

개인 관심분야 추적기법을 이용한 과학기술정보 개인화에 관한 연구*

A Study on Personalization of Science and Technology Information by User Interest Tracking Technique

한 희 준 (Heejun Han)**
최 윤 수 (Yunsoo Choi)***
최 성 필 (Sung-Pil Choi)****

목 차

- | | |
|-------------------|------------------------|
| 1. 서 론 | 4. 개인 관심분야 추적 및 정보 개인화 |
| 2. 이론적 배경 및 관련연구 | 5. 과학기술정보 개인화 성능 평가 |
| 3. 연구의 방법 및 수행 절차 | 6. 결 론 |

초 록

본 연구의 목적은 사용자의 정보 서비스 이용행태를 분석하여 검색하는 의도와 관심분야를 국가과학기술표준분류 기반으로 파악하고 추적하며, 이를 이용해 과학기술정보를 개인화하는 것이다. 즉 과학기술정보 검색 성능을 개선하여 사용자가 원하는 정보를 탐색하는데 효율성과 만족도를 동시에 충족시키고자 하였다. 실시간 관심분야 추적, 관심 태그 클라우드 제공, 관심분야 기반 추천정보 제공, 검색 결과 개인화 네 가지 기능으로 구성된 과학기술정보 개인화 서비스를 개발하여 전문가 실험집단과 통제집단과의 검색 성능 비교를 통해 개인화 정보의 적합성 및 개인화 기능 유용성을 평가하였다. 그 결과 본 연구에서 제안된 개인화 서비스가 비교 대상 서비스보다 검색 성능이 더 우수한 것으로 나타났으며 더 높은 유용성을 제공하는 것을 입증하였다.

ABSTRACT

In this paper, we analyze a user's usage behavior, identify and track search intention and interest field based on the National Science and Technology Standard Classification, and use it to personalize science and technology information. In other words, we sought to satisfy both efficiency and satisfaction in searching for information that users want by improving scientific information search performance. We developed the personalization service of science and technology information and evaluated the suitability and usefulness of personalized information by comparing the search performance between expert experimental group and control group. As a result, the personalization service proposed in this study showed better search performance than comparative service and proved to provide higher usability.

키워드: 개인화 검색, 관심분야 추적, 과학기술정보, 검색성능 평가, 이용 행태
Personalized Search, Interest Tracking, Science and Technology Information,
Search Performance Evaluation, User Behavior

- * 본 논문은 경기대학교 대학원 박사학위 논문의 일부를 요약한 것임.
** 한국과학기술정보연구원 융합서비스센터 선임연구원(hhj@kisti.re.kr) (제1저자)
*** 경기대학교 일반대학원 문헌정보학과 박사과정(cysoo@kgu.ac.kr) (공동저자)
**** 경기대학교 휴먼인재융합대학 문헌정보학과 조교수(spchoi@kgu.ac.kr) (교신저자)
논문접수일자: 2018년 7월 16일 최초심사일자: 2018년 7월 16일 게재확정일자: 2018년 8월 10일
한국문헌정보학회지, 52(3): 5-33, 2018. [http://dx.doi.org/10.4275/KSLIS.2018.52.3.005]

1. 서론

국내 과학기술정보서비스 NTIS¹⁾은 과제, 논문, 특허, 연구보고서, 기술이전, 시설장비 정보 등을 제공하고 연구자, 학생이 폭넓게 사용하는 지식 포털이다. 500만 건 이상의 과학기술 정보에 대해 동일한 질의어에 같은 검색결과를 동일한 형태 또는 순위로 제공하는 것은 관심분야가 제각각인 사용자의 요구를 만족시키기 어렵다. 따라서 본 연구에서는 사용자 관심분야를 국가과학기술표준분류 기반으로 실시간 추적하고, 이를 이용하여 과학기술정보를 개인화하고자 한다. 여기서 개인화란 사용자가 의도하는 바를 미리 파악하여 검색결과를 최적화시켜 주는 것을 의미하며, 검색결과란 입력한 질의어에 적합한 정보뿐만 아니라 입력 의도와 관련된 추천 및 제안 정보도 포함한다. 개인화 검색은 각 사용자의 관심분야를 파악하기 위해 과거의 행위나 미리 저장된 사용자 프로파일 정보를 활용하게 되는데, 사용자 프로파일은 입력받기 어렵고 관심분야는 시간이 지나면서 검색 목적에 따라 변하는 정보이다. 또한 과학기술분야에서 중의어에 대한 검색결과 차별화는 관심분야 반영이 필수적인 문제이다. 개인화 검색과 관련하여 최근까지 이용자 프로파일 분석, 질의어 확장, 협업 필터링, 유사문서 클러스터링 기법 등을 활용한 검색결과 재조정과 정보 추천에 대한 연구는 많이 수행되어 왔다. 하지만 개인의 검색 의도 및 관심분야가 변화하는 상황을 위한 정보 개인화 연구는 부족하

다. Shuo Xu(2014)는 사용자에게 상황에 적합한 맞춤형 서비스를 제공하기 위한 중요한 문제들 중 하나는 대량의 정보들을 이용하여 잠재된 주제와 사용자의 관심분야, 그것들이 변화하는 패턴을 자동으로 감지하는 것이라고 하였다. 즉 사용자의 서비스 이용에서 발견되는 다양한 데이터를 분석하여 변하는 사용자 관심분야를 실시간으로 추적하고, 결정된 관심분야를 이용해 정보를 개인화하는 것은 과학기술정보서비스를 이용하는 많은 연구자에게 절실하고 시급한 문제이다.

본 연구의 목적은 과학기술정보서비스 사용자의 변화하는 관심분야를 추적하고 정보를 개인화하여 정보 이용 측면에서의 만족도를 개선시키는 것이다. 개인화된 검색결과와 정확성을 입증하기 위해 NTIS와의 검색 성능을 비교하여 그 효과를 검증하고 제안된 개인화 서비스의 질적 평가를 수행하였다.

2. 이론적 배경 및 관련연구

2.1 NTIS 개요

한국과학기술정보연구원(KISTI)이 구축·운영하는 NTIS는 사업, 과제, 인력, 연구시설장비, 성과 등 국가연구개발 사업에 대한 과학기술정보를 한 곳에서 서비스하는 세계 최초의 국가R&D정보 지식포털이다. 부처별 국가R&D정보를 총괄하는 대표전문기관, 성과물전담기관,

1) NTIS(National Science & Technology Information Service)는 한국과학기술정보연구원에서 구축·운영하는 국가과학기술지식정보서비스로 사업, 과제, 인력, 연구시설장비, 성과 등 국가연구개발사업에 대한 정보를 제공하는 세계 최초의 국가R&D정보 지식포털이다.

과제관리기관, 출연연 등으로부터 422개 정보 표준을 수집, 가공하여 제공하는데 이는 대국민, 연구원, 기업 대상의 정보전달 기능과 함께 R&D사업조사분석, 예산분배조정, 온라인 평가 등 범부처 활용 기능을 마련하고 있다. NTIS는 사업, 과제, 인력, 연구시설·장비, 성과 정보 등 국가R&D 관련 과학기술정보를 실시간으로 제공하여 과제 기획에서 성과관리, 성과확산, 평가까지 R&D 전주기를 지원하고 과학기술자원에 대한 국가적 활용도 제고를 통해 국가 과학기술가치 극대화를 추구하고 있다. 또한 국가R&D 예산으로 연구과제를 수행하고자 하는 사람은 의무적으로 NTIS를 통해 유사 과제여부를 점검하여 제출해야 하는 등 과제 기획 및 실행 단계에서 학생, 연구자, 기업 등에서 폭넓게 정보를 활용하고 있다.

과학기술정보의 관리 및 유통 효율화를 위해 정부는 연구자가 국가R&D 과제를 기획하여 계획서를 제출할 때와 연구 종료 후 연구보고서를 각 대표전문기관에 제출할 때 국가과학기술표준분류를 부여하도록 규정한다. NTIS에서 수집되는 사업, 과제, 성과, 연구시설·장비 정보에는 국가과학기술표준분류코드와 코드명이 부여되어 있는데, 이는 국가연구개발사업의 연구기획·평가 및 관리, 과학기술예측 및 기술수준평가, 과학기술지식 정보의 관리·유통을 분류 기반으로 효율적으로 수행하기 위함이다. NTIS에서는 국가과학기술표준분류별 연구동향 및 투자현황 파악을 용이하게 하기 위해 국가과학기술표준분류별 과학기술정보를 군집화(clustering)하여 제공한다. 즉, 하나의 중분류를 기준으로 어느 특정기간에 수행한 과제정보와 해당 과제에서 산출된 논문, 특허 등 성과정

보를 파악할 수 있으며, 이는 과학기술정보 분류의 기준으로 활용된다는 의미이다.

2.2 개인화 검색

정보의 과잉 속에서 사용자들은 원하는 정보를 빠른 시간에 정확하게 찾기 위해 많은 노력을 기울이고 있다. 사용자가 선호하는 또는 사용자 관심에 부합하는 정보만을 선별적으로 제공하기 위한 접근 방법 중 하나로 개인화 시스템이 등장하였다. 대부분의 개인화는 사용자 프로파일을 이용하여 맞춤 검색 또는 추천 기법으로 정보를 제공하고 있다. 사용자 프로파일은 일반적으로 통계정보를 포함하는데, 예를 들면 사용자 개인의 관심분야 선호도를 나타내는 성별, 연령대, 국가, 소속기관, 전공분야 등과 같은 요소들이 이용된다. 하지만 요즘 대부분의 웹서비스는 사용자의 관심정보 파악이 가능한 프로파일 정보를 입력받지 않을 뿐만 아니라 사용자 또한 구체적인 프로파일 정보를 입력하지 않기 때문에 최소한의 개인정보 만으로 정보 및 제품을 이용할 수 있도록 한다. 반면, 사용자의 정보 또는 제품 이용 행태를 파악하여 실시간으로 개인화 정보, 개인화 제품을 제공하기 위해 사용자의 온라인상에서의 행위와 로그정보 분석에 집중하고 있다. 실시간 로그분석이라는 것은 갑자기 관심분야가 변경되는 경우, 과거의 로그정보가 거의 존재하지 않는 경우를 보완하기 위해 사용자 최근 행태를 기반으로 개인화하는 방법이다. 개인화는 동일한 질의어에 대해 사용자의 선호도에 따라 서로 다른 검색결과를 제공하는 검색 기법으로 다양한 개인의 정보 요구에 맞춤형된 정보를

제공하여 사용자의 검색 만족도를 높이는 기법이다(Pitcow et al. 2002). 특히, 본 연구에서 의미하는 개인화 검색은 이용자의 관심 분야를 파악하여 검색 결과를 최적화시켜 주는 것을 말하며, 검색 결과는 입력한 질의에 적합한 문서뿐만 아니라 입력 의도와 관련된 추천 정보 및 제안 정보도 포함한다.

2.3 관련연구

웹에서 정보를 찾기 위해 검색 엔진을 사용할 때 서로 다른 사용자는 서로 다른 정보 요구를 갖는다. Liang(2011)는 이러한 요구에 대해 개인화된 검색 기술을 이용하여 제공하는 정보를 개인화하는 것으로 해결할 수 있다고 하였다. 여기서 의미하는 개인화된 검색 기술은 사용자가 입력한 질의어와 문서 간의 적합성뿐만 아니라 사용자 프로파일과 문서의 관련성까지 고려한 것이다. 미리 정의된 사용자 프로파일을 이용해 검색 결과를 재순위화 하는 방법, 사용자 프로파일과 관련된 키워드로 질의를 확장하여 검색 결과의 적합성을 향상시키는 방법, 유사한 검색 행동을 보이는 사용자 그룹이 이용한 문서를 클러스터링하여 검색 결과를 재조합하는 방법 등 다양한 방법과 기술에 의해 개인화 검색이 이루어진다.

