

논문의 중요성 및 품질을 이용한 학술 전문가 검색 기법

Academic Expert Search Method Using Importance and Quality of Papers

이서희*, 박윤정*, 한진수**, 최도진**, 임종태**, 복경수**, 유재수**
충북대학교 빅데이터협동과정*, 충북대학교 정보통신공학과**

Seo-Hee Lee(shl@cbnu.ac.kr)*, Yun-jeong Park(yunjeongp@cbnu.ac.kr)*,
Jin-Su Han(jinsu@cbnu.ac.kr)**, Do-Jin Choi(ycdj91@cbnu.ac.kr)**,
Jong-Tae Lim(jtlim@cbnu.ac.kr)**, Kyoung-Soo Bok(ksbok@cbnu.ac.kr)**,
Jaesoo Yoo(yjs@cbnu.ac.kr)**

요약

사용자들에게 필요한 대표적인 연구 결과물과 조언을 제공할 수 있는 대용량 학술 정보를 이용하여 특정 관심 분야의 전문가를 검색하는 기법에 대한 연구가 요구되고 있다. 기존의 전문가 검색 기법은 사용자 프로필, 최근 활동 분석을 기반으로 전문가를 검색하기 때문에 사용자의 프로필 또는 활동 정보를 파악하지 못할 경우 전문가를 판별하기 어려운 문제점이 있다. 본 논문에서는 논문의 중요성 및 품질을 이용한 전문가 검색 기법을 제안한다. 논문의 중요성은 논문의 희소성과 최근 이슈가 되는 토픽을 고려하여 계산한다. 논문의 품질은 인용 수, 저널의 IF, 최신성, 저자관계를 고려하여 논문의 영향력을 평가한다. 제안하는 기법의 우수성을 보이기 위해 기존 기법과 정확률과 재현율 관점에서 성능평가를 수행한다.

■ 중심어 : | 전문가 검색 | 논문 품질 | 논문의 중요성 | 학술 정보 | 학술 검색 |

Abstract

An expert search method using a large amount of academic data that can provide users with representative research results and advice is required. Since the existing expert search methods perform the expert search based on user profile or activity information, they have a problem that it is hard to discriminate the expert when we do not know the user profile or activity information. In this paper, we propose an academic expert search method using the importance and quality of a paper. The importance of a paper is computed by considering its scarcity and up-to-date topics. The quality of a paper is evaluated by considering the number of citations, IF of Journal, recency and author relations. To show the superiority of the proposed method, we compare it with the existing scheme through the performance evaluation in terms of recall and precision

■ keyword : | Expert Search | Paper Quality | Paper Importance | Academic Information | Academic Search |

* 본 연구는 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원 (No.2015R1D1A3A01015962), 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업(IITP-2016-H8501-16-1013) 및 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2016R1A2B3007527)

* 본 논문은 한국콘텐츠학회 2016 춘계 종합학술대회 우수논문입니다.

접수일자 : 2016년 10월 25일

심사완료일 : 2016년 11월 09일

수정일자 : 2016년 11월 09일

교신저자 : 유재수, e-mail : yjs@cbnu.ac.kr

1. 서론

정보통신 혁명과 인터넷의 발달로 새로운 과학 기술은 우리 생활의 여러 분야에서 끊임없는 발전을 계속하고 있으며, 다양한 분야의 연구가 지속적으로 수행되고 있다. 일반적으로 새로운 과학 기술의 연구를 처음 수행하기 위해서 대부분의 사용자는 기존의 많은 연구 결과물을 활용한다. 기존의 연구 결과물을 이용하는 방법 중 하나는 다양한 학술 검색사이트를 이용하는 것이다[1-3]. 학술 검색 사이트는 사용자가 기술의 진보에 따른 새로운 기술을 적용하거나 현재의 기술을 개선, 발전시킬 수 있도록 기존의 수많은 연구 결과물을 제공한다.

사용자들은 이러한 연구 결과물 중 신뢰할 수 있는 대표적인 연구 결과물을 활용하고, 이를 바탕으로 새로운 연구의 기반을 마련하기를 원한다. 그러나 기존의 다양한 학술 검색 사이트에서 제공하는 학술 정보는 방대해서 관심 분야의 가장 대표적인 연구 결과물을 찾기가 쉽지 않다. 따라서 사용자들에게 필요한 대표적인 연구 결과물과 조연을 제공할 수 있는 대용량 학술 정보를 이용한 특정 관심 분야의 전문가 검색 기법에 대한 연구가 요구되고 있다.

