 개인화 검색시스템을 제안하였다. 웹 이용자들의 질의어 패턴을

1) TScore 기법은 콘텐츠에서 대표 개념을 추출하기 위하여 도메인 온톨로지의 개념별 연관 용어집합과 콘텐츠에서 추출한 대표 용어간의 유사도를 비교하기 위한 기법.

분석하고 주요 관심사를 기반으로 형성된 커뮤니티를 통해서 검색을 수행하는 방식이다. 웹 이용자의 질의어 빈도수를 자동으로 확인 및 랭킹을 부여 하여 데이터베이스로 구축한 후 질의어에 대한 랭킹정보를 통해서 개인의 관심사를 파악하였다. 또한 상위에 순위화된 질의어 정보를 통해 주요 관심사 별 커뮤니티 형성을 가능하게 하고 자신과 관심분야가 유사한 커뮤니티를 검색에 반영하는 개인화 검색시스템이다.

이성직(2009) 등은 개념 네트워크 모델기반의 이용자 프로파일을 구성하였다. 그 모델을 구성하는 기본적인 키워드들 간의 연관 관계를 찾기 위해서는 중요 단어출현빈도를 사용하여 키워드를 추출하는 방법을 제시하였고 이용자가 입력한 검색 키워드들과 방문한 웹 문서들로부터 찾아낸 키워드들 간의 연관 관계가 있는 단어들을 사용한 개념 네트워크기반의 개인화 검색시스템을 제안하였다.

윤기상(2009) 등은 웹 2.0 환경에서 많이 사용되고 있고 앞으로도 더욱 활발히 사용될 것으로 예상되는 태그를 개인화 검색에 활용할 것을 제안하였다. 태그 카테고리를 사용하여 이용자의 관심 프로파일을 생성하고 이용자가 검색어를 입력하면 기존 검색 엔진을 이용하여 상위에 있는 웹 페이지의 주소와 태그를 추출하고 이용자의 관심사항이 반영된 태그 카테고리를 사용하여 순위를 조정하였다.

윤홍준(2010) 등은 개인화 검색을 위해 개념 네트워크 형태의 이용자 프로파일을 구성하고 이용자가 과거에 검색했던 키워드들과 탐색했던 웹 페이지 문서들을 분석하였다. 또한 이용자가 웹 페이지에 지정한 태그 데이터를 이용하여, 개념 네트워크를 확장함으로써 개념 네트워크의 정확성을 더 높고 이렇게 생성된 개념 네트워크를 이용하여 이용자가 검색한 질의어를 확장하여 이용자 별로 검색결과의 순위를 재산정하여, 이용자 개인의 특성을 반영한 검색결과를 제공하였다.

최근 국외의 개인화 검색시스템들은 개념기반의 프로파일에 관련된 많은 연구들을 수행하고 있다. Persona 시스템은 모든 개념을 Open Directory Project의 분류체계를 중심으로 미리 분류된 웹 페이지들을 이용하여 명시적 피드백을 적용하여 tree-coloring[2] 기법을 이용한 개념기반의 프로파일을 구성하였다. tree-coloring 기법을 이용하여 이용자의 관심 정보와 비관심 정보로 나누고 미리 분류된 문서들을 검색할 때 명시적으로 관심 정보인지 아닌지를 피드백 받아서 이용자 프로파일을 갱신하였다(Tanudjaja and Mui 2002).

PVA 시스템은 Open Directory Project의 55개념들을 중심으로 웹 페이지들을 이용하여 암시적기반의 이용자 피드백을 이용하여 분류 개념을 추가/삭제할 수 있는 문서 분류 기법을 이용하였다(Chen, Chen, and Sun 2002).

Foxtrot 추천시스템은 CORA의 97개의 개념들을 중심으로 미리 분류된 연구보고서 페이지들을 이용하여 명시적 및 암시적 피드백을 적용하여 tree-coloring 기법을 이용한 개념기반의 프로파일을 구성하였다. Foxtrot 추천시스템은 Persona와 유사한 접근 방식으로 이용자와 상호작용을 하기 위해서 명시적 및 암시적

2) 하위단계에서 상위단계의 개념으로 적용시키는 기법, 즉 자식노드로부터 상위 부모노드 또는 그 상위의 조상노드까지 적용시키는 방법.

피드백 정보를 수집하고 암시적 피드백에 대한 개념 가중치보다 명시적 개념 가중치에 10배 이상 부여하였다(Middleton et al. 2003).

OBIWAN 시스템과 Misearch 시스템은 Open Directory Project의 2,000 개념들을 중심으로 웹 페이지나 키워드들과 검색결과들을 이용하여 암시적 피드백을 적용하고 문서 분류 기법을 이용하여 개념기반의 프로파일을 구성하였다. ODP목록과 문서들을 학습 데이터로 활용하여 이용자의 브라우징과 같은 활동 정보들을 이용하여 학습 데이터를 분류 후에 이용자의 프로파일에 반영하였다(Chaffee and Gauch 2000; Sieg, Mobasher, and Burke 2004).

Liu와 Yu(2004)는 Open Directory Project의 619 개념들을 중심으로 검색결과와 키워드들을 이용하여 암시적기반의 이용자 피드백을 구성하고 카탈로그를 이용하여 명시적기반의 이용자 피드백을 혼합한 형태를 적용하여 학습에 기반을 둔 문서 분류 기법을 이용하였다.

Bibster 시스템은 전산기계협회(ACM) 토픽 계층구조인 1,287의 개념들을 중심으로 키워드들을 이용하여 이용자의 암시적 피드백과 명시적 피드백으로 개념들을 추가하는 혼합 형태를 적용하여 Tree coloring 기법과 다른 유사한 이용자의 프로파일을 추천하는 기법을 이용하였다(Haase et al. 2005).

Xu, Bao와 Fei(2008)는 이용자들의 태그를 활용한 폭소노미 데이터를 이용하여 개인화 검색을 위한 프레임워크를 제안하였다. 특히 카테고리, 키워드와 구조의 세 가지 폭소노모의 속성을 활용하여 개인화 검색시스템의 프레임워크에서 질의와 웹 페이지의 내용이 일치하는 용어뿐만 아니라 이용자의 관심 정보와 웹 페이지의 주제 정보가 일치하는 문서들을 랭킹에 반영하였다.

Daoud, Tamine-Lechani와 Boughanem(2009)은 온톨로지기반의 이용자 프로파일을 이용하여 개인 성향 정보를 분석하여 검색결과에 반영하였고 HARD 2003 TREC을 이용하여 개인화 검색시스템 성능을 평가하였다.

Carmel(2009) 등은 이용자의 소셜 관계를 기반으로 한 개인화 소셜 검색시스템을 제안하였다. 개인화 소셜 검색시스템은 이용자의 소셜 네트워크에서 개인별 관계에 따라서 검색결과를 재순위화하였다. 이용자가 명시적으로 친근한 관계 사람들, 이용자의 사회적 활동과 유사한 이용자 등과 같은 소셜 네트워크의 관계성을 활용하였다.

3. 연구의 방법

3.1 연구의 설계

본 연구에서는 한국과학기술정보연구원(KISTI)의 과학기술학회마을에서 서비스하는 국내학술논문 DB을 활용하였다. 또한 2010년 6월 12일부터~2010년 8월 17일까지 67일간 수집한 이용자의 로그 파일을 분석하고 이를 바탕으로 컴퓨터와 문헌정보학관련 전문가들이 선정한 실제 이용자의 프로파일들을 구성하였다. 각 분야별로 전문가들이 선정한 질의 셋을 직접 검색하고 평가를 하여 그 성능을 측정하였다. 마지막으로 질적 평가를 위해서 전문가들과 심층면담을 통해서 사용성 평가를 수행하였다. 이와 같이 전문가들이 두 시스템을 직접 비교하고 평가하기 위

해서는 아래와 같은 절차를 통해서 수행하였다.

1) 국내학술논문 80여만 건을 이용하여 K-개인화시스템을 구축한다.
2) 구축된 K-개인화시스템과 기존의 과학기술학회마을의 K-일반시스템에 대해서 이미 분류된 KISTI 표준분류체계를 활용한다.
3) 67일간 수집된 로그데이터를 분석하여 "컴퓨터" 및 "문헌정보학" 분야에 이용자들의 프로파일 중에서 전문가들이 선정한 개별 이용자들의 프로파일을 각각 구성한다.
4) 전문가들이 K-개인화시스템과 K-일반시스템의 성능을 평가하기 위해서 우선 중의성을 가진 질의 셋을 이용하여 검색을 수행하고 적합한 문서에 대해서 역순위평균값(MRR)과 상위 n개에 대한 정확률을 측정한다.
5) 또한 전문가들이 직접 무작위로 일반 검색 키워드들의 질의 셋을 선정한 후 검색을 수행하고 적합한 문서에 대해서 역순위평균값과 상위 n개에 대한 정확률을 측정한다.
6) 마지막으로 K-일반시스템과 K-개인화시스템들에 대해서 선호하는 이유를 조사하고자 질적 평가를 위한 심층면담을 수행한다.