소영준, 박영택(2000)은 사용자 관심 문서를 사용자가 직접 구축하고, 구축된 문서로부터 빈도 기반의 키워드를 추출한 후 이를 질의 확장에 활용하였다. 검색된 문서의 순위 재조정이 아닌 관심 키워드와 키워드별 가중치 기반의 질의 확장 기법의 개인화 검색에 대해 제안하였다.

Shen, Tan and Zhai(2005)는 사용자가 입력한 과거 질의어와 클릭했던 문서로부터 암묵

적 사용자 히스토리를 생성하고, 이를 질의어를 수정하는 요소로 활용하여 검색결과를 재순위화 하였다.

허정, 서희철, 장명길(2006)은 키워드 기반의 질의어 파악에 있어 동음이의어의 중의성을 해소하고자 키워드 간 상호정보량과 복합명사 의미사전을 활용하였으며, 이현주, 김한일(2008)은 데이터 분류의 대표적인 방법인 사용자 태깅 기법에 기반을 둔 폭소노미를 사용하였다. 폭소노미는 사용자에게 유연성과 적응력을 제공해주는 장점이 있으나 동의어, 유의어 등의 관리가 부족한 특성으로 인해 검색정확도가 낮아질 수 있다는 단점이 있다고 밝히면서 질의어 입력 시 사용자의 성향 정보를 이용하여 연관관계가 있는 태그를 찾아낸 후 이를 태그 클라우드 형태의 UI로 제공하는 개인화 기법을 제안하였다.

Bouras, Pouloupoulos and Silintziris(2009)는 사용자가 명시적으로 제공한 개인 정보와 시스템에서 사용자의 검색 및 브라우징 세션동안 사용자의 행위로부터 학습 가능한 정보를 동시에 사용하여 사용자 관심 분류를 뉴스 기사에 접목하여 개인화 정보를 뉴스 정보를 제공하는 방법을 논하였다.

이준, 신동욱, 최종민(2009)은 사용자와 유사한 검색 행동을 갖는 사용자 집단의 검색 기록을 이용하여 사용자의 행동 예측을 수행하였다. 즉, 유사 사용자의 질의어 빈도를 사용자 행동 예측의 키워드로 정의하여 사용자가 보게 될 확률이 가장 높은 문서의 집합을 구성하였다. 그 후 일반 검색 결과와 문서 집합과 사용자가 이용할 확률이 높은 문서 집합 간의 클러스터링을 통해 검색순위를 재조정하였다.

김동욱(2011)은 사용자 자신이 관심 있는 분

아와 관련된 웹 페이지들에 북마킹을 하며, 사용자가 북마크한 URL의 태그들은 사용자가 관심 있는 분야의 단어들이라는 가정하에 사용자가 원하는 웹 문서의 URL과 연관된 태그를 저장할 수 있는 서비스인 딜리셔스 사이트(del.icio.us)를 이용하였다. 다수의 사용자에게 의해 만들어진 URL과 태그 정보를 웹 문서와 연관 태그들을 분류할 수 있는 집단지성 데이터로 활용하여 개인 프로파일을 구성하였으며, 프로파일에 구성된 태그 정보로 질의를 확장하는 방식의 개인화 검색 시스템을 논하였다.

김한준, 노준호, 장재영(2012)은 순위 재조정 기법을 주축으로 질의어 확장 기법을 융합한 방법으로써, 사용자 프로파일에 의해 추천된 확장 질의어로부터 도출된 문서 집합들에 공통적으로 출현하는 문서들의 중첩도를 평가하여, 그 결과 값을 순위 재조정에 활용하였다.

이현재(2012)는 대학 도서관 서비스에 주목한 개인의 명시적 선호정보 파악을 통한 검색 결과 정확성 향상 기법을 논하였다. 국내 연구자 선호정보를 미리 정의된 연구자 프로파일 DB내의 국가과학기술표준분류체계로 정의하였고, 이를 대학 도서관 소장자료의 주제분류 정보인 DDC와 매핑을 통해 콘텐츠와 연계점을 형성하였다.

Xu et al.(2014)는 상황 인식 및 맞춤형 서비스 제공을 위한 주요 문제를 변화하는 사용자 관심분야와 잠재 주제를 발견하는 것이라 정의하고 맵리듀스(MapReduce)²⁾의 주요 아이디

어를 적용한 분산 추론 알고리즘(D-AToT)으로 동적 사용자 관심분야를 탐색하는 모델을 제안하였다. 또한 신경정보처리시스템(NIPS) 데이터셋을 활용한 실험으로 소셜 네트워크 서비스상의 대용량 정보로부터 동적인 잠재 주제와 사용자 관심분야 및 변화 패턴을 감지하였다.

윤성희(2016)는 사용자가 최근 입력한 질의어들과 검색 시 클릭했던 문서들에 자주 나타나는 단어의 출현 빈도와 빈도별 가중치를 동시에 반영하여 등록된 데이터 셋을 개인 프로파일로 정의하였다. 또한 개인 관심분야 키워드 저장 시 중의성을 갖는 단어의 식별을 위해 기 구축된 워드넷의 단어 간 거리를 측정하여 해당 중의어가 어떤 의미를 갖는지 파악했다.

Xing et al.(2011)는 개인화된 서비스를 위해 가장 중요한 것은 개별 사용자의 관심사를 추출하는 것이라 설명하였다. 그는 도메인 온톨로지를 사용하여 과학 문헌의 문서 정보를 분석하였고, 여기에 사용자의 행동과 결합하여 관심분야를 추출하였다. 키워드 기반의 가중치와 거리 계산으로 관심 키워드를 추출한 결과와 동시에 사용자 행동 정보를 고려하였을 경우 관심분야 추출 성능이 향상됨을 설명하였다.

김광영, 심강섭, 박승진(2009)은 논문 학술정보서비스에 대해 사용자가 직접 입력한 질의어 기반으로 개인 관심분야를 결정하였다. 사용자가 논문 검색을 위해 입력한 질의어를 kNN³⁾ 분류기를 이용하여 논문 서지데이터의 분류체계인 DDC 분류 기준으로 식별한 후 개인 관심정보 관

2) 맵리듀스(MapReduce)는 구글에서 대용량 데이터 처리를 분산 병렬 컴퓨팅에서 처리하기 위한 목적으로 제작하여 2004년 발표한 소프트웨어 프레임워크다.

3) kNN(k-Nearest Neighbor): k-최근접 이웃이란 분류되어 있지 않은 정보 또는 문서들을 분류된 정보 또는 문서들 중 가장 비슷한 속성을 가진 쪽으로 할당하는 분류 기법이다.

리를 위해 질의어, 질의 빈도, DDC 값을 자동으로 저장하였다.

최근 연구 및 검색 서비스는 사용자의 의도를 이해하거나 효율적인 검색 결과 전달을 위한 개인화 연구를 지속하고 있다. 하지만 개인 검색 사용자들은 동일한 질의어에 암묵적으로 서로 다른 정보 의도를 갖고 있을 수 있기 때문에 중의성이 크고, 일반적으로 질의어가 짧기 때문에 충분한 정보량을 갖지 못하는 문제점이 존재한다. 이러한 문제를 해결하기 위해 짧은 질의어에 부여된 사용자 검색 의도를 파악하여 가장 사용자 의도에 적합한 문서를 제공해야만 한다. 특히 과학기술분야는 연구 영역의 변화가 다수 발생할 뿐만 아니라 과학기술정보는 다양한 분야에서 서로 다른 의미의 단어들이 많이 사용되는 특성이 있기 때문에 사용자 검색 의도 또는 관심 분야를 정확하게 식별하고 이에 적합한 과학기술정보를 가장 이용하기 쉬운 형태로 제공하기 위한 개인화 연구가 필요하다. 또한 정보서비스는 검색 뿐 만 아니라 다양한 기능을 제공하고 있으며 서비스를 이용하면서 발생하는 사용자의 이용행태를 실시간으로 파악하여 관심분야를 추적하는 것은 매우 중요하다. 특히 김광영, 심강섭, 박승진(2009)은 사용자 질의어를 이용해 DDC 주제분류명으로 관심분야를 식별하였으나, 검색 뿐 아니라 다양한 사용자 행태를 분석해 관심분야를 결정하는 것이 필요하다.

3. 연구의 방법 및 수행 절차

본 장에서는 개인 관심분야 추적 및 과학기술정보 개인화와 성능 평가 전반에 대한 수행

절차와 방법을 제시한다. 과학기술정보서비스 사용자 행태 분석, 과학기술정보 개인화 서비스 구현, 사용자 행태 기반의 관심분야 추적, 관심분야를 활용한 정보 개인화를 위한 구체적인 연구 수행 방법과 내용에 대해 논한다.

3.1 연구 방법

과학기술정보서비스에서 국가과학기술표준 분류 기반으로 변화하는 사용자의 관심분야를 추적하고 이를 이용해 개인화 정보를 제공하는 방법의 효과를 검증하기 위해 다음과 같은 연구 방법을 수립하였다.

첫째, 과학기술정보를 이용하는 사용자 행태는 다양하다. 원하는 정보를 찾기 위해 사용자는 검색을 수행하며 제시된 문서를 검토하고, 더 읽기를 원한다면 원문을 다운로드하거나 공유한다. 이러한 사용자 행태는 정보에 대한 관심정도에 따라 다르게 나타난다. 따라서 과학기술정보서비스 이용행태별 정보에 대한 관심정도 점수를 정의하기 위해 문헌정보학, 정보학 및 정보서비스 개발 및 운영자를 대상으로 설문조사를 수행하고 결과를 분석한다.

둘째, NTIS와 동일한 정보, 같은 검색 엔진으로 과학기술정보 개인화 서비스를 개발하여 설문 참여했던 전문가에게 일정기간 이용토록 하고, 이를 실험집단으로 가정한다. 사용자의 과학기술정보 이용 행태를 분석하고 실시간으로 사용자의 관심분야를 추적하는데, 이 때 사용자 관심분야는 국가과학기술표준분류를 매핑하여 활용한다. 추적된 개인 관심분야를 검색결과 재순위의 입력 정보로 활용하여 정보를 개인화한다.

셋째, 동일한 과학기술정보를 제공하는 NTIS

를 같은 기간, 같은 전문가에 이용토록 하고 이를 통제집단으로 정의한다.

넷째, 과학기술정보 개인화 서비스를 이용하는데 있어 질의어에 대한 제한은 두지 않는다. 단, 사용자 검색 목적과 이에 따른 관심분야 변화 추적의 유의성 판단은 개인에게 맡긴다. 이때 관심분야 변화에 대한 반강제적 유도를 위해 특정 질의어 사용을 권고한다.

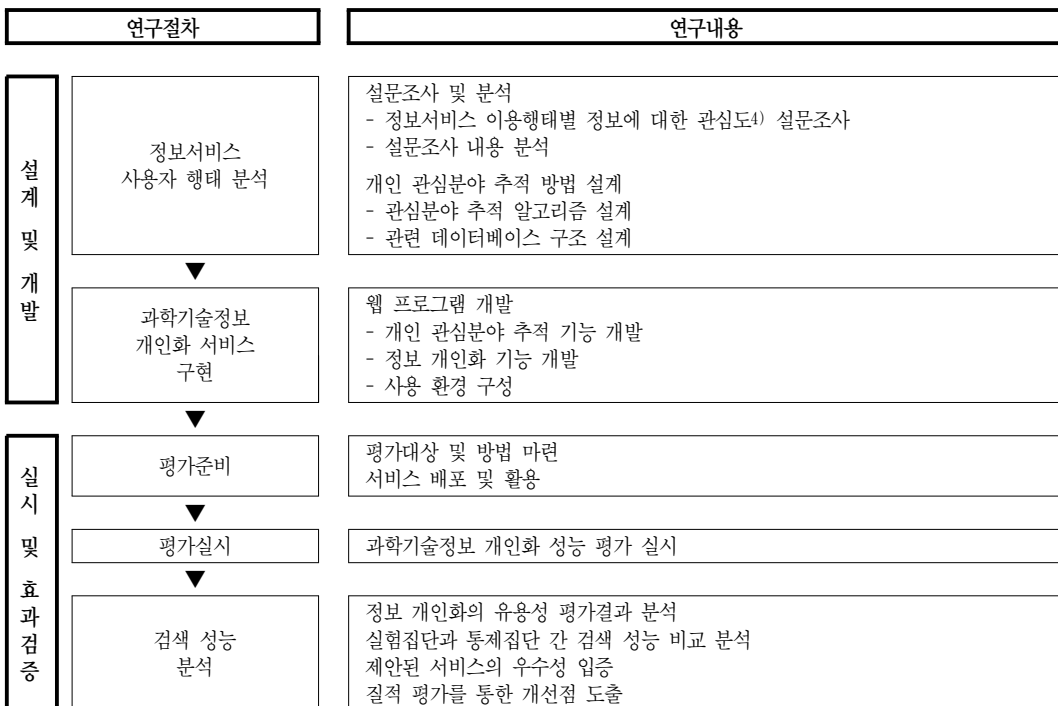
다섯째, 설문을 통해 사용자의 정보 탐색의 목적에 따라 관심분야의 추적이 제대로 이루어졌는지 평가한다. 동시에 제안된 서비스에서 제공하는 개인화 정보의 적합성 판단을 위해 NTIS와의 검색 성능 비교 평가를 수행한다. 검색 성능

평가를 위해 학문적으로 공인되고 널리 활용되는 지표를 사용한다.