기존 전문가 검색 기법은 소셜 네트워크 환경에서 유사한 사용자의 프로필을 이용하는 기법과 사용자의 최근 활동정보를 분석해 전문가를 검색하는 기법이 대부분이다[4-8]. 유사한 사용자의 프로필을 이용하는 기법은 질의와 관련된 키워드가 많이 연결되는 사용자가 전문가로 검색되며 시간과 비용은 적게 들지만 정확도가 낮은 문제점이 있다[4]. 사용자의 최근 활동 정보를 분석해 전문가를 검색하는 기법은 사용자가 최근에 남긴 글과 댓글, 평점 등을 고려하여 전문가를 검색한다[9-12]. 이러한 전문가 검색 기법은 사용자의 프로필 또는 활동 정보를 알지 못하는 경우에 검색하는데 어려움이 존재한다. 또 다른 전문가 검색 기법은 특정 분야 논문의 검토자를 검색하기 위해 인용 수, IF, 최신성을 이용해 전문가를 판단한다[13-15]. 그러나 이 기법은 인용수와 저널의 IF, 최신성만으로 전문가의 전문성을 판단하기 때문에, 저자 관계에 상관없이 논문을 많이

게한 경우에 높은 점수를 부여받게 되는 문제점이 존재한다[16].

본 논문에서는 논문의 중요성 및 품질을 이용한 전문가 검색 기법을 제안한다. 논문의 중요성은 논문의 회소성과 최근 이슈가 되는 토픽을 고려하여 계산한다. 논문의 품질은 인용 수, 저널의 IF, 최신성, 저자관계를 고려하여 논문의 영향력을 평가한다. 특히 사용자가 원하는 분야의 전문가를 검색하기 위해 다수의 유사한 키워드들을 단일의 키워드로 판단할 수 있도록 온톨로지를 구성한다. 또한 연구 결과물의 중요성 및 품질을 기반으로 각 분야별 지수를 도출 한 후, 이를 통해 랭킹 결과를 얻어 사용자에게 적합한 전문가검색 결과를 제공한다.

본 논문의 구성은 다음과 같다. II장에서는 기존에 관련된 연구와 기존의 전문가 검색 기법에 대해 분석하고 III장에서는 전체 전문가 시스템의 구조와 본 논문에서 제안하는 전문가검색 기법의 각 모듈의 역할에 대해 설명한다. IV장에서는 본 논문에서 제안하는 전문가 검색 기법과 기존의 전문가 검색 기법과의 성능평가를 실시한다. V장에서는 본 논문에 대한 결론 및 향후 연구방향을 제시한다.

II. 관련연구

1. 학술 검색 사이트

구글 스칼라(<https://scholar.google.co.kr/>)는 가장 많은 사람들이 이용하는 학술 검색 사이트이다. 구글 학술 검색은 구글에서 제공하는 검색 서비스 중 하나로 논문, 학술지, 간행물 등의 검색을 수행 한다. 구글 스칼라는 정렬 알고리즘을 사용해 낱순, 관련도 순으로 검색을 이용할 수 있다[1].

DBpia(<http://www.dbpia.co.kr/>)는 국내 학술저널, 학술대회, 전문잡지, 전자책 등을 제공하는 온라인 서비스로 논문 원문 및 상세한 서지정보를 검색 열람할 수 있다[2]. 최근에는 DBpia Insight를 통해 연관논문 추천 서비스, 공저자/연관 저자 추천 서비스, 저널의 인용/피인용 저널 추천 서비스, 저널 내 영향력 있는 저자 추천

서비스와 같은 다양한 서비스를 제공하고 있다. 공저자/연관 추천 서비스는 저자의 논문을 함께 저술한 공동 저자와 저작 논문의 추천 논문을 기반으로 연관 저자를 추천해 주는 서비스이다. 저널 내 영향력 있는 저자 추천 서비스는 저널 내 논문의 발행, 인용, 인용 정보를 기반으로 영향력 있는 저자를 추천하는 서비스이다.

KCI 한국학술지인용색인 (<https://www.kci.go.kr>) 는 국내 학술지정보, 논문정보 및 참고문헌을 DB화 하여 논문 간 인용관계를 분석하는 시스템이다[3]. KCI는 국내 학술지 및 게재 논문에 대한 각종 학술정보의 제공과 함께 연구자원 관리에 필요한 각종 통계자료와 인용빈도에 따른 학술지의 영향력을 산출하여 특정 주제분야에서 발행되는 학술지의 질적인 수준을 평가할 수 있는 하나의 도구로 활용될 수 있다. KCI에서는 논문 원문과 저자정보, 인용정보, 통계정보, 학술지정보, 학회정보를 제공하고 있다.