3.2 실험 환경 및 데이터

본 연구에서는 국내학술논문 80여만 건을 K-개인화시스템에 적재 및 구축을 하기 위한 과학기술학회마을 DB의 섹션 구성은 〈표 1〉과 같다. 시스템의 서버 사양은 리눅스 Redhat 4.1.2, 메모리 12G, 2CPU 인텔 Xeon 1.6GHz를 사용하였다.

〈표 1〉 학회마을의 섹션 구성

섹션 이름	설 명
TIK	한글 제목
TIE	영문 제목
ABK	한글 초록
...	...
DDC	듀이십진분류
ACC	KISTI 표준분류
IMF	원문유무

본 연구에서는 〈표 1〉과 같이 국내 논문의 서지 사항과 시스템 관리를 위한 섹션 등으로 구성이 되어있다. 전체 국내학술논문은 874,389만 건이고 한글 제목, 영문 제목, 초록, 영문 초록, 키워드, 저자명 등을 각각 하나의 섹션으로 나누고 이것들이 전체 하나의 문서를 이루는 구조로 적재하고 색인하여 사용하였다.

3.3 전문가 선정

선행연구에서 제시한 전문가 선정 과정을 구체적으로 살펴보면 다음과 같다.

Speratta(2000)는 그의 박사학위 논문에서 개인화 검색시스템과 Google 일반 검색시스템의 결과를 비교 평가하기 위해서 6개월여 동안 대학원생 6명(컴퓨터, 수학, 화학)과 상위 10위까지 문서에 대한 적합성 평가 및 연구를 수행하였다. 또한 가상의 2명의 이용자 프로파일을 학습시키기 위해서 240개의 질의어를 사용하였고 이용자별 5개의 질의어를 이용하여 적합성 테스트 등을 수행하였다.

Liu와 Yu(2002)는 개인화 검색시스템 평가를 위해서 7명의 이용자 프로파일을 구성하여 ODP 분류 개념에 맵핑시키기 위한 다양한 모델

방법을 연구하였다. Shen, Tan 및 Zhai(2004)는 개인화 검색시스템 결과와 Google 검색결과를 비교하기 위해서 대학원생 6명(컴퓨터 3명, 생물학 2명, 화학 1명)을 활용하여 상위 30위까지 문서에 대한 적합성 비교 평가를 수행하였다.

이와 같이 개인화 검색시스템과 일반 검색시스템 간의 적합성을 비교 평가하는 것은 많은 시간과 정확성을 요구하는 작업임으로 일반 이용자들이 참여하여 정확률을 측정하고 평하기에는 어려움이 따른다.

본 연구에서 K-개인화시스템와 K-일반시스템을 평가하기 위해 과학기술학회마을 이용자들 중에 문헌정보학과 컴퓨터 분야 전문가들 중에 각 분야별로 석사 이상의 학위를 소지자로 최소 10년 이상의 근무 경력자를 기준으로 자발적으로 참여를 원하는 이용자를 중심으로 선정하였다.

전문가는 컴퓨터와 문헌정보학 각 분야별로 5명씩 전체 10명을 구성하였다. 10명의 전문가들은 중의성 키워드 5개와 무작위로 입력한 일반 검색 키워드 5개를 이용하여 상위 50위까지 문서들에 대해서 적합성을 판정하였다.

3.4 이용자 프로파일 선정

3.4.1 컴퓨터관련 이용자 프로파일

컴퓨터 분야는 방대하지만 본 연구에서는 KISTI 표준분류체계에서 전산학(NA), 정보통신(NB) 및 정보가공(ND) 분야로 제한하였다. 본 연구에서는 제한된 분야에 대해서 1차에 선정된 이용자들의 검색 키워드를 중심으로 주제분류를 수행한다. 본 연구에서는 제한된 분야로 나타나는 이용자들에 대해서 주제분류 값이 높은 후보 30명의 실제 이용자 프로파일들을 전문가들에게 제공하여 적합한 이용자를 각각 선정하도록 하였다. 따라서 컴퓨터 분야에서는 5명의 이용자 프로파일을 구성하였으며 선정된 컴퓨터 분야 이용자의 프로파일 예는 〈표 2〉와 같다.

〈표 2〉 컴퓨터관련 이용자 예

이용자 IP	실제 이용자 키워드
128.134.xx.xx	음성인식, 키워드 검색, 음성인식 검색, 음성인식 검색, 멀티미디어 검색, 멀티미디어 내용검색, 멀티미디어 내용검색, 녹취록, 검색 키워드, 스파팅 키워드, 멀티미디어 검색, 내용기반 검색, 유전자 알고리듬, 유전자 알고리듬, 임베딩, 자바, 상황인식, 상황인식

3.4.2 문헌정보학관련 이용자 프로파일

문헌정보학관련 이용자 프로파일은 KISTI 표준분류체계에 있는 정보학(NC)분야로 제한하였다. 본 연구에서는 컴퓨터 분야와 같이 문헌정보학 분야 5명의 이용자 프로파일을 구성하였으며 선정된 문헌정보학 분야 이용자의 프로파일 예는 아래 〈표 3〉과 같다.

〈표 3〉 문헌정보학관련 이용자 예

이용자 IP	실제 이용자 키워드
211.184..xx.xx	MARC, 목록, 외주, 아웃 단가, 아웃소싱, 아웃소싱, 공공도서관, 용역, 외주, 공공도서관, 목록, 편목, 아웃소싱, 장서구성, 장서, 장서구성, 장서, 장서통계, 장서통계, 이러닝, 전자도서관, 사이버, 컨텐츠, 콘텐츠, 비도서, 비도서, 공공도서관, 공간운영, 공공도서관, 정부간행물의 출판현황과 도서관의 정부간행물의 운영 실태분석, 정부간행물, 간행물

3.5 전문가 평가 질의 셋 선정

본 연구에서는 선정된 이용자들의 실제 프로파일을 생성한 후 전문가들이 직접 평가하기 위한 질의 셋 선정 방법을 아래와 같이 두 가지 방법을 사용하였다.

3.5.1 중의성 단어 중심으로 평가 질의 셋 선정

중의성 질의 셋을 이용하여 각 분야별로 전문가들이 직접 검색을 수행하여 K-개인화시스템과 K-일반시스템 간의 각 분야별로 적합한 문서들에 대해서 평가를 수행하였다. 평가를 위해서 중의성이 있는 단어들로 구성을 한 이유는 다양한 분야에서 나타날 수 있는 단어들을 선택함으로써 해당 분야의 이용자에게 적합한 문서인지 여부를 전문가들이 쉽게 판단하기 위해서이다.

본 연구에서는 허정 등의 연구(허정, 옥철영 2006; 허정, 서희철, 장명길 2006)에서 제시한 중의성 단어들 및 중의성 관련 논문에서 찾은 것들로 컴퓨터 및 문헌정보학 관련 중의성 단어들로 구성하였다.

〈표 4〉와 같이 선정된 기준은 전문가들이 다양한 분야에서 나타날 수 있는 중의성을 가진 단어들을 선택하였다.

3.5.2 일반 단어 중심으로 평가 질의 셋 선정

전문가들이 무작위로 직접 일반 단어를 선정 후 각 단어들에 대해서 K-개인화시스템과 K-일반시스템 간의 성능을 비교 측정하였다. 즉 각 분야별로 각각의 전문가들이 직접 5개의 키워드들을 무작위로 선정하여 검색하고 평가하였다. 〈표 5〉는 각 분야별로 전문가들이 무작위로 선정한 일반 검색 키워드들이다. 문헌정보학 분야에서 전문가 두 명이 무작위로 선정한 키워드들 중에서 "정보서비스"의 키워드는 중복된 것으로 나타났다.

〈표 4〉 선정된 중의성 키워드

분야	중의성 키워드
컴퓨터	바이러스, 키, IR, cell, 백신
문헌정보학	수서, 판, 기술, LCC, LC

〈표 5〉 무작위로 선정된 일반 검색 키워드

분야	키워드	개수
컴퓨터	알고리즘, transition, 압축, 손실, window, 검색, 언어자원, 구조화, 의미 처리, 음성 개체 식별, 엔진, mining, 유전자 알고리즘, 대용량, OWL, 웹 2.0, 추론, 온톨로지, 시맨틱 시뮬레이션, 주문형, 컴포넌트, 개발 프로세스, 개발 주기	25
문헌정보학	전자책, 대학 도서관, 연구지원, 참고 봉사, 디지털도서관 이용 형태, 서비스 품질, 도서관 마케팅, 정보서비스, SDI, 자료 분석, 정보관리, 정보서비스, 사용성 평가, 도서, 이용자서비스, 기증, 목록, 보존, 봉사 대상, 교육, 외주, 서비스, 폐기, 도서 정책	25

3.6 양적 평가

본 연구에서는 적합한 문서가 상위에 얼마나 분포하는지 여부를 조사하기 위해서 역순위평균(MRR) 평가 방식을 사용하였다. 전문가가 평가한 적합한 문서들에 대해서 상위 10위까지를 정답 문서로 가정하고 K-일반시스템과 K-개인화시스템 간의 MRR값을 측정하여 시스템 간의 성능을 비교하였다.