마지막으로 과학기술정보서비스에서 관심분야 추적 및 정보 개인화의 만족도를 측정하고 개선사항을 도출하며 향후 연구계획을 수립하기 위해 전문가를 통한 질적 평가를 수행한다.

3.2 연구 수행 절차

과학기술정보서비스에서 국가과학기술표준분류 기반으로 개인 관심분야를 실시간으로 추적하고 이를 이용해 정보를 개인화하여 평가하기 위한 연구 수행 절차는 <그림 1>과 같다.



<그림 1> 연구 수행 절차

4) 정보서비스 내에서는 다양한 이용자 행태(검색, 상세정보클릭, 원문다운로드, 북마크 등)가 발생하는데, 특정 문서 또는 정보에 대해 얼마나 관심이 있을 경우 해당 행태를 수행하는지에 대한 정도를 수치로 표현한 것

4. 개인 관심분야 추적 및 정보 개인화

4.1 과학기술정보서비스의 사용자 행태 분석

웹을 통해 과학기술정보(과제, 논문, 특허, 연구보고서 등)를 이용할 때 사용자는 찾고자 하는 문서의 제목을 정확히 알고 검색하는 경우도 있고, 키워드로 검색을 한 후 검색결과로 제시된 문서들 가운데 원하는 정보를 찾기도 하며, 상세정보를 제공하는 화면으로 이동하거나 필요시 해당 정보를 스크랩, 북마크, 공유하기도 하며 인쇄 또는 구독의 행위를 한다. 사용자가 특정 문서에 어느 정도의 관심도가 있을 경우 각각의 행위를 하는지 수치화하고, 사용자가 과학기술정보를 탐색하거나 이용할 때 해당 문서의 국가과학기술표준분류에 수치화된 점수를 부과하여 국가과학기술표준분류 기반으로 개인 관심분야를 추적하고자 한다.

과학기술정보서비스 이용행태별 문서에 대한 관심정도 측정을 위해 서비스 이용 시 발생 가능한 사용자 행태를 14개로 정의하였다. 국내의 대표적인 학술정보서비스(NDSL,⁵⁾ RISS⁶⁾ 등) 및 과학기술정보서비스(NTIS 등)에서 제공하는 기능들을 기반으로 서비스에서 공통적으로 제공하고 있는 사용자 행태이다.

사용자 행태별 관심정도를 수치화하기 위해

10단계의 리커트 척도(Likert scale)⁷⁾를 사용하였으며, 이용하는 문서에 대한 관심정도가 낮은 순(1점)에서 높은 순(10점)으로 어느 정도 일 때 특정 행위를 하는지 조사하였다. 설문 조사는 논문, 특허, 연구보고서, 도서 등 학술정보와 문헌정보를 주로 이용하면서 특히 과학기술정보의 유통과 서비스에 관한 지식이 있는 학생 또는 연구자 100명을 대상으로 하였다. 문헌정보학 또는 정보학을 전공하고 정보서비스 기획, 운영, 개발 분야에 10년 이상 종사했거나 종사하고 있는 정보서비스 전문가 50명, 도서관 및 정보서비스 제공 기관의 근무 경력이 있는 전문가 20명, 그리고 정보서비스 제공자 및 사용자를 모두 고려하여 과학기술정보서비스 NTIS의 회원 30명을 선정하였다. 개인 식별이 가능한 정보는 수집하지 않았으며, 사용자 행태별 문서에 대한 관심정도를 측정하는 질문으로 설문 조사를 수행하였다. 설문조사 방법은 웹 기반의 무료 온라인 설문조사 플랫폼 서베이몽키(SurveyMonkey)를 활용하였고, 일주일 간(2016.5.9.~5.13) 진행된 설문조사 결과 100명의 설문 대상자 가운데 53명이 응답하였다.

〈표 1〉은 과학기술정보서비스 사용자 행태별 정보에 대한 관심정도를 조사하는 설문에 대한 응답결과이다. 〈수식 1〉과 같이 관심정도 점수와 응답 사용자수를 인수로 하여 계산한 가중평균(Weighted Average)으로 각 사용자행

5) NDSL(National Digital Science Library)은 한국과학기술정보연구원이 운영하는 학술정보서비스 플랫폼으로 논문, 특허, 보고서, 동향, 저널/프로시딩, 연구자, 연구기관 등 약 1억 건 이상의 콘텐츠를 제공한다.

6) RISS(Research Information Sharing Service)는 전국 대학이 생산하고 보유하며 구독하는 학술자원을 공동으로 이용할 수 있도록 개방된 대국민 서비스로써 한국교육학술정보원이 운영한다.

7) 태도 측정법의 하나로, 피실험자에게 조사항목에 대한 동의 여부를 묻지 않고, 각각의 항목에 대한 동의 정도를 표시하도록 하는 측정방법으로 명칭은 이 척도 사용에 대한 보고서를 발간한 렌시스 리커트(Rensis Likert)의 이름에서 따온 것이다.

〈표 1〉 과학기술정보서비스 이용 행태별 정보에 대한 관심정도 응답결과

	관심정도별 응답비율 및 응답자수									
	1	2	3	4	5	6	7	8	9	10
문서의 정확한 제목으로 검색	0.00% 0	1.89% 1	5.66% 3	1.89% 1	7.55% 4	13.21% 7	15.09% 8	22.64% 12	13.21% 7	18.87% 10
일반적인 키워드 검색 (검색결과 다수 존재)	15.09% 8	26.42% 14	28.30% 15	5.66% 3	11.32% 6	5.66% 3	3.77% 2	1.89% 1	1.89% 1	0.00% 0
상세보기로 이동	0.00% 0	0.00% 0	0.00% 0	9.43% 5	30.19% 16	7.55% 4	16.98% 9	11.32% 6	18.87% 10	5.66% 3
상세페이지 인쇄	5.66% 3	13.21% 7	5.66% 3	1.89% 1	5.66% 3	1.89% 1	5.66% 3	20.75% 11	15.09% 8	24.53% 13
장바구니 담기	13.21% 7	13.21% 7	3.77% 2	3.77% 2	3.77% 2	3.77% 2	9.43% 5	11.32% 6	15.09% 8	22.64% 12
구매 및 구독하기(결제)	16.98% 9	5.66% 3	5.66% 3	3.77% 2	3.77% 2	1.89% 1	3.77% 2	1.89% 1	5.66% 3	50.94% 27
원문보기	0.00% 0	0.00% 0	1.89% 1	0.00% 0	11.32% 6	5.66% 3	13.21% 7	20.75% 11	16.98% 9	30.19% 16
원문복사신청	1.89% 1	5.66% 3	5.66% 3	1.89% 1	0.00% 0	5.66% 3	7.55% 4	7.55% 4	28.30% 15	35.85% 19
북마크(스크랩 등)	1.89% 1	11.32% 6	0.00% 0	0.00% 0	3.77% 2	13.21% 7	9.43% 5	33.96% 18	13.21% 7	13.21% 7
이메일 보내기	7.55% 4	13.21% 7	5.66% 3	3.77% 2	7.55% 4	7.55% 4	7.55% 4	24.53% 13	5.66% 3	16.98% 9
SNS공유	22.64% 12	5.66% 3	5.66% 3	1.89% 1	9.43% 5	5.66% 3	13.21% 7	5.66% 3	13.21% 7	16.98% 9
좋아요 클릭	26.42% 14	5.66% 3	3.77% 2	1.89% 1	7.55% 4	11.32% 6	11.32% 6	9.43% 5	7.55% 4	15.09% 8
유사문서 보기 실행	1.89% 1	13.21% 7	26.42% 14	7.55% 4	30.19% 16	7.55% 4	3.77% 2	0.00% 0	5.66% 3	3.77% 2
댓글달기	32.08% 17	5.66% 3	7.55% 4	0.00% 0	3.77% 2	3.77% 2	11.32% 6	3.77% 2	9.43% 5	22.64% 12

〈수식 1〉 $WA_q = \frac{\sum_{i=1}^{10} i \times \text{응답자수}_i}{\text{전체응답자수}}$, where q is the number of questions

태별 관심정도를 수치화하였고, 〈표 2〉는 사용 행태별 관심정도 점수의 내림차순 정렬 결과를 나타낸다. 〈표 2〉에서 보는 바와 같이 사용자가 어떤 문서에 관심이 있는 경우, 원문보기, 원문 복사신청, 정확한 제목으로 검색, 북마크 한 경우 순서로 관심정도의 점수가 높다는 결과가

나왔으며 상대적으로 키워드 검색, 유사문서보기, 댓글달기는 관심정도 점수가 낮다. 사용자 이용 행태별 점수는 특정 문서에 부여된 국가 과학기술표준분류의 점수로 산정되어 정보 개인화에 활용된다.

〈표 2〉 과학기술정보서비스 이용 행태별 정보에 대한 관심정도

이용 행태		관심정도 점수
1	원문보기	8.09
2	원문복사신청	7.98
3	문서의 정확한 제목으로 검색	7.40
4	북마크(스크랩 등)	7.11
5	구매 및 구독하기(결제)	6.92
6	상세페이지 인쇄	6.83
7	상세보기로 이동	6.70
8	장바구니 담기	6.26
9	이메일 보내기	6.19
10	SNS공유	5.66
11	좋아요 클릭	4.43
12	댓글달기	3.98
13	유사문서보기 실행	3.47
14	키워드 검색 (검색결과 다수 존재)	2.32

4.2 과학기술정보 개인화 서비스 구현

과학기술정보 개인화 서비스 구현을 위해 NTIS에서 가장 많이 활용되면서 각 레코드에 국가과학기술표준분류가 부여된 정보를 대상으로 하였다. 개인화 서비스 구현의 대상 정보는 〈표 3〉과 같다.

과학기술정보 개인화 서비스 검색 기능은 NTIS에서 제공하는 공개API를 활용하였다. 공개API를 이용하면 서비스 디자인과 데이터들

분리함으로써 서비스 확장 및 수정이 용이하다. 중요한 것은 동일한 과학기술정보를 제공하는 개인화 서비스를 개발하여 검색결과의 정확성을 NTIS와 비교할 수 있다는 것이다. 과학기술정보 개인화 서비스 구현을 위해 사용자정보, 관심정보 등을 관리하기 위한 데이터베이스는 오라클 11g를 이용하였고 NTIS 공개API로부터 검색결과를 요청하여 응답받는 WAS(Web Application Server)는 톰캣(Tomcat) 8.0 버전을 활용하였으며 개발 언어는 JAVA를 이용

〈표 3〉 과학기술정보 개인화 서비스 대상 정보 ('17년 7월 1일 기준)

DB종류	설 명	건수
과제	2002년부터 수행한 국가연구개발 과제 정보	626,901
논문	2007년부터 국가연구개발과제에서 산출된 논문 정보	1,058,017
특허	2007년부터 국가연구개발과제에서 산출된 특허 정보	327,671
연구보고서	2002년부터 수행한 국가연구개발 과제 연구보고서 정보	93,404
기술요약	2002년부터 국가연구개발과제에서 산출된 기술요약정보	29,802
연구시설장비	1987년부터 구축된 연구시설장비 정보	103,378
합 계		2,239,173

하였다.