2. 전문가 검색 기법

가장 기본적인 전문가 검색 기법은 소셜 네트워크 환경에서 유사한 사용자의 프로필을 이용하는 기법이다 [4-8] 기법은 사용자의 프로필을 수집한 후, 사용자와 유사한 프로필을 가지는 사용자의 성향을 이용하여 검색을 수행한다. 일반적으로 사용자의 성향은 시간에 따라 점차 변화되지만 프로필을 변경하는 사용자는 거의 존재하지 않는다. 따라서 프로필이 유사한 사용자 성향을 이용하여 검색할 경우 실제 사용자의 최신 성향과 다른 결과가 생성될 수 있다. 또한, 사용자가 기술한 프로필을 이용하여 전문가를 검색할 경우, 검증되지 않은 허위정보를 기재할 가능성이 있어서 검색 결과의 신뢰성이 저하된다.

프로필 기반의 전문가 검색 기법의 문제점을 해결하기 위해 최근에는 사용자의 최근 활동 정보를 분석해 전문가를 검색하는 기법이 많이 연구되고 있다[9-12]. 소셜 활동 정보 기반의 전문가 검색 기법은 소셜 네트워크 환경에서 사용자가 SNS에서 활동한 행위, 게시물, 댓글, 다운로드 등과 같은 리소스들을 분석하여 전문가를 판별한다. 이 기법에서는 소셜 네트워크에 가입된 사용자가 질의를 요청하면 서버는 질의를 요청한 사용

자와 전문가 후보자들을 연결한다. 사용자와 친구 관계에 있는 사용자들을 전문가 후보로 간주하고 사용자들이 남긴 글이나 프로필을 분석한 후 전문가 후보자들의 전문지식을 판단한다. 전문지식 점수에 따라 순위를 부여해 질의를 요청한 사용자에게 검색 결과를 제공한다. 그러나 이러한 전문가 검색기법은 사용자의 성향을 판별하기 위해 많은 시간이 소요되며, 사용자의 프로필 또는 활동 정보가 부족한 경우에 검색하는데 어려움이 존재한다.

최근 논문의 정보 중 인용 수, IF, 최신성을 이용해 전문가를 계산하는 기법이 연구되었다[12]. 논문이 학회나 저널에 출판되기 위해선 그 분야의 저명한 전문가에게 검토를 받아야 한다. 이 기법에서는 검토할 논문의 제목 및 초록, 참고문헌으로부터 키워드를 추출해 논문의 토픽을 파악한 후, 연구 분야의 관련성을 파악하고 전문가의 전문성을 판별하기 위해 인용 수, IF, 최신성을 이용한다. 그러나 인용 수, IF, 최신성만 고려할 경우 전문가의 전문성을 판단하는데 어려움이 존재한다. 이 기법에서는 저자관계에 관계없이 논문을 많이 게재한 경우에 높은 점수를 부여받게 되는 문제점이 존재한다. 따라서 각 분야별 전문가를 판별하기 위해 논문의 중요성 및 품질을 고려한 기법이 필요하다.

III. 제안하는 전문가 검색 기법

1. 제안하는 기법의 특징

기존의 학술 검색 사이트는 특정 분야에서 어떤 연구가 최근에 떠오르고 있는지, 어떤 결과물이 중요하고 믿을 만한 연구인지 알기 위해 사용자가 직접 판별해야 하는 문제점이 있다. 사용자에게 신뢰할만한 대표적인 연구 결과물을 제공하기 위해서 본 논문에서는 관련된 분야에 상당한 지식과 경험을 가진 전문가를 검색하는 분야별 전문가 검색 기법을 제안한다. 분야별 전문가 검색을 위해 논문의 중요성 및 품질을 이용한다. 논문의 중요성은 각 분야별로 희소성이 있는 연구와 최신 토픽을 연구하는 논문으로 판별하고, 논문의 품질은 인용수, IF, 최신성, 저자관계를 이용해 판별한다.