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i}$$

또한 1~50위까지의 문서 순위 중에서 얼마나 적합한 문서들이 상위로 분포되는지를 보기 위해서 순위에 따라 역순위값으로 점수를 부여하고 1~10위, 1~20위, 1~30위, 1~40위, 1~50위 등과 같이 구간별 평균RR(reciprocal rank) 값을 측정하였다. 즉 전문가들이 평가한 적합한 문서들이 구간별로 어떻게 분포되는지 측정하였다.

마지막으로 본 연구에서는 상위 50개까지에 대해서 K-일반시스템과 K-개인화시스템 간의 검색결과에 상위 n 정확률을 아래 수식을 이용하여 측정하였다.

$$Top-n\ Precision = \frac{\#\ of\ relevant\ retrieved\ within\ n}{n}$$

$$Top-n\ Recall = \frac{\#\ of\ relevant\ retrieval\ within\ n}{total\ \#\ of\ revevant\ documets}$$

본 연구에서는 학회마을 논문 80여만 건에 대해서 각 질의에 대한 전체 적합한 문서 값을 구하지 못함으로 정확한 재현율 값은 측정할 수가 없다. 따라서 상위 50개에 대한 정확률만 측정하였다.

3.7 질적 자료 수집

본 연구에서 K-개인화시스템과 K-일반시스템의 선호도 차이를 비교 분석하기 위한 질적 자료는 전문가들이 테스트 후 연구자와 전문가들 간의 반 구조화된 면담을 통해 수집하였다. 반 구조화된 면담은 질문의 문항과 순서가 미리 정해져 있지만 질문의 문항은 개방형으로 구성되어 있으며, 면담 과정에서 필요한 경우 구체적인 사항에 대해 연구자가 심층적인 의견을 지속적으로 물어보는 면담기법이다.

면담은 개별적으로 이루어졌고, 1회에 걸쳐 약 30~40분 동안 진행되었다. 면담 내용은 피면담자의 동의하에 녹음하였고, 면담 내용의 정확한 기록을 유지하기 위해 핵심 내용은 연구자가 면담 설문지에 기록하였다. 녹음된 자료는 면담 종료 후 곧바로 원본의 내용을 그대로 유지하기위해서 워드프로세스로 기록하는 문서화 과정을 거쳤으며 기록된 자료는 분실을 방지하기 위해 복사해 두었다.

K-개인화시스템과 K-일반시스템의 선호도와 그 차이를 비교 분석하기 위한 면담 항목은 〈표 6〉과 같다.

〈표 6〉 면담 질문지 구성의 범위와 내용

면담 범위	면담 개요
시스템의 선호도 차이	K-개인화시스템과 K-일반시스템의 선호도와 그 이유
시스템의 사용성 경험	K-개인화시스템과 K-일반시스템을 사용하면서 좋았던 점, 불편했던 점 K-개인화시스템과 K-일반시스템을 사용하면서 느낀 구체적인 의견 제시

4. 시스템 설계 및 개발

4.1 개요

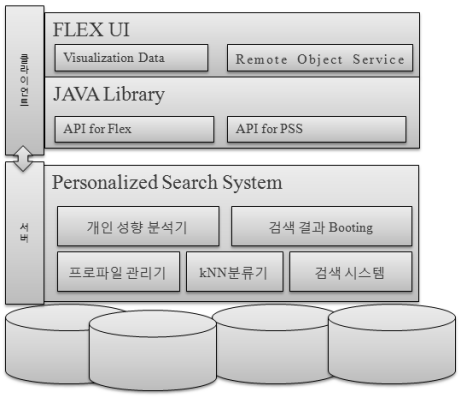
본 연구에서는 개인의 관심이나 기호 등을 분석하기 위해서는 주제분류와 하이브리드기반의 이용자 프로파일을 이용한 개인화 검색시스템을 구성하고 이것을 이용하여 이용자의 성향을 분석한다.

〈그림 1〉 이용자와 상호작용 개인화 검색시스템

본 연구에서 중요한 접근점은 〈그림 1〉과 같이 이용자와 상호작용하는 개인화 검색시스템을 제공하는 것이다. 즉 이용자의 키워드나 클릭 등의 행위를 관찰, 분석, 추론하여 이용자에게 적합한 개인화된 검색결과와 다양한 추천시스템들을 제공하고 다시 이용자의 행위정보를 받아서 분석하여 검색결과에 지속적으로 반영한다. 이용자의 키워드들은 입력 데이터 값이 적기 때문에 콘텐츠와 분류기의 성능에 따라 상이한 결과를 제시할 수 있으므로 부모 개념을 적용하여 네트워크기반의 프로파일을 구성한다. 또한 클릭한 문서들의 행위정보는 입력 데이터 값이 많고 이미 분류된 값을 가지고 있으므로 자식 개념을 적용하여 계층적 개념기반의 프로파일을 구성한 하이브리드기반의 이용자 프로파일을 구성하여 이용자와 상호작용하는 내용들을 프로파일에 계속 반영하여 개인화 검색시스템을 관리한다.

4.2 개인화 검색시스템 구조

본 연구에서는 주제분류 기반의 개인화 검색시스템에 최적화된 구조로 시스템을 설계하여 개발하였다.



〈그림 2〉 개인화 검색시스템 구조

〈그림 2〉와 같이 개인화 검색시스템의 구조는 일반 문서들을 적재하고 색인 및 검색할 수 있는 검색시스템과 개인의 성향정보를 기반으로 개인 성향에 맞는 문서들을 재순위화 처리를 담당하는 개인화 검색시스템, 이용자 인터페이스 및 서버와 클라이언트 간의 연결을 위한 자바 클라이언트 API와 이용자들에게 다양한 정보 제공과 쉽게 사용할 수 있는 Flex[3] UI 구조로 되어있다.

검색시스템은 다양한 콘텐츠들의 검색을 위해서 색인 DB와 원문 정보 등의 관리를 담당한다. 콘텐츠 유형별로 또는 정형화된 데이터, 반정형 데이터 및 비정형화된 데이터들을 처리할 수 있으며 다양한 콘텐츠별로 검색을 수행할 수 있도록 다양한 색인 타입 등을 제공하는 시스템이다.

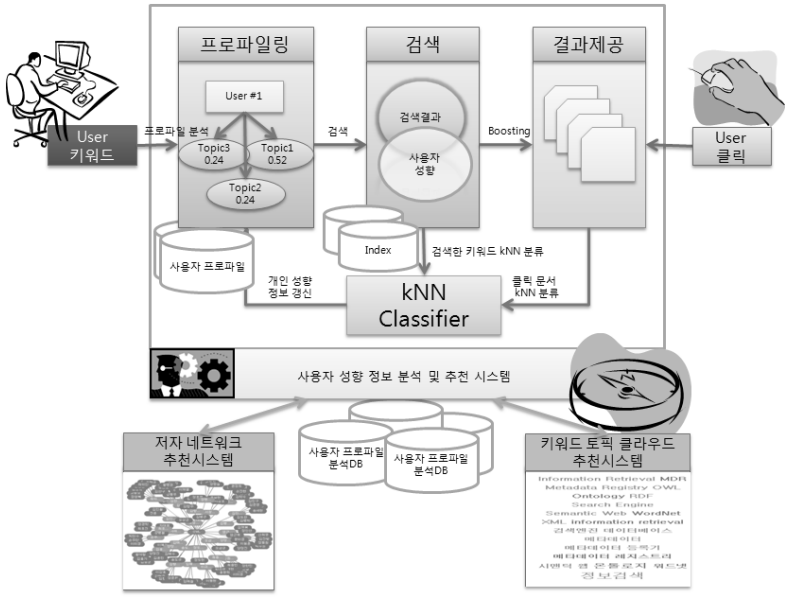
개인화 검색시스템은 검색시스템에서 제공하는 결과를 이용자의 성향 정보에 따라 재구성한다. 이를 위해서 이용자 프로파일 관리기를 사용한다. 이용자 프로파일 관리기는 기본적으로 이용자가 검색을 하기 위해 입력한 키워드와 클릭한 문서 정보를 중심으로 하이브리드기반의 프로파일을 구성한다. 키워드 및 클릭 정보를 이용하여 범주화된 각 콘텐츠 정보들을 중심으로 분류된 값도 함께 저장한다. 이 분류된 값들은 각 콘텐츠의 주제분류의 값에 따라 다를 수가 있다. 예를 들면 DDC, IPC 분류, KISTI 표준분류 등의 콘텐츠 정보에 따라 달라질 수가 있다.

자바 API는 이용자의 UI 정보와 개인화 검색시스템 간의 연결을 담당한다. 직접 검색시스템을 수행할 수도 있고 개인화 검색시스템과 연동하여 처리할 수 있다. 이용자의 다양한 인터페이스 서비스 처리를 위해서 Flex와 직접 연동하여 수행한다. 이용자 인터페이스의 다양한 서비스를 지원하기 위해서 직접적 개인화 검색시스템과 연동한다.