검색API를 활용하기 때문에 별도의 과학기술정보 관리를 위한 DB는 필요 없으나, 사용자 정보, 이용행태별 관심정도 점수, 사용자 관심분야 등의 정보관리를 위해 데이터베이스에 대한 설계가 필요하다. <표 4>는 과학기술정보 개인화 서비스를 위해 설계된 4개 DB 테이블과

설명을 나타낸다.

<표 5>는 각 테이블 별 필드의 구성을 나타낸다. 사용자 정보 테이블(P_USER_INFO)은 사용자 식별을 위한 최소한의 정보인 이름, 아이디, 이메일, 소속기관, 등록일로 구성된다. 여기서 이메일은 향후 개인화 서비스의 성능평가 요청을 위해 관리하는 정보이고, 소속기관

<표 4> 과학기술정보 개인화 서비스 DB 테이블

테이블명	설 명
P_USER_INFO	사용자 정보(이름, ID, 이메일 등)를 관리한다.
P_INTEREST_CODE	이용 행태별 관심정도 점수를 관리한다.
P_USER_INTEREST	사용자별 관심분야(국가과학기술표준분류와 점수)를 관리한다.
P_USE_LOG	사용자별 이용 행태 로그와 질의어 히스토리를 관리한다.

<표 5> 과학기술정보 개인화 서비스 DB 테이블 스키마

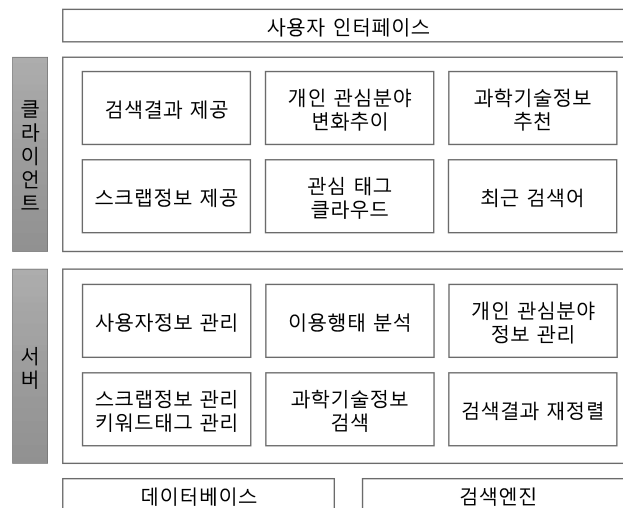
테이블명	필드명	데이터 타입	설명
P_USER_INFO	USER_ID	varchar2(20)	사용자 아이디
	USER_NM	varchar2(20)	사용자 이름
	USER_EMAIL	varchar2(20)	사용자 이메일
	ORG_NM	varchar2(20)	소속기관명
	REGIST_DT	date	가입일시
P_INTEREST_CODE	CODE	varchar2(2)	이용행태 코드
	CODE_NM	varchar2(50)	이용행태
	CODE_WEIGHT	number	관심정도 점수
P_USER_INTEREST	SEQ	number	시퀀스 번호
	USER_ID	varchar2(20)	사용자 아이디
	SCIENCE_NM	varchar2(200)	관심분야명
	SCORE	number	관심정도 점수
	DAY_SCORE	number	관심정보 점수 일일 누적
	TOTAL_SCORE	number	관심정보 점수 전체 누적
	REGIST_DT	date	이용행태 발생일시
P_USE_LOG	SEQ	number	시퀀스 번호
	USER_ID	varchar2(20)	사용자 아이디
	ACT_FLAG	varchar2(2)	이용행태 코드
	ACT_SCORE	number	관심정도 점수
	DB_TYPE	varchar2(10)	이용 DB
	USER_QUERY	varchar2(200)	질의어
	ACT_DT	date	로그일시

명은 향후 기관별 질의어 특성 분석 또는 국가 과학기술표준분류 연관성 파악을 위해 저장하였다. 그러나 정보 추천 및 개인화를 위한 프로파일 구성을 위해 입력받는 정보는 없는 것이 특징이다. 이용 행태별 관심정도 점수 테이블(P_INTEREST_CODE)은 3.2장의 설문조사 결과를 입력한 정보이고, 별도의 DB 테이블로 구성하여 향후 사용자 이용행태의 수정, 입력, 삭제 등의 확장이 용이하도록 하였다. 사용자별 관심분야 테이블(P_USER_INTEREST)은 사용자 ID별 국가과학기술표준분류명과 해당 점수를 누적 계산하여 저장, 관리한다. 사용자가 어떤 문서에 대해 이용 행태별 관심정도 점수 테이블(P_INTEREST_CODE)에 정의된 이용 행태를 할 때마다 해당 문서가 지니고 있는 국가과학기술표준분류의 중분류명에 관심정도 점수가 누적 계산되어 저장되며, 이 테이블에서 관리하는 정보는 검색결과 개인화 및 정보 개인화에 활용된다. 로그 테이블(P_USE_LOG)은 과학기술

정보 개인화 서비스에서 발행하는 모든 사용자 이용 행태와 질의어를 시간별로 기록하여 사용 로그 분석 및 최근 검색어 정보를 제공하는데 활용된다.

과학기술정보 검색 공개API를 이용하여 정보를 검색하고, 제공된 정보에 대한 사용자 이용행태를 분석하여 관심분야를 추적하며 결정된 개인 관심분야를 다시 정보를 개인화하는데 활용하는 일련의 선순환 프로세스를 구현하는 개인화 서비스는 <그림 2>와 같은 구성도를 가진다.

서버측은 회원정보를 관리하는 '사용자정보 관리', 과학기술정보를 이용하는 이용 행태별 로그를 분석하여 관리하는 '이용행태 분석', 분석된 결과를 이용해 국가과학기술표준분류에 관심도 점수를 계산하여 관리하는 '개인 관심분야정보 관리', 스크랩 정보와 개인 관심 키워드를 저장하는 '스크랩정보 관리'와 '키워드태그 관리', NTIS 공개API를 이용하여 검색결과



<그림 2> 과학기술정보 개인화 서비스 구성도

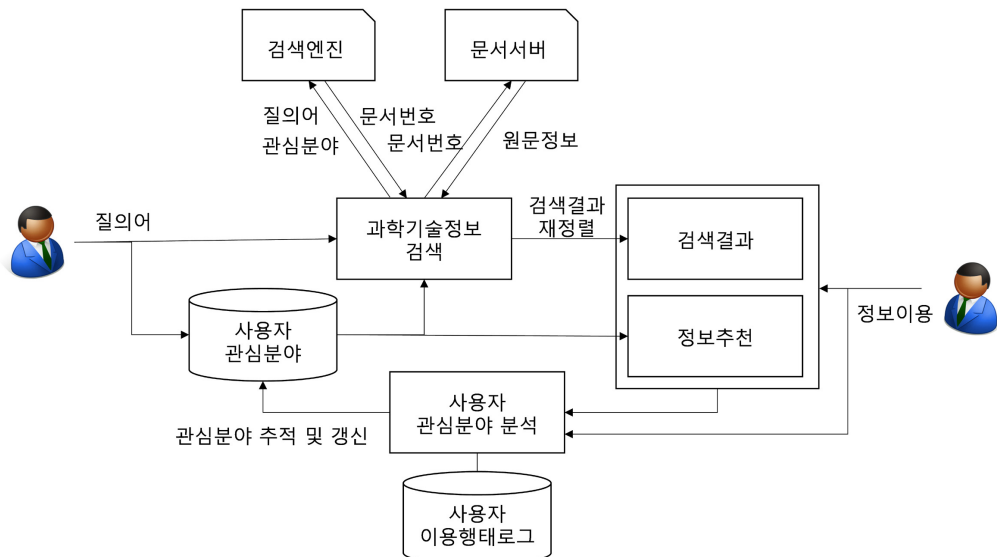
로 메타데이터와 원문정보를 가져오는 '과학기술정보 검색', 개인 관심분야에 따라 검색결과를 재순위화하는 '검색결과 재정렬' 부분으로 구성된다.

클라이언트 측은 개인화된 과학기술정보를 제공하는 '검색결과 제공', 정보 이용 행태 기반의 개인 관심분야(국가과학기술표준분류) 비율을 제시하는 '개인 관심분야 변화추이', 검색 기능을 수행하지 않고도 관심분야 기반의 정보를 제공하는 '과학기술정보 추천', '스크랩 정보 제공', 사용자가 자주 이용하는 과학기술정보가 지닌 키워드를 빈도순으로 저장하여 제공하는 '관심 태그 클라우드', 검색 히스토리를 통해 검색을 바로 수행하도록 도와주는 '최근 검색어' 기능으로 구성된다.

사용자 개인에게 맞춤형, 즉 개인 관심분야에 적합한 과학기술정보를 제공하기 위해 가장 중요한 요소는 이용행태 분석을 통한 개인 관

심분야정보 추적 및 갱신, 관심분야를 적용한 정보의 개인화이다. <그림 3>은 개인 관심분야를 추적하고 정보를 개인화하는 프로세스 흐름도를 나타낸다.

사용자가 원하는 정보를 찾기 위해 로그인 후 질의어를 입력하면 해당 질의어와 기 추적되어 저장된 사용자 관심분야(국가과학기술표준분류)는 검색 기능에 전달된다. 검색엔진과 문서서버로부터 얻은 검색결과와 문서서버로부터 얻은 검색결과와 개인 관심분야에 적합한 순으로 재정렬되어 사용자에게 전달된다. 사용자가 검색을 수행하지 않고 개인화 서비스에 로그인 한 경우에도 기 저장된 사용자 관심분야와 최근 검색어 기반의 과학기술 정보가 추천된다. 이 때 사용자는 정보를 보거나 구매하거나 스크랩하는 등의 다양한 행태를 취하게 되고 로그정보가 저장되는 동시에 관심분야가 추적되고 갱신된다.



<그림 3> 관심분야 추적 및 정보 개인화 프로세스 흐름도

4.3 사용자 행태 기반의 관심분야 추적

사용자 관심분야 추적을 위해 4.1장에서 설명한 설문조사 결과인 '사용자 이용행태별 관심정도 점수'를 활용한다. 단, 관심정도 계산을 용이하게 하기 위해 각 점수를 100점으로 정규화 하였으며, 소수점 아래는 반올림하였다. 그리고 관심분야로 활용할 분류정보는 국가과학기술표준분류의 중분류명을 사용한다. 대분류명은 분류명이 의미하는 범위가 너무 포괄적이어서 개인별 관심분야로 활용할 때 모호성이 크며, 세부분류명은 전체 과학기술정보의 50% 이상이 빈 필드로 존재하기 때문이다.

〈표 2〉에서 정의된 14개의 서비스 이용 행태가 발생할 때마다 사용자 관심분야에 해당하는

국가과학기술표준분류 중분류명의 관심도 점수는 누적되어 갱신되는데, 서비스를 지속적으로 이용하면서 행위가 다양하게 계속됨에 따라 관심분야에 해당하는 중분류명의 개수와 점수는 증가한다. 이 때, 관심분야별 점수 갱신은 일일 기준으로 산정되며, 특정 사용자의 관심분야로 나타난 모든 중분류명의 점수를 이용해 상대 비율을 계산한다.