[그림 1]은 전체 전문가 검색 시스템 구성도를 나타낸다. 데이터는 인용 수, IF, 논문정보를 수집하여 논문 정보의 논문 제목을 이용해 키워드를 추출한 후, 수집된 데이터를 저장한다. 저장된 논문정보와 IF, 인용수를 이용해 논문의 중요성 및 품질을 판별한다. 판별 후 전문가 지수를 도출하여 전문가 랭킹을 판별한다. 사용자가 검색 하여 질의를 입력 받으면 질의 분석을 수행한다. 온톨로지는 다음과 같은 과정을 통해 구성된다. 논문의 제목을 형태소 분석기를 이용해 분석하여 키워드로 분류한 후, 분석한 키워드를 count한 후 의미 있는 키워드만 추출한다. 각 키워드는 온톨로지를 통해 키워드가 내포하는 다양한 단어들을 하위 클래스를 이용하여 관리한다. 추출한 키워드를 기반으로 온톨로지를 구성하여 사용자가 검색하고자 하는 분야를 판별 할 수 있다. 최종적으로 제안하는 전문가 검색 시스템은 사용자의 질의에 따라 조회된 분야별 전문가 검색 결과를 사용자에게 제공한다. 본 논문에서 제안하는 기법으로 사용자의 질의에 따라 알맞은 분야별 전문가 검색을 수행할 수 있다.

[그림 2]는 제안하는 기법의 전문가 랭킹 처리 과정을 나타낸다. 논문의 중요성을 고려하기 위해 논문으로부터 연도 별 회소한 연구 분야를 판별하여 회소성 점수를 계산하고, 최근에 떠오르는 분야에 가중치를 부여하기 위해 최신 토픽을 판별한다. 논문의 품질을 고려하기 위해 인용지수계산, IF계산, 저자관계 판별, 최신성 판별을 수행한다. 논문의 품질 및 중요성을 이용해 전문가 지수를 판별하고 랭킹 분석을 수행한다.

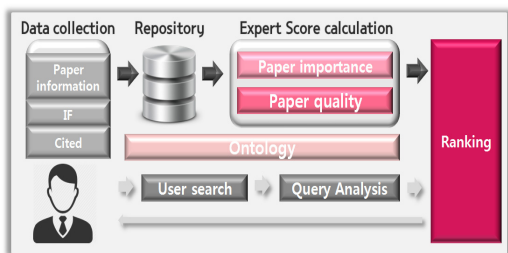


그림 1. 전문가 검색 시스템 구성도

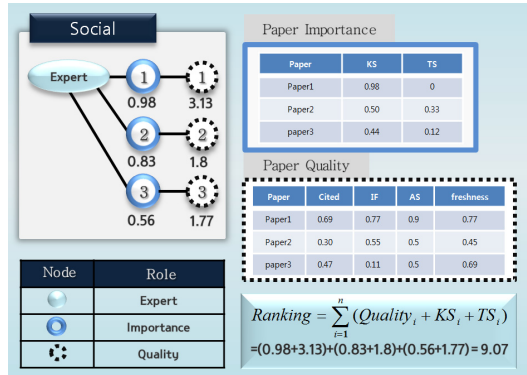


그림 2. 전문가 랭킹 처리 과정

2. 논문의 중요성 판별

제안하는 전문가 검색 기법은 각 분야별로 회소성이 있는 연구와 최신토픽을 연구하는 전문가가 중요성이 높다고 판단한다. 사용자들이 처음 연구를 시작할 때 어떤 연구가 최근에 많이 연구되고, 분야 내에서 중요한 연구인지 알기 어려운 문제점이 있다. 본 논문에서 중요성의 의미는 최근에 떠오르는 분야를 연구한 전문가와 회소한 분야의 연구를 한 전문가를 높게 평가하기 위함이다. 식 (1)은 키워드의 회소성을 계산하는 식이다. 첨자 i 는 논문을 나타내고 첨자 y_i 는 논문 i 의 연도를 나타낸다. 키워드의 회소성은 논문이 발행된 연도의 총 키워드 수와 논문이 출판된 연도의 총 논문 수를 이용하여 계산 한다.

$$KS_i = \log\left(\frac{Number\ of\ Paper_{y_i}}{Number\ of\ keyword_i}\right) \quad (1)$$

식(2)는 최신 토픽을 이용하여 키워드 별 가중치를 계산 하는 식이다. 최신에 떠오르는 분야에 가중치를 부여하기 위해 최신 토픽은 키워드별로 연도 당 키워드 수, 즉 토픽의 변화량을 계산하여 판단한다. 과거에 떠올랐던 토픽이 최신에 다시 떠올랐을 경우에는 상대적으로 낮은 점수를 부여한다. 현재연도는 출판된 논문이 많지 않기 때문에 현재연도 이전의 논문부터 계산한다. 토픽의 변화량은 연도 별로 출판된 논문의 수에 따라 달라진다. 논문이 적게 출판된 해에 상대적으로 낮은 수치의 토픽 변화량이 무시되지 않도록 연도 별로 총