Flex 인터페이스는 자바 API를 이용하여 개인화에 적합한 인터페이스 구성을 담당하게 된다. Flex는 웹 브라우저에 독립적으로 동작을 하며 자바 API를 이용하여 개인화 검색시스템을 위한 시각화를 담당한다. Flex 이용자 인터페이스는 개인에게 필요한 키워드 클라우드, 검색결과의 2차원, 논문 저자 토픽 클라우드 추천, 저자 내비게이션 등의 기능들을 이용하여 시각화 표현을 할 수 있도록 설계한다.

프로세스별 개인화 검색시스템은 〈그림 3〉과 같다. 이용자가 검색을 하기 위해서 키워드나 클릭한 문서들에 대해서 첫 번째로 프로파일 정보를 분석 처리한다. 두 번째는 이용자의 성향 정보를 기반으로 개인화 검색을 수행하고 실제 검색한 키워드는 kNN분류기를 이용하여 주제분류 후 이용자 프로파일에 반영한다. 마지막으로 개인화 검색결과는 이용자 성향 정보를 반영하여 검색한 결과 값을 재순위화 처리하여 검색결과 순위를 재조정한다. 이와 같이 이용자가 검색한 결과들을 중심으로 저자들 간의 네트워크 형태를 제공함으로써 이용자가 찾고자하는 관련 저자들을 쉽게 찾을 수 있고 해당 저자를 클릭할 경우에 이용자의 프로파일에 다시 반영된다. 또한 저자 키워드 토픽 클라우드 추천시스템정보를 제공함으로써 이용자가 찾고자 하는 관련 주제어들을 쉽게 찾을 수 있도록 추천하며 해당 주제어 정보를 클릭할 경우에 이용자의 프로파일에 다시 반영된다.

3) 생산성이 뛰어난 무료 오픈 소스 프레임워크로 Adobe Flash Player와 Adobe AIR 런타임을 활용하여 모든 주요 브라우저, 데스크톱 및 운영 체제에 일관되게 배포할 수 있는 웹 애플리케이션을 구축하고 유지 관리할 수 있음. Flex 애플리케이션은 Flex 프레임워크만을 사용해서 구축할 수 있지만 Adobe Flash Builder™(이전 Adobe Flex® Builder™) 소프트웨어를 사용하면 지능적 코딩, 인터랙티브한 단계별 디버깅 및 유저 인터페이스 레이아웃의 시각적 디자인과 같은 기능을 통해 개발 시간을 단축할 수 있음.



〈그림 3〉 프로세스별 개인화 검색시스템 흐름도

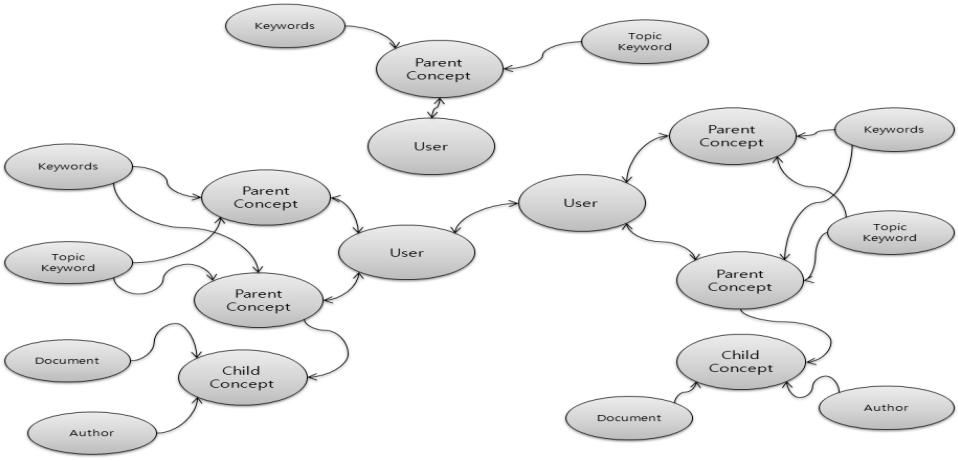
4.3 개인화 검색시스템 기술

4.3.1 하이브리드기반의 이용자 프로파일

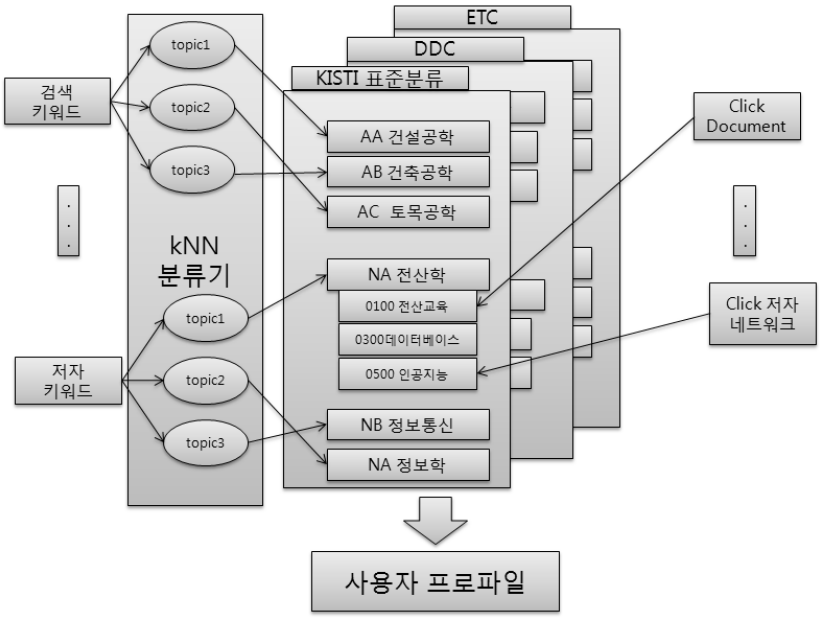
본 연구에서 제시한 하이브리드기반의 이용자 프로파일의 구성은 〈그림 4〉와 같다.

개인화 검색시스템은 이용자의 행위 중에서 검색 키워드들과 저자 토픽 키워드를 클릭한 경우에는 부모개념을 중심으로 주제분류를 수행한다. 이용자의 검색 키워드 및 저자 토픽 키워드 클라우드 정보는 입력 데이터 값이 적기 때문에 콘텐츠와 분류기의 성능에 따라 상이한 결과를 제시할 수 있으므로 부모 개념을 적용한다. 즉 보다 넓은 개념으로 이용자의 성향 정보를 접근하는 방식이다. 따라서 부모 개념으로 분류된 여러 개의 값들을 이용하여 다양한 주제 개념을 적용할 수 있는 네트워크기반으로 접근한다. 예를 들면 이용자가 "바이러스"라고 검색 키워드를 입력할 경우에는 "바이러스"는 다양한 분야에서 사용되는 용어로 검색결과 문서들을 중심으로 포괄적인 개념으로 분류한다. 즉 의학/생물/컴퓨터 등과 같이 포괄적인 개념으로 네트워크기반의 프로파일로 구성한다. 그러나 검색결과 문서들 중에 특정 문서를 클릭할 경우에는 그 문서는 이미 분류가 되어 있으므로 좁은 범위의 자식 개념의 분류체계를 적용할 수가 있으며 분류된 값이 없을 경우에는 주제분류 후에 적용한다. 이와 같이 좁은 범위로 축소함으로써 향후 이용자가 자주 보는 문서, 클릭한 문서들에 대한 정확한 추천시스템을 제공할 수가 있다.

이용자의 검색 키워드와 클릭한 문서들과 같이 이용자의 행위 정보를 이용하여 이용자 프로파일을 구성하는 방법은 〈그림 5〉와 같다. 키워드의 입력 정보는 이용자가 검색한 키워드들



〈그림 4〉 하이브리드기반의 이용자 프로파일



〈그림 5〉 주제분류 기반 이용자 프로파일 구성

에 대해서 kNN분류기를 이용하여 주제분류한다. 이렇게 분류된 주제 정보는 이용자 프로파일에 저장된다. 또한 이용자가 개인화 검색결과 중에서 특정 문서를 클릭한 경우에 주제분류 정보를 이용하여 이용자 프로파일 정보를 구성한다.

만약에 해당 콘텐츠의 분류체계가 계층적인 구조를 가지지 않을 경우에는 위와 같은 접근방식을 사용할 수가 없고 똑같은 주제범위로 할당한다.

4.3.2 주제분류 및 개인화 가중치

본 연구에서 제안한 주제분류 기반의 개인화 검색시스템을 위한 이용자 프로파일 구성은 KDC, DDC 등의 다양한 분류체계를 이용하여 이용자 프로파일을 구성할 수가 있다. 예를 들면 KISTI에서 운영하는 학회마을의 국내 논문들을 이용하여 하이브리드 프로파일을 구성하기 위해서는 KISTI 표준분류체계를 적용할 수가 있다. KISTI 분류체계는 대분류 37개, 중분류 217개, 소분류 904개로 〈표 7〉과 같이 구성된다. 이와 같이 KISTI 표준분류체계를 이용하여 이용자 프로파일을 구성할 경우에는 대/중/소분류를 활용하여 구성할 수가 있다.