사용자의 관심분야는 시간이 지남에 따라 또는 정보의 활용 목적에 따라 변화하기 마련이며, 변하는 관심분야를 추적하여 적합한 과학기술정보를 제공하는 것이 연구의 목적이다. 사용자가 과학기술정보를 이용함에 따라 개인 관심분야는 아래와 같이 추적된다. 관심분야 비율 계산은 〈수식 2〉와 같다.

$$\text{〈수식 2〉} \quad \text{관심분야비율}_p = \frac{DAYSCORE_p}{\sum_{i=1}^s DAYSCORE_i} \times 100, \text{ where } s \text{ is the number of SEQ}$$

- ① 홍길동(id:someone)은 컴퓨터공학을 전공한 후 IT분야에서 데이터 분석 업무를 수행하는 연구자이고, 빅데이터 및 머신 러닝에 대한 논문 정보를 이용하고자 '딥러닝' 질의어로 검색을 수행하였고, 총 89건의 논문 검색결과를 얻었다. 이 때 논문 89건의 국가과학기술표준분류 중분류명 군집 결과 로봇제어/지능화 기술, 인공지능, SW솔루션이 상위 3위를 차지하였다.

[관심분야(P_USER_INTEREST) 정보 갱신]

SEQ	USER_ID	SCIENCE_NM	DAY_SCORE	REGIST_DT
1	someone	로봇제어/지능화기술	23	2017/07/28 오후 5:55:26
2	someone	인공지능	23	2017/07/28 오후 5:55:26
3	someone	S/W솔루션	23	2017/07/28 오후 5:55:26

[관심분야 비율]

순번	관심분야	계산식	관심비율
1	로봇제어/지능화기술	{23 / (23+23+23)} * 100	33%
2	인공지능	{23 / (23+23+23)} * 100	33%
3	S/W솔루션	{23 / (23+23+23)} * 100	33%

- ② 홍길동은 검색결과를 검토한 후, '사물인터넷을 이용한 노년층 질환 예측에 관한 연구' 논문의 상세보기를 클릭하고 논문 원문을 다운로드하였다. 이 때 해당 논문의 메타데이터 국가과학기술표준분류 중분류명은 인공지능이다.

[관심분야(P_USER_INTEREST) 정보 갱신]

SEQ	USER_ID	SCIENCE_NM	DAY_SCORE	REGIST_DT
1	someone	로봇제어/지능화기술	23	2017/07/28 오후 5:55:26
2	someone	인공지능	23+67+81=171	2017/07/28 오후 6:02:10
3	someone	S/W솔루션	23	2017/07/28 오후 5:55:26

[관심분야 비율]

순번	관심분야	계산식	관심비율
1	로봇제어/지능화기술	{23 / (23+171+23)} * 100	11%
2	인공지능	{171 / (23+171+23)} * 100	79%
3	S/W솔루션	{23 / (23+171+23)} * 100	11%

- ③ 마침 홍길동이 근무하는 연구소에서 데이터 분석을 통한 치매 질병예측 과제를 기획하던 중이었다. 다운로드 받은 논문을 읽은 후 다음날 홍길동은 치매의 원인과 진단, 예방을 위한 빅데이터 분석 관련 국가R&D과제를 찾아보기로 하였다. 질의어 '치매 빅데이터'로 과제 검색을 수행하였고, 총 66건의 과제 검색결과를 얻었다. 이 때 과제 89건의 국가과학기술표준분류 중분류명 군집 결과 뇌공학, 소프트웨어, 정보이론이 상위 3위를 차지하였다.

[관심분야(P_USER_INTEREST) 정보 갱신]

SEQ	USER_ID	SCIENCE_NM	DAY_SCORE	REGIST_DT
1	someone	로봇제어/지능화기술	23	2017/07/28 오후 5:55:26
2	someone	인공지능	171	2017/07/28 오후 6:02:10
3	someone	S/W솔루션	23	2017/07/28 오후 5:55:26
4	someone	뇌공학	23	2017/07/29 오후 2:15:02
5	someone	소프트웨어	23	2017/07/29 오후 2:15:02
6	someone	정보이론	23	2017/07/29 오후 2:15:02

[관심분야 비율]

순번	관심분야	계산식	관심비율
1	로봇제어/지능화기술	{23 / (23+171+23+23+23+23)} * 100	8%
2	인공지능	{171 / (23+171+23+23+23+23)} * 100	60%
3	S/W솔루션	{23 / (23+171+23+23+23+23)} * 100	8%
4	뇌공학	{23 / (23+171+23+23+23+23)} * 100	8%
5	소프트웨어	{23 / (23+171+23+23+23+23)} * 100	8%
6	정보이론	{23 / (23+171+23+23+23+23)} * 100	8%

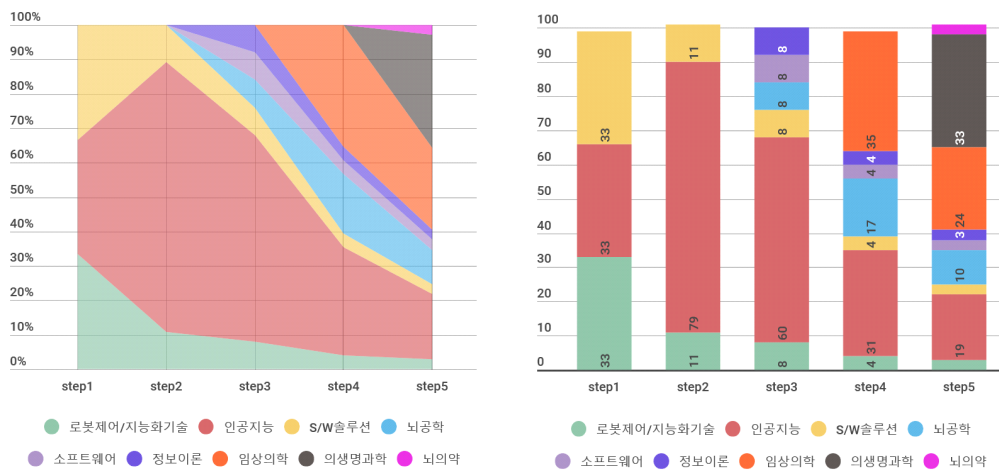
사용자는 과학기술정보를 다양한 행태로 이용하게 되고, 이용하는 정보가 지닌 국가과학기술표준분류와 행태별 관심도 점수는 이용 행태가 발생할 때마다 계속하여 갱신된다. <그림 4>는 사용자가 과학기술정보 개인화 서비스에서 정보를 이용하거나 검색을 수행함에 따라 추적되는 관심분야 결과의 예이다. 사용자 관심분야는 초기에 인공지능, 로봇제어/지능화기술이었으나 검색 및 정보 이용이 계속됨에 따라 관심분야가 점점 의생명과학, 임상의학, 뇌공학 분야로 이동하는 것을 볼 수 있다.

4.4 관심분야를 활용한 과학기술정보 개인화

관심분야가 서로 다르게 추적된 사용자가 검색을 수행하였을 때 검색결과 내의 문서들 중에서 각자의 관심분야로 결정된 국가과학기술표준분류에 해당하는 문서들을 검색 상위에 노출시킨다. 서로 다른 사용자가 동일한 질의어를 입력하였다고 하더라도 각자의 관심분야에

적합한 문서를 이용하기 먼저 제공하는 것이다. 관심분야를 활용한 정보 개인화를 위해 검색결과 개인화, 관심 태그 클라우드, 관심분야 기반 추천정보의 세 가지 측면에서 논한다.

국가과학기술표준분류 중분류명을 이용해 추적된 개인 관심분야는 검색결과 순위조정에 반영된다. 검색 엔진에 질의어와 개인 관심분야가 전달되면 검색 엔진은 질의어에 적합한 결과 셋(result set)을 생성한 후 검색 결과 셋의 문서들 중에 특정 관심분야 정보를 가지는 문서를 상위에 위치시킴으로써 재정렬을 수행한다. 단, 검색 포털에서 사용자가 첫 페이지의 상위 3건 내의 정보에 대한 클릭률이 50%가 넘고 원하는 정보의 탐색 노력이 셋째 페이지 이상에서는 거의 이루어지지 않는다는 조사 결과에 기반하여, 검색 결과 셋(set) 가운데 개인 관심분야 국가과학기술표준분류 중분류명을 가지는 상위 문서 30건에 가중치를 부여하여 전체 검색 결과 순위 30위에 포함되도록 재순위화하는 방법을 적용하였다.



<그림 4> 국가과학기술표준분류 기반 개인 관심분야 추적결과 예시

사용자가 정보를 이용하는 다양한 행태를 일으킬 때마다 개인화 시스템은 최근까지 추적된 관심분야 DB 테이블에서 비율이 높은 순으로 3개 국가과학기술표준분류 중분류명을 질의어와 함께 검색 엔진에 전달한다. 그 다음 검색 엔진은 질의어에 적합한 결과 문서 집합을 생성한다. 그 후 관심분야를 반영한 검색결과 재순위를 실행하기 위해 상대 질의 부스팅(relative query boosting) 매커니즘을 적용한다. 상대 질의 부스팅은 검색 결과 문서들 중 특정 질의어를 포함한 문서의 랭크(rank)값⁸⁾을 조정하여 검색결과 내에서 상대적 위치를 조정하는 기법이다. 검색결과 문서들 중 특정 개인의 관심분야에 해당하는 국가과학기술표준분류 중분류명을 메타데이터로 가진 상위 30위 문서에 전체 문서들 중 1위부터 60위 문서 랭크 값의 평균값을 가중치(weight)로 부여하여 검색 순위를 재

조정한다.

비교 대상의 NTIS는 사용자가 어떤 분야에 관심이 있는지, 어떤 자료에 대한 선호도가 있는지 파악하지 않고 단지 질의어와 색인과의 적합성을 판단해 검색결과를 제공한다. 요구와 관심분야가 제각각인 사용자들에게 모두 같은 검색결과를 이용하라는 것이다. 단적이고 직관적인 예로 유전자공학자가 입력하는 'cell'⁹⁾과 반도체 전문가나 정보통신 공학자가 입력하는 'cell'은 서로 다른 검색 결과를 기대한다. <표 6>은 사용자가 개인화 서비스에 회원 가입 후 관심분야가 추적되지 않은 상태에서 첫 질의어로 'cell'로 과제 검색을 수행한 결과이며, 이는 NTIS에서 제공하는 검색결과와 동일하다.

하지만 사용자가 개인화 서비스를 이용하면서 개인 관심분야는 국가과학기술표준분류 기반으로 추적된다. 관심분야가 분자세포생물학,

<표 6> NTIS 및 개인화 서비스 비로그인의 경우 질의어 'cell'의 과제검색 결과

	R&D과제 제목	중분류명
1	세포분열저해 저독성 항암제 개발	N/A
2	분자제어에 의한 나노바이오 프로브 개발	N/A
3	고형암에서 방사선 치료를 이용한 효과적인 CAR T cell therapy 연구	방사선기술
4	글로벌 줄기세포/재생의료 연구개발촉진센터	기타 생명과학
5	중앙 특이적 분자영상 및 방사선치료 효율 증진을 위한 다기능성 중공 메조다공성 실리카 나노입자 개발	방사선기술
6	프리온 free 소 모델을 이용한 프리온 기능 연구	유전학/유전공학
7	Human Long-lived NK 세포와 중앙 특이적 기억 T 세포 유도기술개발	의약품/의약품개발
8	소뇌실조증 질환동물모델에 대한 줄기세포치료제의 유효성 및 기전연구	의약품/의약품개발
9	중배엽 및 외배엽 표지인자 도입 연구용 인간배아줄기세포주 확립 및 세포주분양과 교육	발생/신경생물학
10	화합물계 나노소자공정 실습교육과정 운영	반도체소자/시스템

8) FAST 검색엔진은 질의어에 적합한 문서들의 검색결과 셋을 생성하는데, 이 때 각각의 문서는 rank value가 부여된다. 이는 검색 엔진 자체의 적합성 판정 알고리즘에 의해 문서의 적합성 순위를 표현한 숫자이다.
 9) cell은 생물학(biology)에서는 유기체의 기본 구조 및 활동 단위인 세포를 의미하며, 이동통신공학에서는 이동통신에서 하나의 기지국이 포괄하는 지역을 가리키는 개념이다. 또한 반도체나 소프트웨어 공학 영역에서 cell은 메모리에서 데이터를 저장하는 최소단위를 뜻한다.