논문수를 고려해 토픽 변화량을 보정해준다. 논문 수를 고려해 보정한 토픽 변화량을 최대 토픽 변화량으로 나누어서 0에서 1사이로 정규화 해준다.

$$T_i = KOfNumber_{y_{c-1}} - KOfNumber_{Y_{c-2}}$$

$$TS_i = \frac{T_i}{T_{max_i}} \quad (0 \leq TS_i \leq 1) \quad (2)$$

3. 논문의 품질 판별

본 논문에서는 논문의 품질을 판별하기 위해 인용 수, IF, 최신성, 저자관계를 이용한다. 본 논문에서는 사용자가 다수의 논문들 중 품질이 높은 논문에 관심을 갖고 인용하기 때문에 인용수를 논문의 품질에 반영하는 것이 의미 있다고 판단한다. 하지만 인용 지수만을 이용해 논문의 품질을 평가 할 경우, 최근 논문들이 상대적으로 낮은 품질로 평가될 수 있기 때문에, 최신성을 고려하고, 품질을 평가하기 위한 속성으로 논문이 게재된 학회의 평판을 고려하는 IF와 저자의 관계를 이용해 논문 품질을 계산한다. 식(3)은 논문의 품질을 위한 인용수를 계산하는 식이다. 인용 수는 Google 학술 검색을 이용해 수집하여 계산한다.

$$Cited_i = \log(Cited_i + 2) \quad (3)$$

식(4)는 논문이 게재된 학회의 평판을 고려하기 위한 점수이다. IF는 KCI 한국 학술지인용색인으로부터 영향력지수를 수집하여 이용하였다. KCI영향력지수는 특정기간동안 한 학술지에 수록된 하나의 논문이 다른 논문에 인용된 평균 횟수로 동일 분야 저널의 상대적 중요성을 비교 평가하는 방법이다. KCI로부터 수집한 IF 값들을 0에서 1사이로 정규화 하기 위해 IF최대값으로 나눠준다.

$$IF_i = \frac{IF_i}{IF_{max}} \quad (4)$$

식(5)는 논문의 최신성을 계산하는 식이다. 최신에 출판되는 논문들의 품질이 인용 수 등을 고려한 기존 기

법에서 낮게 평가되는 점을 고려하여 최신 출판 연도에 가중치를 부여한다. 이를 통해 최근에 논문을 많이 게재한 전문가에게 가중치 점수를 부여할 수 있다. 식 (5)의 Y_c 는 현재연도를 나타내고, Y_i 는 논문 I의 연도를 나타낸다.

$$freshness_i = \frac{1}{\log(2 + Y_c - Y_i)} \quad (5)$$

식(6)은 전문가의 전문성을 판단하기 위한 식이다. 기존의 전문가 검색 기법은 저자관계에 관계없이 논문을 많이 게재한 경우에 높은 점수를 부여받게 되는 문제점이 존재한다. 제안하는 기법에서는 논문의 주저자 또는 공저자의 관계를 파악해 상대적으로 영향력이 큰 저자에게 점수를 부여하는 저자점수를 계산한다. 저자수가 1인인 경우, 제 1저자에게 100% 즉 1점을 부여한다. 저자수가 2인 이상인 경우 2인 공동으로 작성 시 70%, 3인 공동으로 작성 시 50%, 4인 이상 작성 시 30%를 부여하고 저자가 2인 이상인 경우 1저자에게는 20%의 가중치를 부여한다.

$$AS_i = \sqrt{\sum_{i=1}^n Author_i} \quad (6)$$

$$Author_i = \left\{ \begin{array}{l} \text{제1저자} = 100\% \\ \text{공저자} \left\{ \begin{array}{l} 2인공동 = 70\% \\ 3인공동 = 50\% \\ 4인이상 = 30\% \\ \text{*제1저자} = +20\% \end{array} \right. \end{array} \right\}$$

식(7)은 인용수, IF, 최신성, 저자관계를 고려해 논문의 품질을 판별하는 식이다.