〈표 8〉은 이용자가 입력한 키워드를 중심으로 KISTI 표준분류체계에서 중분류를 중심으로 분류를 수행하여 이용자 프로파일을 구성한 것이다. 빈도는 이용자가 입력한 키워드나 클릭한 문서 빈도를 나타낸다. 〈표 8〉에서는 키워드 빈도 또는 가중치 중심으로 가장 높은 분류값들을 선택하여 이용자 성향 정보로 판단하여 개인화 검색결과에 반영한다. 이와 같이 분류정보 값은 키워드 빈도와 분류기에서 분류된 가중치를 함께 계산한 후 같은 분류 정보 값으로 병합하여 최대값을 구하여 이용자 성향 정보를 분별하는데 사용한다.

아래의 〈수식 1〉은 주제분류별 빈도에 따른 가중치를 계산하는 수식이며 w값은 분류기에서 분류된 가중치 값을 나타내며 분류된 값들 중에서 가중치가 높은 값들을 선택한다. 이후 같은 분류 값들은 합쳐서 전체 이용자 주제별 프로파일 정보를 구성한다. 그 중에서 최대값을 선정하여 정규화한 후 이용자의 성향 정보로 추론하여 개인화 검색결과에 반영한다. 실제 몇 개를 선정했을 때 가장 좋은 결과가 나오는지는 실

〈표 7〉 KISTI 표준분류체계

분류단위	코드체계	비고
대분류	(영문2자리) - 총 37개 AA, AB, AC, AD, AE, AF, BA, BB, BC, BD, BE, BF, BG, BH, BL, ...	- 첫 자리에 의미부여
중분류	(숫자) 2자리 01~99	
소분류	(숫자) 2자리 01~99	
분류사례	(대분류) ND: 정보가공 (중분류) ND09: 데이터베이스 및 DBMS 일반 (소분류) ND0903: DBMS	

〈표 8〉 KISTI 표준분류를 이용한 이용자 키워드

키워드	KISTI 표준 대분류	빈도(TF)	분류가중치
암	BM:의학	19	0.483
	BF:식품		0.121
	BD:약학		0.112
항암화학요법	BM:의학	13	0.194
	BD:약학		0.016
...	...	...	...

험을 통해서 그 개수를 조정할 수도 있고 분류 체계, 콘텐츠 등에 따라 달라질 수도 있다.

〈수식 1〉 $Class(C_i) = \frac{tf_i}{\sum_{k=1}^{n} tf_k} \times w_i$

검색된 문서들에 대해서 기본 가중치 계산은 벡터 공간 모델의 가중치를 사용하며 〈수식 3〉과 같다. 즉 일반검색에서 사용되는 벡터 가중치로 검색한 키워드별로 문서에 대한 가중치를 TF-IDF 값을 이용하여 계산한다. a값은 〈수식 1〉과 같이 이용자의 전체 주제별 가중치를 계산한 값이다. 같은 주제별 가중치 값은 병합한 값 중에서 최대값으로 정렬을 한 후에 각 주제별로 높은 값들을 선택하고 전체 가중치는 1.0을 넘지 않도록 정규화작업을 수행한다. β값은 〈수식 2〉와 같이 a값으로 선정된 주제분류(C_i)에 포함되는 용어(t)에 대해서 가중치가 높은 값을 선정한다. 용어의 빈도(TF) 값들은 이용자가 입력하는 단어의 빈도수가 높아질수록 자주 사용하는 키워드들로 반영된다.

최종 문서 가중치는 〈수식 4〉와 같이 이용자의 주제분류 a값, 해당 주제에서 자주 사용하는 키워드 가중치 β값과 일반검색에서 사용하는 벡터 가중치 값을 합산하여 처리한다.

〈수식 2〉 $Weight_{i,t \in C} = \frac{tf_i}{\sum_{k=1}^{n} tf_k}$

〈수식 3〉 $w_{t,d} = tf_t \log \frac{|D|}{|t \in d|}$

〈수식 4〉 $Weight_{total}(d) = \frac{(\alpha + \beta + w_{t,d})}{3}$

$0 < \alpha \leq 1.0, 0 < \beta \leq 1.0$

최종 문서 가중치는 주제분류 a값(0~1.0), 해당 주제에서 가장 자주 출현하는 키워드 가중치 β값(0~1.0)과 벡터 모델의 가중치 값(0~1.0)을 합산 후 전체를 3로 나누어 전체 0~1.0의 가중치를 가지도록 한다.

5. 연구 결과 분석 및 평가

본 연구에서는 개인화 검색시스템의 성능을 평가하기 위해서 KISTI가 서비스 중인 과학기술학회마을에서 로그데이터를 수집하고 수집된 로그데이터를 기반으로 실제 이용자의 프로파일을 구성하였다. 구성된 프로파일을 기반으로 중의성 및 일반 키워드 질의 셋을 이용하여 전문가 10명이 직접 평가하고 실험을 수행하였다.

5.1 전문가 적합성 측정 결과

5.1.1 중의성 단어에 대한 컴퓨터관련 이용자 측정 결과

본 연구에서는 각 질의별로 상위 10위까지가 정답문서라고 가정하고 〈표 9〉와 같이 역순위 평균값을 측정하였다. 평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 MRR은 0.925이고 K-일반시스템의 MRR은 0.285로 K-개인화시스템이 더 높은 성능을 보였다. 즉 K-개인화시스템이 K-일반시스템보다 0.64로 더 우수한 것으로 나타났다. 평균 5인의 변화율(%)은 K-개인화시스템이 K-일반시스템보다 224.5% 더 높게 측정되었다.

〈표 10〉은 상위 50위 문서에 대한 정확률을 측정한 결과이다. 평가자 5인이 5개의 질의에

〈표 9〉 컴퓨터관련 상위 10위까지 중의성 키워드 MRR값

10위 MRR	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
바이러스	1	0.333	1	0.333	1	0.333	0.977	0.333	1	0.333
키	1	0.989	1	0.989	1	0.989	0.989	0.989	1	1
IR	0.654	0	0.654	0	0.681	0	0.808	0	0.775	0
cell	0.947	0	0.947	0	0.878	0	0.878	0	0.963	0
백신	1	0.125	1	0	1	0	1	0.125	0.977	0.250
MRR	0.920	0.289	0.920	0.264	0.912	0.264	0.930	0.289	0.943	0.317
향상률	0.631		0.656		0.648		0.641		0.626	
변화율(%)	218.3		248.4		245.4		221.7		197.4	
5인 평균 MRR	K-개인화시스템		0.925		K-일반시스템		0.285			
5인 평균 향상률	0.640									
5인 평균 변화율(%)	224.5									

〈표 10〉 컴퓨터관련 중의성 키워드 상위 50에 대한 정확률

50위 정확률	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
바이러스	0.780	0.060	0.840	0.060	0.960	0.060	0.900	0.060	0.920	0.060
키	1	0.820	1	0.740	1	0.9	0.960	0.840	0.960	0.800
IR	0.560	0	0.440	0	0.260	0	0.320	0	0.260	0
cell	0.980	0.040	0.760	0	0.920	0	0.900	0	0.920	0
백신	0.280	0.300	0.280	0.220	0.320	0.280	0.300	0.280	0.340	0.280
평균 정확률	0.720	0.244	0.664	0.204	0.692	0.248	0.676	0.236	0.680	0.228
향상률	0.476		0.460		0.444		0.440		0.452	
변화율(%)	195.0		225.4		179.0		186.4		198.2	
5인 평균 정확률	K-개인화시스템		0.686		K-일반시스템		0.232			
5인 평균 향상률	0.454									
5인 평균 변화율(%)	195.6									

대한 전체 K-개인화시스템의 정확률은 0.686이고 K-일반시스템의 정확률은 0.232로 K-개인화시스템이 더 높은 정확률을 보였다. 즉 K-개인화시스템이 K-일반시스템보다 0.454로 더 우수한 것으로 나타났다. 5인의 평균 변화율(%)은 K-개인화시스템이 K-일반시스템보다 195.6% 더 높게 측정되었다.

5.1.2 일반 단어에 대한 컴퓨터관련 이용자 측정 결과

본 연구에서는 각 질의별로 상위 10위까지가 정답문서라고 가정할 경우에 〈표 11〉과 같이 역순위평균값(MRR)을 측정하였다.

평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 MRR은 0.955이고 K-일반시스템의

〈표 11〉 컴퓨터관련 상위 10위까지 일반 키워드 MRR값

10위 MRR	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
K#1	0.989	0.276	1	0.243	0.880	0.693	0.989	0.500	1	0
K#2	0.875	0.267	1	0.786	1	0.989	1	1	1	0.643
K#3	1	0.938	0.928	0	1	0.431	1	0.431	1	0.664
K#4	1	0	1	0.697	1	0.427	1	1	1	0.236
K#5	0.841	0.486	1	0.250	1	0	1	0.629	0.381	0
MRR	0.941	0.393	0.986	0.395	0.976	0.508	0.998	0.712	0.876	0.308
향상률	0.548		0.591		0.468		0.286		0.568	
변화율(%)	139.4		149.6		92.1		40.1		184.4	
5인 평균 MRR	K-개인화시스템		0.955		K-일반시스템		0.463			
5인 평균 향상률	0.492									
5인 평균 변화율(%)	106.2									

MRR은 0.463로 K-개인화시스템의 성능이 더 높게 나타났다. 즉 K-개인화시스템이 K-일반시스템보다 0.492로 더 우수한 것으로 나타났다. 평균 5인의 변화율(%)은 K-개인화시스템이 K-일반시스템보다 106.2% 더 높게 측정되었다.