농생물학, 의생명과학으로 추적된 사용자의 경우 질의어 'cell'에 대해 <표 7>의 논문 검색 결과가 제시된다. 반면 이동통신, U-컴퓨팅의 관심분야를 지닌 사용자에게는 <표 8>의 논문 검색 결과가 제시된다.

정보 개인화를 위해 개인화 서비스는 관심 태그 클라우드 기능을 통해 30개의 관심 키워드 후보군을 제공한다. 사용자가 과학기술정보의 어떤 문서의 상세정보를 이용하거나, 스크랩, 인쇄, 원문보기 등 검색을 제외한 다른 모든 행태를 할 때마다 해당 문서의 메타데이터 중 키워드 한글, 키워드 영문의 정보를 추출한다. 단, 특히, 기술요약, 시설장비 정보는 국가과학기술표준정보에 키워드 항목이 정의되지 않아 수집 대상이 아니므로 제외한다. 과제, 논문, 연구보고서 정보를 사용자가 이용할 때마다 해당 문서의 키워드 정보는 별도의 테이블에 저장되는데 이때 중복된 키워드는 빈도수를

누적한다. 키워드는 해당 과학기술정보를 가장 함축적이고 정확하며 간결하게 요약하여 놓은 단어들의 집합이다. 관심분야에 해당하는 과학기술정보들을 대표하는 키워드들은 질의어 후보 가능성이 높으며, 이는 사용자가 관심분야의 문서들을 쉽게 접근할 수 있도록 도와주는 역할을 한다.

사용자가 검색을 수행하기 전에 선호할 만한 과학기술정보를 제공하는 기능이 관심분야 기반 추천정보이다. 추천정보 역시 사용자 행태가 발생될 때마다 바뀌어 제공되는데, 이는 사용자 최근 질의어 3개와 관심분야 상위 3개 국가과학기술표준분류 중분류명을 이용해 추천정보를 생성하기 때문이다. 최근 질의어 3개로 과제, 논문, 특히, 연구보고서, 기술요약, 시설장비를 검색하고 관심분야 정보로 부스팅한 결과를 제공하며 <그림 5>는 관심분야 기반 추천정보의 예시 화면이다.

<표 7> 생물학 관심분야 사용자의 'cell'에 대한 논문 검색 결과

	논문 제목	중분류명
1	Dihydroergotamine Tartrate Induces Lung Cancer Cell Death through Apoptosis and Mitophagy	달리 분류되지 않는 분자세포 생물학
2	Hypoxia-induced expression of cellular prion protein improves the therapeutic potential of mesenchymal stem cells	분자세포생물학
3	Promising Therapeutic Strategies for Mesenchymal Stem Cell-Based Cardiovascular Regeneration: From Cell Priming to Tissue Engineering	줄기세포생물학
4	The prognostic value of interim and end-of-treatment PET/CT in patients with newly diagnosed peripheral T-cell lymphoma	분자세포생물학
5	Bioinspired tuning of glycol chitosan for 3D cell culture	줄기세포생물학
6	An 8-gene signature for prediction of prognosis and chemoresponse in non-small cell lung cancer	분자세포생물학
7	MAPK Cascades in Guard Cell Signal Transduction	시스템생물학
8	A Genetic Screen Reveals Novel Targets to Render Pseudomonas aeruginosa Sensitive to Lysozyme and Cell Wall-Targeting Antibiotics	미생물/기생생물학
9	Biphasic activation of extracellular signal-regulated kinase (ERK) 1/2 in epidermal growth factor (EGF)-stimulated SW480 colorectal cancer cells	시스템생물학
10	Stepwise inhibition of T cell recruitment at post-capillary venules by orally active desulfated heparins in inflammatory arthritis	분자세포생물학

〈표 8〉 이동통신, U-컴퓨팅, 소프트웨어 관심분야 사용자의 'cell'에 대한 논문 검색 결과

	논문 제목	중분류명
1	A Cell Searching Technique without Double Counting for a Mobile Station with Multiple Antenna Arrays in Millimeter Wave Cellular Communication Systems	U-컴퓨팅 플랫폼/ 응용기술
2	An Enhanced Clustering Method Based on Grid-shaking	소프트웨어
3	Stability of SiNX/SiNX double stack antireflection coating for single crystalline silicon solar cells	소프트웨어
4	자바 애플릿을 이용한 2차원 혼합형 비정렬격자 생성 프로그램의 개발	소프트웨어
5	위상차 현미경 영상 내 푸리에 묘사자를 이용한 암세포 형태별 분류	소프트웨어
6	A Statistical Inter-Cell Interference Model for Uplink Cellular OFDMA Networks Under Log-Normal Shadowing and Rayleigh Fading	U-컴퓨팅
7	Virtual Cell Beamforming in Cooperative Networks	소프트웨어
8	NDRG2 is one of novel intrinsic factors for regulation of IL-10 production in human myeloid cell	U-컴퓨팅
9	Library Support in an Actor-Based Parallel Programming Platform	소프트웨어
10	SCAMP5 Links Endoplasmic Reticulum Stress to the Accumulation of Expanded Polyglutamine Protein Aggregates via Endocytosis Inhibition	소프트웨어

관심분야 기반 추천정보

과제	논문	특허	연구보고서	기술요약	시설장비
Protection method of stack upon emergency shut down or black out in the solid oxide fuel cell system Cross-flow membrane test cell and membrane test device with the same CELL -BASED TRANSPARENT SENSOR CAPABLE OF REAL-TIME OPTICAL OBSERVATION OF CELL BEHAVIOR, METHOD FOR MANUFACTURING THE SAME AND MULTI-DETECTION SENSOR CHIP USING THE SAME 종양 특이적 프로모터 및 이를 포함하는 종양 살상 바이러스 벡터 (Tumor-specific promoter and oncolytic virus vector comprising the same) 곰우단백질 자실체로부터 분리된 신규한 파라티페닐 화합물을 유효성분으로 포함하는 인플루엔자 바이러스 감염 질환의 예방 또는 치료용 조성물					

〈그림 5〉 관심분야 기반 추천정보 화면 예시

5. 과학기술정보 개인화 성능 평가

개인 관심분야 추적과 정보 개인화에 대한 성능을 평가하였다. 첫째, 사용자가 과학기술정보 개인화 서비스를 통해 정보를 검색하거나 이용할 때 자신의 검색의도 혹은 목적 대비하여 관심분야 추적결과가 어느 정도 부합하는지에 대한 정확성을 평가하였다. 둘째, 관심분야 추적

결과 후의 검색 결과 정확성을 측정하였다. 이때 관심분야 추적 전을 논리적 통제집단으로 간주하고 NTIS의 검색 결과 정확성을 측정하고, 관심분야 추적 후는 과학기술정보 개인화 서비스를 통한 검색 결과의 정확성을 측정하여 상호 비교 평가하였다. 셋째, 개인화 서비스에서 제시하는 관심 태그 클라우드와 관심분야 기반 추천정보의 유용성 판단을 위해 각 기능의 만족도

를 평가하였으며, 마지막으로 NTIS와 본 연구에서 제시한 과학기술정보 개인화 서비스에 대해 각각 선호하는 이유와 개선 필요사항에 대해 조사하고자 질적 평가를 수행하였다.

과학기술정보 개인화 성능평가는 사용자 행태 분석의 대상이 되었던 100명의 전문가 중 해당 영역에 10년 이상 근무하였고 특히 정보학, 전산학, 컴퓨터공학, 생명공학 분야를 전공하였거나 해당 분야에 연구를 수행하고 있는 자 40여명을 후보로 하여 <표 9>와 같이 최종 12명으로 선정하였다. 성능 평가에 참여 의사를 보낸 전문가 가운데 최종적으로 평가 당시 연구 수행 분야 및 관심분야 기준, 컴퓨터 및 정보학 분야 6명, 생명공학 및 바이오/의료 분야 6명을 선정하여 진행하였다. 개인화 검색 성능의 평가를 위한 실험집단과 통제집단의 구성은 동일인이며 단, 논리적 통제집단을 NTIS 서비스 사용자로 가정하였다.

5.1 평가 방법

과학기술정보 개인화 서비스의 성능 평가를 위한 관심분야 추적결과 정확성, 개인화 정보의 정확성, 관심 태그 클라우드 유용성, 관심분야 기반 추천정보 유용성과 서비스 개선 중심 질적 평가는 아래의 방법으로 이루어졌다. 검

색어에 대한 검색 정확성 보다는 결과로 제시되는 문서가 사용자에게 얼마나 유용한지에 대한 문서 적합성 평가로 개인화 서비스의 효과를 입증하기 위해 다양한 검색성능 평가지표를 활용해 평가하고 그 결과를 분석하였다.

- ① 평가대상 전문가 12명으로 하여금 일정기간 (2017.6.1.부터 2017.9.8.까지 약 100일간) 과학기술정보 개인화 서비스를 이용하도록 하였다. 개인 관심분야 변화 추이에 대한 정보를 수집하고 제공하기 위해 컴퓨터 및 정보학 분야 전문가에게는 아래의 바이오/의학 분야 용어를 간헐적으로 질의어로 활용하고, 바이오/의학 분야 문서를 이용토록 권고하였다. 반대로 바이오 및 생명의료 분야 전문가에게는 컴퓨터공학 및 IT분야의 용어를 간헐적으로 질의어로 활용하고, 해당 분야 문서를 이용토록 권고하였다. 하지만 해당 기간 동안 자신의 연구에 필요한 과학기술정보를 NTIS 대신 제안한 서비스에서 찾아 이용하도록 하였다.

구분	사용 권고 질의어
바이오 및 의학 분야	알레르기, 유전자, cancer, 세균, 세포배양, 마취, RNA, 관절염, 진단키트, 뇌신경, 염증, 항체, 혈액, 아토피 등
컴퓨터 및 IT 분야	인공지능, IoT, 반도체, 다이오드, MPEG, SSD, 메모리, 트랜지스터, 블록체인, 암호화, LTE, 안드로이드 등

<표 9> 과학기술정보 개인화 성능평가 대상 전문가

구분	컴퓨터 및 정보학 분야 연구자	생명공학, 바이오/의료 분야 연구자
출연(연) 연구자	3	1
대학 연구자	2	2
NTIS이용자	1	1
전문의	0	2
합계	6	6

- ② 서비스 사용기간 후반에 '과학기술정보 개인화 성능 평가지'를 12명의 전문가에게 배포 (2017.8.28.)하였다.
- ③ 평가 질문을 통하여 관심분야 추적결과의 정확성에 대한 응답을 수집하였다. 서비스를 이용하는 시점마다의 본인의 관심영역 또는 연구영역과 개인화 서비스가 추적하여 제시한 관심분야 간 부합도를 5점 리커트 척도로 평가하였다.
- ④ 평가 질문을 통하여 관심 태그 클라우드 유용성을 만족도 기반의 5점 리커트 척도로 평가하였다.
- ⑤ 평가 질문을 통하여 관심분야 기반 추천정보 유용성을 만족도 기반의 5점 리커트 척도로 평가하였다.
- ⑥ 평가 질문을 통하여 과학기술정보 개인화 서비스의 검색 결과 정확성과 검색 결과 문서의 적합성을 판단하였다. NTIS와 제안된 개인화 서비스 간 비교평가를 위해 평가시점 이전 3개월간 NTIS에서 가장 많이 사용된 질의어, 여러 연구 분야에서 공통으로 널리 활용되는 단어와 중의성을 가진 단어를 포함한 10개 질의어(바이러스 감염, 로봇, cell, 인공지능, 빅데이터 분석, 치매 진단, 머신러닝, 사물인터넷, 충돌, program)를 제시하였다.
- ⑦ 마지막으로, 본 연구의 과학기술정보 개인화 서비스와 NTIS에 대한 각각의 선호 여부, 이유, 장단점, 개선요청사항을 조사하여 질적 평가를 수행하였다.