$$Quality_i = (Cited_i * IF_i) + freshness_i + AS_i(7)$$

4. 랭킹 분석

사용자의 검색 요청에 따라 적합한 전문가 랭킹 결과를 제공하기 위해, 전문가 검색 랭킹 계산은 다음 식 (8)과 같이 계산된다. 이때, KS는 키워드 점수이고 TS는 최신토픽 가중치이다. 각 논문마다 점수를 계산하기 때문에 분야별로 저자의 논문 편수도 함께 고려할 수 있

다. 이 식을 통해 분야별 전문가 랭킹 점수를 계산한다. 각 점수의 효율적인 비율을 위해 품질점수와 중요성점수를 각각 0에서 1사이로 정규화 해준다.

$$Ranking = \sum_{i=1}^n (Quality_i + KS_i + TS_i) \quad (8)$$

IV. 성능평가

제안하는 분야별 전문가 검색 기법과 기존의 기법의 성능 비교 평가를 통해 제안하는 기법의 우수성을 증명한다. [표 1]은 성능평가 파라미터를 나타낸다. 제안하는 기법에서 사용한 실험 환경은 다음과 같다. Intel core i5-3570k CPU 3.40GHz, 8GB 메모리를 가지는 시스템을 사용했고, 데이터베이스는 PostgreSQL을 활용하였다. 실험 데이터로는 논문데이터, 논문 인용 수, IF를 수집했다. 논문데이터는 DBpia에서 제공하는 Open API를 이용해 수집했다. 제안하는 전문가 검색 기법의 성능평가를 위해 수집된 총 논문의 수는 106095편이며, 저자의 수는 147326명으로 구성된 데이터셋을 사용하였다. 본 논문의 성능평가에서는 전체 데이터 셋 중 소셜 네트워크 분야의 논문 1791편과 저자 2363명을 이용했다. 논문에 대한 인용 수는 Google 학술 검색을 이용해 수집하였다. IF는 KCI한국 학술지인용색인 으로부터 영향력지수를 이용하여 수집했다.

표 1. 성능평가 파라미터

파라미터	값
프로세서	3.40GHz
메모리	8GB
운영체제	CentOS6.8
사용 데이터	DBpia-논문정보, 구글-인용 수, KCI-IF
논문 수 (소셜/전체)	1791/106095
저자 수 (소셜/전체)	2363/147326
검색대상 전문가수(소셜)	88

분야별 전문가를 검색하기 위해 각 연도 별 회소성과 최신 토픽 점수를 계산하여 논문의 중요성을 평가 하였고, 인용 수, IF, 최신성, 저자관계를 고려하여 논문의

품질을 평가한 후, 전문가 지수를 판별하여 랭킹을 분석하였다.

[표 2]는 전문가 검색 기법을 이용한 소셜 네트워크 분야의 전문가 랭킹 결과이다. 사용자가 검색한 질의가 소셜 네트워크 분야로 판별 되었다고 가정 한 후, 소셜 네트워크 분야 내에서 성능평가를 실시하였다. 논문의 품질 + 중요성을 고려한 기법과 품질만 고려한 기법과 기존기법과의 랭킹 결과의 차이를 확인하기 위해 성능평가를 실시하였다. 품질+중요성을 고려한 기법과 품질만 고려한 기법은 유사한 랭킹 결과가 나오지만 랭킹 결과의 다소 차이가 있는 것을 볼 수 있다. 기존기법과 품질+중요성을 고려한 기법을 비교한 결과, 기존기법에서 랭킹에 포함되지 않은 expert7와 같은 전문가들이 품질+중요성을 고려한 기법에서는 상위에 랭크되어 있는 것을 확인할 수 있다. 왜냐하면 expert7은 저자관계, 회소성, 최신토픽 등에서 높은 점수를 가지고 있지만, 기존 기법은 인용 수, IF, 최신성만을 고려하기 때문에 expert7와 같은 전문가들의 전문가 지수가 낮게 평가되기 때문이다.

논문의 품질만 고려한 기법과 품질+중요성을 함께 고려한 기법의 비교 결과, 품질+중요성을 고려한 기법에서 랭킹에 포함된 expert2와 같은 전문가들이 제안하는 기법에서는 20위 안에 포함되지 못한 것을 확인할 수 있다. expert2은 회소성과 최신토픽에서 높은 점수를 가지고 있지만, 품질만 고려했을 경우에는 낮게 평가되기 때문이다.

논문의 품질만 고려한 기법과 기존기법의 비교 결과, 제안하는 기법에서 상위에 랭크된 expert1과 같은 전문가가 기존기법에는 10위안에 포함되지 못한 것을 확인할 수 있다. 기존기법은 인용 수, IF, 최신성을 이용해 전문가를 판단하지만 논문의 품질만 고려한 기법은 추가적으로 저자관계를 고려한다. 이를 통해 expert1은 저자관계에서 높은 점수를 부여 받은 것을 확인할 수 있다.