〈표 12〉는 상위 50위 문서에 대한 정확률을 측정한 결과이다. 평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 정확률은 0.84이고 K-일반시스템의 정확률은 0.434로 K-개인화시스템의 성능이 더 높게 나타났다. 즉 K-개인화시스템이 K-일반시스템보다 0.406로 더 우수한 것으로 나타났다. 5인의 평균 변화율(%)은 K-

〈표 12〉 컴퓨터관련 일반 키워드 상위 50에 대한 정확률

50위 정확률	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
K#1	0.980	0.300	1	0.160	0.480	0.360	0.740	0.700	0.740	0
K#2	0.480	0.160	0.62	0.620	1	0	1	1	1	0.920
K#3	0.980	0.440	0.62	0.020	1	0.820	0.980	0	0.980	0.900
K#4	0.640	0.160	0.780	0.440	1	0.260	1	0.960	1	0.44
K#5	0.800	0.400	0.800	0.660	0.940	0.080	1	0.940	0.440	0.120
평균 정확률	0.776	0.292	0.764	0.38	0.884	0.304	0.944	0.72	0.832	0.476
향상률	0.484		0.384		0.580		0.224		0.356	
변화율(%)	165.7		101.0		190.7		31.1		74.7	
5인 평균 정확률	K-개인화시스템		0.840		K-일반시스템		0.434			
5인 평균 향상률	0.406									
5인 평균 변화율(%)	93.5									

개인화시스템이 K-일반시스템보다 93.5% 더 높게 측정되었다.

5.1.3 중의성 단어에 대한 문헌정보학관련 이용자 측정 결과

본 연구에서는 각 질의별로 상위 10위까지가 정답문서라고 가정할 경우에 〈표 13〉과 같이 MRR값 측정하였다. 평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 MRR은 0.806이고 K-일반시스템의 MRR은 0.211로 K-개인화시스템의 성능이 더 높게 나타났다. 즉 K-개인화시스템이 K-일반시스템보다 0.595로 더 우수한 것으로 나타났다. 평균 5인의 변화율(%)은 K-개인화시스템이 K-일반시스템보다 281.9% 더 높게 측정되었다.

〈표 14〉는 상위 50위 문서에 대한 정확률을 측정한 결과이다. 평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 정확률은 0.3976이고 K-일반시스템의 정확률은 0.0536로 K-개인화시스템의 성능이 더 높게 나타났다. 즉 K-개인화시스템이 K-일반시스템보다 0.344로 더 우수한 것으로 나타났다. 5인의 평균 변화율(%)은 K-개인화시스템이 K-일반시스템보다 641.7% 더 높게 측정되었다.

5.1.4 일반 단어에 대한 문헌정보학관련 이용자 측정 결과

본 연구에서는 각 질의별로 상위 10위까지가 정답문서라고 가정할 경우에 〈표 15〉와 같이 MRR값을 측정하였다.

평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 MRR은 0.901이고 K-일반시스템의 MRR은 0.36로 K-개인화시스템의 성능이 더 높게 나타났다. 즉 K-개인화시스템이 K-일반시스템보다 0.541로 더 우수한 것으로 나타났다. 평균 5인의 변화율(%)은 K-개인화시스템이 K-일반시스템보다 150.3% 더 높게 측정되었다.

〈표 13〉 문헌정보학관련 상위 10위까지 중의성 키워드 MRR값

10위 MRR	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
수서	1	0.878	1	0.878	1	0.878	0.953	0.700	0.077	0.037
판	1	0	0.100	0	0.100	0	0.645	0	0.082	0.001
기술	1	0	1	0	1	0	0.933	0	0.035	0
LCC	0.558	0	0.558	0	0.558	0	0.833	0	0.060	0
LC	0.944	0.167	0.944	0.167	0.944	0.167	0.659	0	0.055	0
MRR	0.901	0.209	0.721	0.209	0.721	0.209	0.805	0.14	0.884	0.286
향상률	0.692		0.512		0.512		0.665		0.598	
변화율(%)	331.1		244.9		244.9		475.0		209.0	
5인 평균 MRR	K-개인화시스템		0.806		K-일반시스템		0.211			
5인 평균 향상률	0.595									
5인 평균 변화율(%)	281.9									

〈표 14〉 문헌정보학관련 중의성 키워드 상위 50에 대한 정확률

50위 정확률	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
수서	0.240	0.160	0.620	0.280	0.480	0.220	0.560	0.220	0.620	0.240
판	0.660	0.020	0.700	0.060	0.560	0.020	0.520	0.020	0.780	0.020
기술	0.420	0	0.560	0	0.400	0	0.220	0	0.400	0
LCC	0.120	0.020	0.240	0	0.300	0	0.220	0	0.320	0
LC	0.160	0.020	0.20	0.040	0.200	0	0.220	0	0.220	0
평균 정확률	0.320	0.044	0.464	0.076	0.388	0.048	0.348	0.048	0.468	0.052
향상률	0.276		0.388		0.34		0.3		0.416	
변화율(%)	627.2		510.5		708.3		625.0		800.0	
5인 평균 정확률	K-개인화시스템		0.3976		K-일반시스템		0.0536			
5인 평균 향상률	0.344									
5인 평균 변화율(%)	641.7									

〈표 15〉 문헌정보학관련 상위 10위까지 일반 키워드 MRR값

10위 MRR	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
K#1	1	1	0.963	0.393	0.917	0.111	0.878	0	1	1
K#2	1	0.604	1	0	0.538	0.380	0.833	0.1	1	0
K#3	0.833	0	0.789	0.760	1	0.216	1	0.222	1	0
K#4	1	0.947	0.963	0.394	0.716	0	1	0	1	0
K#5	1	0.888	0.787	0.393	0.375	0.125	1	1	0.928	0.477
MRR	0.967	0.688	0.900	0.388	0.709	0.166	0.942	0.264	0.986	0.295
향상률	0.279		0.512		0.543		0.678		0.691	
변화율(%)	40.6		132.0		327.1		256.8		234.2	
5인 평균 MRR	K-개인화시스템		0.901		K-일반시스템		0.36			
5인 평균 향상률	0.541									
5인 평균 변화율(%)	150.3									

〈표 16〉은 상위 50위 문서에 대한 정확률을 측정한 결과이다. 평가자 5인이 5개의 질의에 대한 전체 K-개인화시스템의 정확률은 0.589이고 K-일반시스템의 정확률은 0.215로 K-개인화시스템의 성능이 더 높게 나타났다. 즉 K-개인화시스템이 K-일반시스템보다 0.374로 더 우수한 것으로 나타났다. 5인의 평균 변화율(%)은 K-개인화시스템이 K-일반시스템보다 174.0% 더 높게 측정되었다.

〈표 16〉 문헌정보학관련 일반 키워드 상위 50에 대한 정확률

50위 정확률	평가자1		평가자2		평가자3		평가자4		평가자5	
	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템	K-개인화 시스템	K-일반 시스템
K#1	0.920	0.280	0.600	0.100	0.380	0.100	0.680	0.580	1	0.020
K#2	0.980	0.460	0.700	0.020	0.340	0.160	0.140	0.140	0.060	0
K#3	0.120	0.120	0.440	0.360	0.800	0.160	1	0.300	1	0
K#4	0.960	0.820	0.840	0.140	0.360	0	0.640	0.040	0.100	0
K#5	1	0.340	0.280	0.160	0.040	0.080	0.660	0.320	0.680	0.680
평균 정확률	0.796	0.404	0.572	0.156	0.384	0.1	0.624	0.276	0.568	0.14
향상률	0.392		0.416		0.284		0.348		0.428	
변화율(%)	97.0		266.7		284.0		126.1		305.7	
5인 평균 정확률	K-개인화시스템		0.589		K-일반시스템		0.215			
5인 평균 향상률	0.374									
5인 평균 변화율(%)	174.0									

5.2 전문가 질적 평가 분석 결과

본 연구에서는 K-개인화시스템과 K-일반시스템의 선호도 차이를 비교 분석하기 위해 전문가들이 테스트 후 심층 면담조사를 하였다. 수집된 질적 자료에 대한 내용 분석은 Lincoln과 Guba(1985)가 제시한 단위화와 부호화 과정을 거쳐 범주화 분석방법을 사용하였다.

본 연구자가 면담 내용을 반복하여 읽으면서 줄, 문단/문장, 전체자료 의 분석을 실시하여 의미단위를 찾아내어 번호를 부여하였고, 유사한 내용을 모아 공통된 내용으로 개념화한 '좋은 점'과 '불편한 점' 등 2개의 영역으로 하위 범주화하였다. 다시 하위범주를 재분류하고 추상화하여 학습의 용이성, 검색기능의 유용성, 검색결과의 만족도 등 3개의 범주로 구성하였다.