5.2 평가 결과 및 분석

과학기술정보 개인화 서비스의 성능 평가를 위한 관심분야 추적결과 정확성, 개인화 정보의 정확성, 관심 태그 클라우드 유용성, 관심분

야 기반 추천정보 유용성과 서비스 개선 중심 질적 평가를 수행하였다. 특히, 검색어에 대한 검색 정확성 보다는 결과로 제시되는 문서가 사용자에게 얼마나 유용한지에 대한 문서 적합성 평가로 개인화 서비스의 효과를 입증하기 위해 다양한 검색성능 평가지표를 활용해 평가하고 그 결과를 분석하였다.

먼저, 전문가 자신의 연구분야 및 관심분야에 개인화 서비스가 추적하여 제시한 관심분야가 어느 정도 부합하는지로 정확성을 판단하였다. 그리고 제시한 관심 태그 클라우드에 대해 얼마나 만족하는지와 추천정보를 얼마나 만족하는지로 그 유용성을 측정하였다. <표 10>에서 보는 바와 같이 관심분야 추적의 정확성은 4.583점, 관심 태그 클라우드의 유용성은 4.083점, 추천정보의 유용성은 3.417점으로 측정되었다. 이는 100점으로 환산하면 각각 92점, 82점, 68점 정도이다. 표에서와 같이 각 지표 내 평가점수 간 표준편차는 0.5점에서 0.7점 내로 매우 작은 값으로써 모든 평가자가 정확성과 유용성에 높은 점수를 부여하였다. 이는 사용자 정보 이용 행태를 분석하여 추적한 개인별 관심 연구분야가 정확성을 가진 것이며, 추적된 관심분야 기반으로 제공된 태그 클라우드와 추천 과학기술정보가 사용자에게 유용하게 활용된다는 것을 입증한다.

검색 성능 개선은 개인의 관심분야와 질의어에 적합한 문서를 검색 결과 상단에, 그리고 최대한 많은 적합 문서를 제공하는데 목적을 둔다. 따라서 이용자가 관심분야와 질의어에 적합하다고 판단하는 문서를 정답 문서로 가정하여 검색 결과 정확성을 측정하였다. Precision@10을 측정하여 정답 문서가 얼마나 많이 상위 10위

〈표 10〉 관심분야 추적 정확성 등 기능 유용성 평가 결과

	관심분야 추적 정확성	관심 태그 클라우드 유용성	관심정보 기반 추천정보 유용성
평균	4.583	4.083	3.417
표준편차	0.515	0.669	0.669

안에 포함되는지, MAP를 측정하여 얼마나 많은 정답 문서가 상위 10위 안에서도 상단에 랭크되는지를 평가하였다. 그리고 제시된 문서가 얼마나 사용자에게 유용한지에 대한 인지 수준을 적용하여 관련도 높은 문서가 얼마나 많이 상위에 분포하는지를 판단하는 DCG와 nDCG를 측정하여 검색 성능을 평가하였다.

Precision@K는 검색 결과 상위 K개 가운데 정답 문서가 몇 개인지로 측정한다. 12명의 평가자는 각 10개의 평가 질의어에 대해 검색을 수행한 후 검색 시점의 본인 관심분야, 입력한 질의어, 검색결과 문서들 간 모두 관련성이 있는 문서를 T(true)로, 아닌 문서를 F(false)로 판단하였다. NTIS에서는 평균 상위 10개의 결과 문서 중 3.59개의 문서만이 정답으로 판단되었고, pNTIS¹⁰⁾에서는 평균 상위 10개의 결과 문서 가운데 정답이 6.91개 정도로 식별되었다. 〈표 11〉은 질

의별 Precision@10 평가 결과와 NTIS와 pNTIS 비교 시 향상률을 나타낸다. Precision@10의 성능은 결과적으로 NTIS 대비 pNTIS가 0.332(92.47%) 향상되었다.

특히, ‘바이러스 virus’, ‘cell’, ‘충돌’, ‘program’ 질의어는 중의성을 가지거나 여러 연구 분야에 널리 활용되는 단어이다. 특히 ‘바이러스 virus’와 ‘cell’은 일반적으로 생물학 및 바이오, 유전학 정보에 많이 포함된 단어이며, ‘충돌’, ‘program’은 컴퓨터 및 정보학 외에서 더 많이 쓰인다. 때문에 컴퓨터 및 정보학의 관심분야를 둔 평가자 6명의 경우 해당 질의어에 대해 〈표 12〉와 같이 높은 향상률을 보였다. 이는 NTIS에서는 해당 질의어로 검색되지 않았던 문서가 개인 관심분야를 반영하여 개인화 정보를 제공하는 pNTIS에서는 검색 결과 상위 10위안에 5개 이상 더 많이 제시되었다는 것을 나타낸다.

〈표 11〉 Precision@10 평가 결과 및 향상률

	질의어									
	바이러스 virus	로봇	cell	인공 지능	박테이터 분석	치매 진단	머신러닝	사물인터넷	충돌	prog-ram
NTIS	0.300	0.358	0.333	0.383	0.417	0.483	0.433	0.408	0.158	0.317
	(평균) 0.359									
pNTIS	0.600	0.733	0.825	0.658	0.725	0.642	0.633	0.658	0.683	0.750
	(평균) 0.691									
질의별 향상률	0.300	0.375	0.492	0.275	0.308	0.158	0.200	0.250	0.525	0.433
	(평균) 0.332									

10) pNTIS(personalized NTIS)는 본 연구에서 제안한 과학기술정보 개인화 서비스의 약칭으로 정의한다.

〈표 12〉 컴퓨터 및 정보학 전문가에 대한 특정 질의어 Precision@10 향상률

	바이러스 virus		cell		충돌		program	
	NTIS	pNTIS	NTIS	pNTIS	NTIS	pNTIS	NTIS	pNTIS
평가자1	0.100	0.600	0.000	0.700	0.200	0.600	0.100	0.700
평가자2	0.000	0.300	0.100	0.900	0.300	0.700	0.100	0.800
평가자3	0.100	0.500	0.200	0.800	0.200	0.800	0.100	0.900
평가자4	0.000	0.600	0.100	0.500	0.100	0.500	0.300	0.600
평가자5	0.000	0.200	0.000	1.000	0.400	1.000	0.000	0.800
평가자6	0.200	0.200	0.000	1.000	0.400	0.900	0.400	0.900
평균	0.067	0.400	0.067	0.817	0.267	0.750	0.167	0.783
질의별 향상률	0.333		0.750		0.483		0.617	
전체 향상률	0.546							

Precision@K 평가 시의 정답 문서들이 검색 결과 상위 10위 내에서 얼마나 상단에 위치하고 있는지를 판단하고자 MAP를 측정하였다. MAP는 10개의 질의어별 상위 10개 검색결과 중 정답 문서로 체크된 순위 위치 $n(n=1\sim 10)$ 에서의 Precision@n을 모두 구하여 평균을 구한 후 그것들의 전체 평균으로 계산한다. 〈표 13〉은 MAP 평가 결과와 향상률을 나타낸다. 평가 결과 NTIS 대비 pNTIS가 0.312(63.03%) 향상되었다.

마찬가지로 ‘바이러스 virus’, ‘cell’, ‘충돌’,

‘program’ 질의어는 중의성을 가지거나 여러 연구 분야에 널리 활용되는 단어이다. 또한, ‘바이러스 virus’와 ‘cell’은 일반적으로 생물학 및 바이오, 유전학 정보에 많이 포함된 단어이며, ‘충돌’, ‘program’은 컴퓨터 및 정보학 외에서 더 많이 쓰인다. 때문에 ‘바이러스 virus’는 NTIS 대비 pNTIS의 MAP가 0.419(111.14%) 향상, ‘cell’은 0.437(89.73%) 향상, ‘충돌’은 0.470(135.83%), ‘program’은 0.434(102.6%)의 높은 향상률을 보였다.

〈표 13〉 MAP 평가 결과 및 향상률

		NTIS	pNTIS	향상률
질의어	바이러스 virus	0.377	0.796	0.419
	로봇	0.421	0.831	0.410
	cell	0.487	0.924	0.437
	인공지능	0.460	0.732	0.272
	빅데이터 분석	0.509	0.789	0.280
	치매 진단	0.624	0.716	0.092
	머신러닝	0.577	0.758	0.181
	사물인터넷	0.731	0.854	0.123
	충돌	0.346	0.816	0.470
	program	0.423	0.857	0.434
평균	0.495	0.807	0.312	

검색 결과로 제시된 문서나 정보가 사용자에게 얼마나 유용한지에 대한 인지 수준을 평가하는 것은 문서가 단순히 정답인지 아닌지(관련성이 있는지 아닌지)를 판단함을 넘어 문서 적합성 차원의 검색 성능을 판단하는데 중요한 요소이다. 검색 결과 상위 10건의 문서들 중 적합성 점수(4:적합, 3:적합가능, 2:중간, 1:부적합가능, 0:부적합)가 높은 문서가 얼마나 많은지를 판단하기 위해 DCG를 측정하였고, 적합성 점수가 높은 문서가 얼마나 상위에 랭크되어 있는지를 판단하기 위해 DCG를 IDCG(이상적인 문서 적합성 점수 순위에서의 DCG: IdealDCG)로 정규화 한 nDCG를 측정하였다.

〈표 14〉는 질의별 DCG 평가 결과와 NTIS와 pNTIS 비교 시 향상률을 나타낸다. DCG의 성능은 NTIS 대비 pNTIS가 6.08(128.54%) 향상되었다. 특히 '바이러스 virus', 'cell', '충돌', 'program'과 같이 중의성을 가지거나 여러 분야에서 활용되는 단어로 검색했을 때 적합성이 높은 문서가 검색되었으며, '바이러스'와 'cell'의 경우 컴퓨터 및 정보학 전문가의 경우 더 높은 향상률을 보였다. 이는 해당 단어로 검색할 때 유전학 및 생물학 분야의 문서가 많이 검색될

뿐만 아니라 상위로 제시되었던 것이 pNTIS에서 검색하면 컴퓨터 및 IT분야에서 의미하는 '바이러스', 'cell'을 포함한 검색 결과가 상위에 제시되기 때문이다.

앞에서 DCG로 문서 적합성 점수가 높은 검색 결과가 얼마나 많이 상위 10위 안에 포함되었지를 평가하였다. 이 때 적합성 점수를 가진 문서들이 10위 내에서도 상위에 랭크되어 있는지를 평가하기 위해 nDCG를 산출하였다. 〈표 15〉는 질의별 nDCG 평가 결과와 NTIS와 pNTIS 비교 시 향상률을 나타낸다. nDCG의 성능은 NTIS 대비 pNTIS가 0.24(40.67%) 향상되었다.

본 연구에서 제안한 과학기술정보 개인화 서비스와 동일한 정보로 대국민에 서비스 중인 NTIS 간의 선호도와 그 이유를 조사하고, 각 서비스의 장단점 및 개선사항을 도출하고자 질적 평가를 수행하였다. 평가자의 피드백 내용을 반복하여 읽으면서 내용을 분석하여 유사한 것끼리 그룹화하여 의미단위로 표현하였다. 〈표 16〉은 과학기술정보 개인화 서비스와 NTIS의 장단점 평가결과 및 개선 요청사항을 나타낸다. 평가결과 모든 사용자가 NTIS보다 제안된 서비스를 더 선호하는 것으로 나타났다.