실험 결과를 통해 본 논문에서 제안하는 기법의 저자관계와 회소성, 최신토픽을 고려함으로써 랭킹 결과가 다르게 나타나는 것을 확인하였다.

표 2. 전문가 랭킹 결과

순위	품질+중요성	품질	기존기법
1	expert1	expert1	expert2
2	expert7	expert7	expert3
3	expert4	expert4	expert13
4	expert5	expert5	expert6
5	expert9	expert10	expert8
6	expert10	expert9	expert11
7	expert13	expert14	expert9
8	expert14	expert19	expert15
9	expert19	expert24	expert18
10	expert6	expert13	expert20
11	expert23	expert27	expert22
12	expert24	expert6	expert25
13	expert18	expert23	expert23
14	expert27	expert30	expert28
15	expert30	expert18	expert1
16	expert22	expert22	expert4
17	expert2	expert36	expert31
18	expert8	expert37	expert33
19	expert36	expert32	expert12
20	expert37	expert38	expert16

분야별 전문가 검색 기법의 성능 비교를 위해 재현율(Recall), 정확률(Precision), F-score을 사용하였다. 재현율, 정확률, F-score를 평가하기 위한 데이터는 (품질+중요성), (기존기법), (인용 수) 3가지의 방법으로 전문가를 랭킹한 후, 3가지 랭킹결과에 모두 속한 전문가를 인용 수가 많은 기준으로 선별 한 데이터 집합을 사용했다.

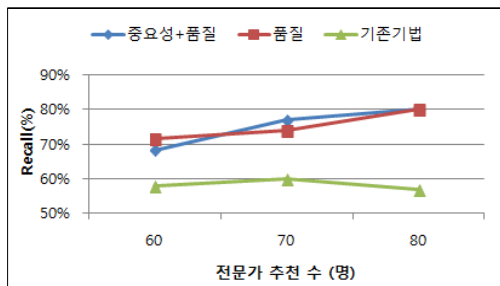


그림 3. 전문가 추천 수 변화에 따른 Recall(재현율)

[그림 3]은 전문가 추천 수 변화에 따른 재현율을 나타낸다. 전문가 추천 수는 60명에서 80명까지 변화시켜가며 측정하였다. 성능평가 결과, 중요성과 품질 모두를 고려한 기법이 기존 기법에 비해 우수한 성능을 보이는 것을 확인하였다. 60명의 전문가를 추천하였을 경우, 중

요성과 품질을 모두 고려한 기법이 기존기법에 비해 10% 높은 재현율을 보였다. 그 이유는 제안하는 기법이 저자관계, 희소성, 최신토픽 등에서 높은 점수를 가지고 있지만, 기존 기법은 인용 수, IF, 최신성만을 고려하기 때문에 전문가 지수가 낮게 평가되기 때문이다.

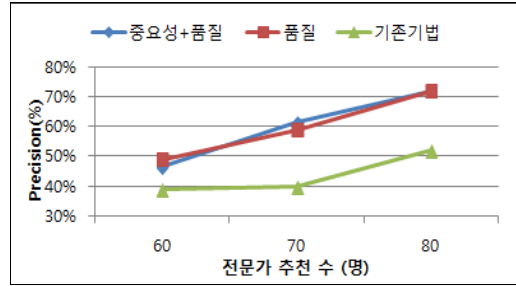


그림 4. 전문가 추천 수 변화에 따른 Precision(정확률)

[그림 4]는 전문가 추천 수 변화에 따른 정확률을 나타낸다. 전문가 추천 수는 60명에서 80명까지 변화시켜가며 측정하였다. 성능평가 결과, 중요성과 품질 모두를 고려한 기법이 기존 기법에 비해 우수한 성능을 보이는 것을 확인하였다. 60명의 전문가를 추천하였을 경우, 중요성과 품질을 모두 고려한 기법이 기존기법에 비해 8% 높은 정확률을 보였다. 이유는 제안하는 기법이 저자관계, 희소성, 최신토픽 등에서 높은 점수를 가지고 있지만, 기존 기법은 인용 수, IF, 최신성만을 고려하기 때문에 전문가 지수가 낮게 평가되기 때문이다.