K-개인화시스템과 K-일반시스템의 선호도 차이에 대한 피면담자들의 면담 내용은 〈표 17〉과 같이 'K-개인화시스템과 K-일반시스템의 사용성 경험 유형'으로 범주화되었다.

5.2.1 학습의 용이성

학습의 용이성은 K-개인화시스템과 K-일반시스템의 사용법을 얼마나 쉽게 배울 수 있는지 면담조사를 통해서 분석하였다. K-개인화시스템과 K-일반시스템에 대한 학습의 용이성에서 사용성 경험 유형을 살펴보면 K-개인화시스템은 '개인화 검색시스템이지만 일반 검색시스템과 비슷하고 친숙하여' 좋았다는 긍정적 평가와 '다양한 추천시스템에 대한 기능 설명 부족하여' 불편했다는 부정적 평가를 나타냈다. K-일반시스템은 '일반적인 정보서비스 형태임으로 이용하기에 친숙하여' 좋았다는 긍정적 평가와 '검색결과에 초기 화면에 초록 요약 정보가 없어 내용 파악 어려워서' 불편했다는 부정적 평가를 나타냈다.

피면담자는 K-개인화시스템과 K-일반시스템의 인터페이스가 각각 상이하지만, 두 시스

템의 사용법이 일반 웹 검색시스템들과 비교적 유사하여 쉽게 사용된 것으로 나타났다. 시스템은 이용자의 수많은 경험을 기반으로 별도의 학습이 없이도 곧바로 사용이 가능한 직관적으로 인지할 수 있는 인터페이스 설계가 중요한 것으로 분석되었다.

5.2.2 검색기능의 유용성

검색기능의 유용성은 K-개인화시스템과 K-일반시스템에서 검색결과를 제공하는 기능이 얼마나 유용한지 면담조사를 통해서 분석하였다. K-개인화시스템과 K-일반시스템에 대한 검색기능의 유용성에서 사용성 경험 유형을 살펴보면 K-개인화시스템은 '개인화 검색결과와 관련된 연구논문 주제 및 연구 저자 등 다양한 맞춤 정보를 제공함으로 정보 획득이 용이하여' 좋았다는 긍정적 평가와 '개인화 검색시스템을 사용하기 위해서 로그인해야 하는 것이' 불편했다는 부정적 평가를 나타냈다. K-일반시스템은 '논문, 학술지 및 학회별로 그룹별의 결과를 제시하여' 좋았다는 긍정적 평가와 '키워드 클라우드 등과 같은 추천 기능 없어서' 불편했다는 부정적 평가를 나타냈다.

피면담자는 다양한 검색 의도에 적합한 정보를 빠르게 찾을 수 있기를 원하고 있고, 다양한 추천시스템을 요구하며 검색결과를 그룹화하여 정리된 것을 선호하는 것으로 분석되었다.

5.2.3 검색결과의 만족도

검색결과의 만족도는 K-개인화시스템과 K-일반시스템에서 검색질문의 검색결과에 얼마나 만족하는지 면담조사를 통해서 분석하였다. K-개인화시스템과 K-일반시스템에 대한 검색결과의 만족도에서 사용성 경험 유형을 살펴보면 K-개인화시스템은 '관심정보를 우선 검색결과를 제공함으로 정보탐색시간 감소하여', '중의적인 의미를 가진 키워드들에 대한 분별력이 좋다'와 '관심정보를 직접 입력하지 않고 자동으로 나의 관심분야 분석/추론함으로 편리하여' 좋았다는 긍정적 평가와 '검색한 키워드, 클릭한 문서, 토픽 클라우드 등 전체적으로 개인의 성향을 반영되는 것과 세부분야 개인의 성향 정보가 어떻게 반영이 되는지 표시가 없어서' 불편했다는 부정적 평가를 나타냈다. K-일반시스템은 '저자, 학회지 정보, 원문정보 등 상세한 정보를 충분히 제공하여' 좋았다는 긍정적 평가와 '원하는 정보 찾기 위해서 다양한 키워드들을 조합해야 하므로 검색시간이 많이 걸리고', '다양한 분야별 문서들이 중첩되어 검색되기 때문에 원하는 정보를 찾기가 어려워서' 불편했다는 부정적 평가를 나타냈다.

피면담자는 K-개인화시스템을 사용하면서 자신이 원하는 정보를 몇 번 검색한 후에 자신의 성향 정보 맞는 검색결과와 다양한 추천시스템들을 활용하여 새로운 정보를 획득함으로써 정보 획득과 정보 탐색 시간이 효율적인 것으로 나타났다. 피면담자는 K-일반시스템을 사용하면서 자신이 원하는 정보를 탐색하기 위해서는 많은 키워드들을 조합해야 하며 중첩된 검색결과를 필터링하기 위해서 결과 재검색, 소팅 등의 방법을 활용해야 하므로 정보 탐색시간을 많이 소비하는 것으로 분석되었다. 따라서 이용자의 자신의 성향에 맞는 적합한 문서들을 제공하고 다양한 추천시스템을 제공하는 것은 이용자에게 정보 획득과 정보 탐색 시간에 매우 효율적이다.

〈표 17〉 K-개인화시스템과 K-일반시스템의 사용성 경험 유형

구분	범주	하위범주	의미단위	빈도
K-개인화 시스템	학습의 용이성	좋은 점	• 개인화 검색시스템이지만 일반 검색시스템과 비슷하고 친숙함	3
		불편한 점	• 다양한 추천시스템에 대한 기능 설명 부족	2
	검색기능의 유용성	좋은 점	• 개인화 검색결과와 관련된 연구논문 주제 및 연구 저자 등 다양한 맞춤 정보를 제공함으로 정보 획득이 용이함	6
		불편한 점	• 개인화 검색시스템을 사용하기 위해서 로그인해야함	2
	검색결과의 만족도	좋은 점	• 관심정보를 우선 검색결과를 제공함으로 정보탐색시간 감소함	9
			• 중의적인 의미를 가진 키워드들에 대한 분별력이 좋음	6
			• 관심정보를 직접 입력하지 않고 자동으로 나의 관심분야 분석/추론함으로 편리함	5
		불편한 점	• 검색한 키워드, 클릭한 문서, 토픽 클라우드 등 전체적으로 개인의 성향을 반영되는 것과 세부분야 개인의 성향 정보가 어떻게 반영이 되는지 표시가 없음	3
K-일반 시스템	학습의 용이성	좋은 점	• 일반적인 정보서비스 형태임으로 이용하기에 친숙함	2
		불편한 점	• 검색결과에 초기 화면에 초록 요약 정보가 없어 내용 파악 어려움	3
	검색기능의 유용성	좋은 점	• 논문, 학술지 및 학회별로 그룹별의 결과를 제시함	4
		불편한 점	• 키워드 클라우드 등과 같은 추천 기능 없음	5
	검색결과의 만족도	좋은 점	• 저자, 학회지 정보, 원문정보 등 상세한 정보를 충분히 제공함	4
		불편한 점	• 원하는 정보 찾기 위해서 다양한 키워드들을 조합해야함으로 검색시간이 많이 걸림	9
			• 다양한 분야별 문서들이 중첩되어 검색되기 때문에 원하는 정보를 찾기가 어려움	5

6. 결론 및 제언

본 연구에서는 도서관 및 학술정보센터의 학술정보 이용자를 위해서 자신의 성향 정보에 맞는 정확한 검색결과를 제공하기 위한 주제분류 기반의 개인화 검색시스템을 설계, 구현, 평가하여 효율성을 입증하였다. 이용자의 행위정보를 이용하여 개인 성향 정보를 분석고자 주제분류 방식과 하이브리드 기반의 이용자 프로파일을 구성하였다.

본 연구에서는 과학기술학회마을 논문 80여만 건을 이용하여 전문가들이 일반 키워드와 중의성을 가진 키워드들을 이용 직접 적합성을 평가하였다. 주제분류 기반 개인화 검색시스템을 선호하는 이유를 조사하고자 질적 평가를 위한 심층면담을 전문가들을 대상으로 진행하였다.

본 논문의 연구 결과를 요약하면 다음과 같다.

첫째, 전문가들이 과학기술학회마을 80여만 건에 대해서 직접 평가한 결과를 보면 "컴퓨터"와 "문헌정보학" 분야에서는 중의성을 가진 키워드들의 정확률은 개인화 검색시스템이 일반 검색시스템보다 더 높은 정확률을 나타냈으며 일반 키워드들에 대해서도 높은 정확률을 보였다.

둘째, 질적 평가인 심층면담을 통한 평가에서도 개인화 검색시스템이 이용자가 선호하는 문서들을 우선적으로 제공함으로써 정보검색

시간 및 정보 획득에 용이하다고 응답하였다. 또한 다양한 추천시스템을 제공함으로써 새로운 정보 탐색에 도움이 된다 하였으며, 개인화 검색시스템이 일반 검색시스템보다 이용자의 만족도가 높음을 알 수 있었다. 하지만 개인화 검색시스템은 이용자를 위한 시스템이므로 다양한 추천시스템에 대한 기능 설명과 직관적인 화면 설계, 검색 결과의 그룹화, 자동로그인 기능 등과 같은 이용자를 위한 편리한 기능 제공을 위한 노력이 더욱 필요로 하였다.