〈표 14〉 DCG 평가 결과 및 향상률

	질의어									
	바이러스 virus	로봇	cell	인공지능	빅데이터 분석	치매 진단	머신러닝	사물인터넷	충돌	program
NTIS	3.43	4.05	4.97	4.39	5.30	6.64	6.42	5.76	2.53	3.76
	(평균) 4.73									
pNTIS	10.96	10.79	13.52	10.77	11.31	10.12	10.11	10.19	10.13	10.25
	(평균) 10.81									
질의별 향상률	7.53	6.74	8.55	6.38	6.01	3.48	3.69	4.43	7.60	6.49
	(평균) 6.08									

〈표 15〉 nDCG 평가 결과 및 향상률

	질의어									
	바이러스 virus	로봇	cell	인공 지능	빅데이터 분석	치매 진단	머신 러닝	사물 인터넷	충돌	prog- ram
NTIS	0.46	0.54	0.47	0.58	0.64	0.70	0.62	0.83	0.47	0.55
	(평균) 0.59									
pNTIS	0.84	0.85	0.94	0.80	0.81	0.77	0.79	0.86	0.83	0.83
	(평균) 0.83									
질의별 향상률	0.38	0.31	0.47	0.22	0.17	0.07	0.17	0.03	0.36	0.28
	(평균) 0.24									

〈표 16〉 NTIS와 pNTIS 사용성에 따른 질적 평가

구분		의미단위
pNTIS	장점	<ul style="list-style-type: none"> • 개인화된 과학기술정보 제공 • 중의어로 검색 시 적합성 높은 문서 제시 • 검색포털과 같이 검색중심의 기능이 친숙하고 편리함 • 개인 관심분야 비율 가시화의 유용함
	단점	<ul style="list-style-type: none"> • 로그인해야만 관심분야가 추적됨 • 논문, 연구보고서 원문 이용시 NTIS 로그인 불편
	바라는 점	<ul style="list-style-type: none"> • 딥러닝 등을 이용한 이용자 요구 파악 기능 • 과제정보 이용시 위탁, 공동연구 현황정보 제공 요청 • 관심분야를 언제든지 초기화 할 수 있는 기능 필요
NTIS	장점	<ul style="list-style-type: none"> • 과학기술정보 다양성 및 통계자료 제공
	단점	<ul style="list-style-type: none"> • 메뉴의 복잡성 • 모든 이용자에게 동일한 검색결과 • 원하는 정보 탐색이 어려움 • 회원가입 시 많은 정보 요구
	바라는 점	<ul style="list-style-type: none"> • 내가 관심있는 키워드를 제시했으면 함 • 검색기능 중심의 개편 필요 • 개인 맞춤형 과학기술정보 제시 필요

5.3 평가결과 요약

사용자의 과학기술정보서비스 이용행태를 분석하여 정보를 검색하거나 이용하는 시점의 개인 연구분야 혹은 관심분야를 추적해 제공한 ‘개인 관심분야 추적결과’의 정확성은 5점 만점에 4.6점으로 매우 높은 것으로 나타났다. 또한 검색을 수행하지 않고 제시된 키워드에 접근함으로써 정보를 탐색할 수 있는 기능인 ‘관심 태

그 클라우드’는 5점 만점에 4.1점으로 유용성이 높게 평가되었다. 그리고 이용자 관심분야와 최근 검색어 기반으로 제공된 ‘관심분야 기반 추천정보’는 5점 만점에 3.4점으로 다소 유용한 것으로 나타났다.

다음으로 개인화 검색 성능을 평가하고자 검색성능 평가 지표로 널리 활용되고 있는 Precision@K, MAP, DCG, nDCG를 각각 산출하여 평가하고 분석하였다. 분석 결과 모든 평가지표에서 과학

기술정보 개인화 서비스가 비교대상인 NTIS 대비하여 점수가 높은 것으로 나타났고, 이는 관심분야 정보를 이용해 검색결과를 재순위화하여 제공한 정보의 유용성과 만족도가 높다는 의미이다. 즉, 검색 결과 상위 10위 안에 적합도가 높은 문서가 랭크되어 사용자가 원하는 정보를 쉽게 접근할 수 있음을 입증하였다. 중의어나 여러 연구분야에 혼용하여 사용되는 단어를 질의어로 사용할 경우 검색결과 적합성 향상률이 매우 높았다.

마지막으로 관심분야에 적합한 문서가 검색상단에 먼저 제시되어 개인화 서비스를 더 선호하며, 검색 기능 위주의 편리함이 우수하다는 질적 평가를 얻었다. 반대로 동일한 질의어 입력 시 모든 사용자에게 동일한 검색결과를 제공하는 NTIS는 원하는 정보를 탐색하는데 시간과 탐색 노력이 많이 들어 불편하다는 의견이 나타났다.

6. 결 론

증가하는 과학기술정보의 양만큼 적합한 정보를 쉽게 찾아 이용하고 싶은 사용자 요구도 증가하고 있다. 이러한 사용자의 정보 요구를 충족시키기 위해 과학기술정보를 제공하는데 있어 단순히 질의어와 문서 간의 적합도만으로 정보를 제공하는 서비스 보다는 사용자의 검색 의도 혹은 관심분야를 파악하여 적합한 정보를 이용하기 쉬운 형태나 검색 순위에 제시하는 것이 필요하다.

본 연구에서는 과학기술정보 사용자를 위해 연구자 성향이나 관심분야에 부합하고 적합성

이 높은 정보를 제공하기 위한 국가과학기술표준분류기반의 관심분야 추적 및 개인화 정보 서비스를 설계, 구현, 평가하였으며, 그 효율성을 입증하였다. 개인 관심분야 추적을 위해 다양한 사용자 이용행태를 파악하여, 행태별 정보에 대한 관심점수를 설문 조사하여 알고리즘을 설계하였고, 미리 정의된 사용자 프로파일 없이 사용자의 서비스 이용 행태만으로 관심분야를 추적 및 결정하였다. 그리고 결정된 개인의 관심분야를 반영하여 과학기술정보를 개인화하여 제공하였다. 개인화 검색 성능 평가를 위해 전문가를 선정하였고, 중의성을 가지거나 여러 연구 분야에서 활용되는 단어를 질의어로 구성하여 검색 결과 적합성을 평가하였다. 제안된 개인화 서비스와 NTIS의 검색 성능 비교 평가를 수행하였으며, 그 결과 제안된 개인화 서비스에서 제공하는 검색 성능이 더 우수한 것으로 나타났다. 그 결과를 정리하면 다음과 같다.

첫째, 서비스 이용행태를 추적하여 결정한 관심분야는 사용자가 정보를 이용하는 시점의 연구영역이나 관심영역과 부합하는 것으로 나타났다.

둘째, 관심 태그 클라우드 제공 기능은 개인의 관심 키워드를 먼저 제시하는 기능으로써, 원하는 과학기술정보를 쉽게 탐색할 수 있도록 도와주는 것으로 평가되었다.

셋째, 관심분야 기반 추천정보는 질의어를 입력하지 않고도 맞춤형 과학기술정보를 이용하기 쉽게 제공한다는 측면에서 유용한 것으로 나타났다.

넷째, 추적된 관심분야 기반으로 검색결과를 개인화하여 제공하였고, 비교 대상 서비스보다

검색 적합성이 우수한 것으로 나타났다. 특히 중의성을 가지거나 여러 연구 분야에 서로 다른 의미로 활용되는 단어를 질의어로 사용할 경우 연구 영역이 서로 다른 사용자마다 최적화된 검색결과를 제공하였으며 사용자 요구에 적합한 정보를 검색 결과 상단에 효과적으로 제공한다는 것이 입증되었다.

다섯째, 질적 평가 결과 과학기술정보 탐색의 관점에서 평가자는 검색 기능 중심의 제안

된 서비스를 NTIS보다 선호하는 것으로 나타났다. 무엇보다도 원하는 정보를 검색결과 첫 페이지 내에서 찾을 수 있도록 하는 정보 개인화에 만족하였다. 더불어 과학기술정보를 개인화 하는데 있어서 유사 연구 분야의 사용자 서비스를 이용한 행태와 로그정보, 또는 키워드의 주제별 군집화 및 딥러닝을 통해 사용자 질의어를 미리 예측하는 연구 방법에 대한 제안도 존재하였다.

참 고 문 헌

- [1] 김광영, 심강섭, 박승진. 2009. 분류와 사용자 질의어 정보에 기반한 개인화 검색시스템. 『한국문헌정보학회지』, 43(3): 163-180.
- [2] 김동욱. 2011. 『사용자 프로파일에 기반한 개인화 웹 검색 시스템 설계』. 석사학위논문, 한양대학교 대학원 전자컴퓨터통신공학과.
- [3] 김한준, 노준호, 장재영. 2012. 웹 검색 개인화를 위한 개념네트워크 프로파일 기반 순위 재조정 기법. 『한국인터넷방송통신학회논문지』, 12(2): 69-76.
- [4] 소영준, 박영택. 2000. 사용자 프로파일 기반 개인 웹 에이전트. 『정보과학회논문지: 소프트웨어 및 응용』, 27(3): 248-256.
- [5] 윤성희. 2016. 질의기반 사용자 프로파일을 이용하는 개인화 웹 검색. 『한국산학기술학회논문지』, 17(2): 690-696.
- [6] 이준, 신동욱, 최종민. 2009. 사용자 행동 예측을 이용한 개인화 기반 검색결과 순위 재조정. 『한국정보과학회 한국컴퓨터종합학술대회 논문집』, 36(1B): 354-359.
- [7] 이현재. 2012. 『연구자료의 검색효율 증대를 위한 개인화 검색 기법에 대한 연구: 대학도서관 자료 검색을 중심으로』. 석사학위논문, 고려대학교 교육대학원 컴퓨터교육학과.
- [8] 이현주, 김한일. 2008. 태그 클라우드를 이용한 개인화된 검색 시스템 설계. 『한국멀티미디어학회 추계학술발표논문집』, 165-168.
- [9] 허정, 서희철, 장명길. 2006. 상호정보량과 복합명사 의미사전에 기반한 동음이의어 중의성 해소. 『정보과학회논문지』, 33(12): 1073-1089.
- [10] Bouras, C., Pouloupoulos, V. and Silintziris, P. 2009. "Personalized News Search in www:

- Adapting on User's Behavior." *Internet and Web Applications and Services, Fourth International Conference on*, IEEE, 125-130.
- [11] Liang, C. 2011. "User Profile for Personalized Web Search." *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, IEEE, 1847-1850.
- [12] Pitcow, J. et al. 2002. "Personalized Search." *Communications of the ACM*, 45(9): 50-55.
- [13] Shen, X., Tan, B. and Zhai, C. 2005. "Implicit user modeling for personalized search." *Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM, 824-831.
- [14] Xing, K. et al. 2011. "Behavior based User Interests Extraction Algorithm." *Internet of Things (iThings/CPSCoM), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing*, IEEE, 448-452.
- [15] Xu, S. et al. 2014. "A Dynamic Users' Interest Discovery Model with Distributed Inference Algorithm." *International Journal of Distributed Sensor Networks*, 10(4), 280892.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kim, Kwang-Young, Shim, Kang-Seop and 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.
- [2] Kim, Dongwook. 2011. *Design of Personalized Web Search System based on User Profile*. M.A. thesis, Hanyang University Graduate School.
- [3] Kim, Han-joon, Noh, Joonho and Chang, Jaeyoung. 2012. "A New Re-ranking Technique based on Concept-Network Profiles for Personalized Web Search." *The Journal of the Institute of Internet, Broadcasting and Communication*, 12(2): 69-76.
- [4] So, Young-Jun and Park, Young-Tack. 2000. "User Profile based Personalized Web Agent." *Journal of KIISE*, 27(3): 248-256.
- [5] Yoon, Sung Hee. 2016. "Personalized Web Search using Query based User Profile." *Journal of the Korea Academia-Industrial Cooperation Society*, 17(2): 690-696.
- [6] Lee, Joon, Shin, Dongwook and Choi, Joongmin. 2009. "Personalized Search Result Re-ranking Using User Behavior Prediction." *The Korean Institute of Information Scientists and Engineers, Korea Computer Conference*, 36(1B): 354-359.
- [7] Lee, Hyunjae. 2012. *A Study on the Personalized Retrieval Method for Increasing Search*

Efficiency of Research Data: Focusing on university library materials search. M.A. thesis, Korea University Graduate School.

- [8] Lee, Hyun-Joo and Kim, Han-il. 2008. "Design for Personalized Search System Using Tag Cloud." *Proceedings of the Korea Multimedia Society Conference*, 165-168.
- [9] 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): 1073-1089.