향후 연구과제로는 단일 분류기에서 주제분류를 잘못 분류할 경우를 고려하여 다중 분류기를 이용한 투표 방식으로 주제분류 기반의 개인화 검색시스템에 관한 연구와 고찰이 필요하다. 분류체계를 가지고 있지 않은 콘텐츠에 대해서 특정 분류체계를 활용하여 직접 주제분류 작업을 수행하는 것보다 해당 콘텐츠를 직접 클러스터링한 후에 개인화 검색시스템을 적용하는 연구가 필요하다. 대용량의 특허, 웹 데이터 등과 같은 다양한 콘텐츠에 개인화 검색시스템을 적용하여 각 콘텐츠별로 적합한 개인화 검색시스템 모델에 대해서 연구와 각 콘텐츠별로 다양한 추천시스템에 관련된 요소 연구가 필요하다. 또한 다양한 추천시스템의 정확도를 높이기 위해서 전거데이터와 같은 식별된 데이터를 기반으로 한 개인화 검색 기술 활용에 관한 연구도 필요하다.

참 고 문 헌

[1] 김경중, 조성배. 2001. 퍼지 개념 네트워크를 이용한 개인화된 링크기반 검색엔진의 개발.『정보과학회논문지』, 7(3): 211-219.

[2] 김광영, 심강섭, 곽승진. 2009. 분류와 이용자 키워드 정보에 기반한 개인화 시스템,『한국문헌정보학회지』, 43(3): 163-180.

[3] 문현전, 이수진, 김영지, 우용태. 2007. 도메인 온톨로지를 이용한 개인화된 개념기반 검색 기법.『한국전자거래학회지』, 12(3): 269-280.

[4] 박건우, 이상훈. 2008. 질의어 패턴 자동분석을 통한 커뮤니티 기반 개인화 검색.『정보과학회논문지』, 36(4): 321-326.

[5] 윤기상, 윤광호, 김재광, 이지형. 2009. 태그를 이용한 개인화 검색 시스템.『한국정보과학회 학술발표대회』, 320-324.

[6] 윤홍준, 노준호, 김한준, 이병정, 강수용, 장재영. 2010. 개념 네트워크 기반 개인화 검색 시스템의 설계 및 구현.『한국인터넷정보학회 학술발표대회』, 147-152.

[7] 이성직, 김한준, 이병정, 강수용. 2009. 개인화 웹검색을 위한 개념 네트워크 기반 이용자 프로파일 구성 기법.『한국정보과학회 학술대회』, 203-208.

[8] 허정, 서희철, 장명길. 2006. 상호정보량과 복합명사 의미사전에 기반한 동음이의어 중의성 해소. 『정보과학회논문지』, 33(12): 1073-1088.

[9] 허정, 옥철영, 2000. 사전의 뜻풀이말에서 추출한 의미정보에 기반한 동형이의어 중의성 해결 시스템. 『정보과학회논문지』, 28(9): 688-698.

[10] Carmel, David, Zwerdling, Naama, Guy, Ido, Ofek-Koifman, Shila, Harel, Nadav, Rone, Inbal, Uziel, Erel, Yogev, Siva, & Chernov, Sergey. 2009. "Personalized social search base on the user's social network." In *Proceeding of CIKM'09*, 1227-1236.

[11] Chaffee, J., & Gauch, S. 2000. "Personal ontologies for web navigation." In *Proceeding of 9th International Conference on Information Knowledge Management*, 227-234.

[12] Chen, C., Chen, M., & Sun, Y. 2002. "PVA: A self-adaptive Personaled View Agent." *Journal of Intelligent Information System*, 173-194.

[13] Chirita, P. A., Nejdl, W., Paiu, R., & Kohlschutter, C. 2005. "Using ODP metadata to personalize search." *Proceedings of the 28th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval, SIGIR*, 178-185.

[14] Daoud, Mariam, Tamine-Lechani, Lynda, & Boughanem, Mohand. 2009. "A session based personalized search using an ontological user profile." In *Proceedings of the 2009 ACM Symposium on Applied Computing*, 1732-1736.

[15] Glover, E. J., et a1. 2000. "Web Search - Your Way." *Communications of the ACM*, 44(12): 97-102.

[16] Haase, P., Hotho, A., Schmidt-Thieme, L., & Sure, Y. 2005. "Collaborative and usage-driven evolution of personalized ontologies." In *Proceeding of 2th European Semantic Web Conference*, 486-599.

[17] Jeh, G., & Widom, J. 2003. "Scaling personalized web search." *Proceedings of the 12th International Conference on World Wide Web*, 271-279.

[18] Jiang, Xing, & Tan, Ah-Hwee. 2009. "Learning and inferencing in user ontology for personalized Semantic Web search." *Information Sciences*, 179(16): 2794-2808.

[19] Liu, F., & Yu, C. 2004. "Personalized web search for improving retrieval effectiveness." *IEEE Transaction on Knowledge and Data Engineering*, 28-40.

[20] Lincoln, Yvonna S., & Guba, Egon G. 1985. *Naturalistic Inquiry*. Beverly Hills, Calif.: Sage.

[21] Middletion, S. E., Shadbolt, N. R., & De Roure, D. C. 2003. "Capturing interest through inference and visualization: Ontological user profiling in recommender system." *International Conference on Knowledge Capture*, 62-69

[22] Qiu, Feng., & Cho, Junghoo. 2006. "Automatic identification of user interest for personalized

search." *Proceeding of the WWW '06 Proceedings of the 15th International Conference on World Wide Web*, 727-736.

[23] Radlinski, F., & Dumais, S. 2006. "Improving personalized web search using result diversification." *Research and Development in Information Retrieval*, 691-692.

[24] Shahabi, Cyrus, & Chen, Yi-Shin. 2003. "Web information personalization: Challenges and approaches." In *Proceedings of 3rd Workshop on Databases in Networked Information System*, 5-15.

[25] Shen, Xuehua., Tan, Bin, & Zhai, Cheng Xiang. 2005. "Implicit user modeling for personalized search." In *Proceeding of CIKM'05*, 824-831.

[26] Sieg, Ahu, Mobasher, Bamshad, & Burke, Robin. 2007. "Web search personalization with ontological user profiles." *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, 525-534.

[27] Speretta, Mirco, & Gauch, Susan. 2005. "Personalizing search based on search histories." In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 622-628.

[28] Speretta, Mirco. 2000. *Personalized Search Based on User Search Histories*, Ph.D. diss., Udine University, Udine, Italy.

[29] Tanudjaja, F, & Mui, L. 2002. "Persona: A contextualized and personalized web search." In *Proceeding of 35" Hawaii International Conference on System Sciences*, 195-230.

[30] Trajkova, J., & Gauch, S. 2004. "Improving ontology-based user profiles." *Proceedings of the Recherched' Information Assistear Ordinateur, RIAO*, 380-389.

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

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

[1] Kim, Kyung-Joog, & Cho, Sung-Bae. 2001. "Development of a personalized link-based search engine using fuzzy concept network." *Journal of KISS: computing practices*, 7(3): 211-219.

[2] Kim, Kwang-Young, Shim, Kang-Seop, & Kwak, Seung-Jin. 2009. "A personalized retrieval system based on classification and user query." *Journal of the Korean Society for Library and Information Science*, 43(3): 163-180.

[3] Mun, Hyeon Jeong, Lee, Soo Jin, Kim, Young Ji, & Woo, Yong Tae. 2007. "A personalized concept-based retrieval technique using domain ontology." *The Journal of Society for e-Business Studies*, 12(3): 269-280.

[4] Park, Gun Woo, & Lee, Sang Hoon. 2008. "Personalized search based on community through automatic analysis of query patterns." *Journal of KISS: Databases*, 36(4): 321-326.

[5] Yoon, Ki Sang, Yoon, Kwang Ho, Kim, Jae Kwang, and Lee, Jee Hyong. 2009. "Personalized web search system using tag." *Proceeding of the Korean Institute of Information Scientists and Engineers*, 36(2): 320-324.

[6] Yune, Hong June, Noh, Joo Ho, Kim, Han-Joon, & Lee, Byung-Jeong. 2010. "Design and implementation of concept network-based personalized search systems." *Proceeding of Korean Society for Internet Information*, 6: 147-152.

[7] Lee, Sung Jick, Kim, Han-Joon, Lee, Byung-Jeong, & Kang Soo Yong. 2009. "Concept network-based user profile construction for personalized web search." *Proceeding of the Korean Institute of Information Scientists and Engineers*, 203-208.

[8] Heo, Jeong, Seo, Hee-Chel, and Jang, Myung-Gil. 2006. "Homonym disambiguation based on mutual information and sense-tagged compound noun dictionary." *Journal of KISS: Software and Application* , 33(12): 1073-1088.

[9] Heo, Jeong, & Ock, Cheol-Young. 2000. "A homonym disambiguation system based on semantic information extracted from dictionary definitions." *Journal of KISS: Software and Application*, 28(9): 688-698.