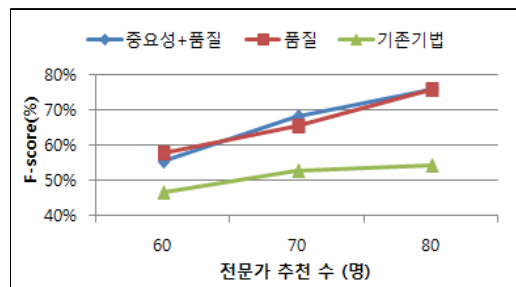


그림 5. 전문가 추천 수 변화에 따른 F-Score

[그림 5]는 제안하는 전문가 검색 기법과 제안하는 기법 중 논문의 품질만 이용한 기법, 기존의 전문가 검색기법의 F-score 비교 결과를 보여주고 있다. F-score

는 정확성을 측정하기 위한 통계적인 방법으로 정확률과 재현율에 가중치가 적용된 평균을 구함으로써 정확률 또는 재현율을 모두 고려하여 성능을 평가할 수 있다. F-score를 이용하여 기존기법과 제안하는 기법의 성능을 비교한 결과 전문가 추천 수 60명, 70명, 80명에 따라 11%, 13%, 21%의 F-Score값이 높아졌다. 기존의 전문가 검색 기법을 기반으로 전문가를 검색할 경우 인용 수, IF, 최신성만을 고려하기 때문에 전문가의 전문성을 보장하기 어렵다. 하지만 제안하는 분야별 전문가 검색 기법의 경우 기존기법에 추가적으로 논문의 중요성과 저자관계를 추가적으로 고려한다. 이로 인해 전체적으로 정확도, 재현율이 모두 향상되었음을 볼 수 있다.

V. 결론

본 논문에서는 논문의 중요성 및 품질을 고려하여 전문가를 검색하는 기법을 제안하였다. 논문의 중요성은 회소성이 있는 연구 분야와 최근에 많이 연구되는 분야에 가중치를 부여하여 중요성을 판단하였다. 논문의 품질은 해당 논문의 인용수와 출판·발행한 학회의 IF 연구의 최신성 및 저자관계를 통해 판단하였다. 성능평가 결과 제안하는 전문가 검색은 동일한 논문점수를 받더라도 중요성과 논문의 품질의 가중치에 의해 전문가가 기존 기법과 다르게 판단될 수 있음을 확인하였다. 제안하는 전문가 검색 기법을 통해 사용자가 새로운 과학 기술의 연구를 시작하고, 기존의 연구에 새로운 기술을 적용하고 개선, 발전시켜나가는데 활용할 수 있다. 향후 연구에서는 제안하는 기법의 수식을 보완하고, 사용자의 선호도에 대한 피드백 정보를 분석하여 각 속성에 대한 적합한 가중치를 검색 결과에 반영할 수 있는 연구를 수행할 계획이다.

참고 문헌

- [1] <https://scholar.google.co.kr/>
 [2] <http://www.dbpia.co.kr/>

- [3] <https://www.kci.go.kr>
 [4] A. Omidvar, M. Garakani, and H. R. Safarpour, "Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis," *Information Technology and Management*, Vol.15, No.1, pp.51-63, 2014.
 [5] T. Vu and A. Baid, "Ask, Don't Search: A Social Help Engine for Online Social Network Mobile Users," *Proc. IEEE Samoff Symposium*, pp.1-5, 2012.
 [6] S. D. Gollapalli, P. Mitra, and C. L. Giles, "Ranking Experts using Author-Document-Topic graphs," *Proc. ACM/ IEEE-CS joint conference on Digital libraries*, pp.87-96, 2014.
 [7] 송창우, 김종훈, 정경용, 류중경, 이정현, "시맨틱 웹에서 개인화 프로파일을 이용한 콘텐츠 추천 검색 시스템," *한국콘텐츠학회논문지*, 제8권, 제1호, pp.318-327, 2008.
 [8] 한희준, 예용희, 류범중, "학술정보서비스에서 인명검색 고도화 방법," *한국콘텐츠학회논문지*, 제10권, 제2호, pp.490-498, 2010.
 [9] S. Wang, S. Xie, X. Zhang, Z. Li, S. Y. Philip, and X. Shu, "Future Influence Ranking of Scientific Literature," *Proc. SIAM International Conference on Data Mining*, pp.749-757, 2014.
 [10] S. K. Rani, K.Raju, and V. V. Kumari, "Expert Finding System using Latent Effort Ranking in Academic Social Networks," *International Journal of Information Technology and Computer Science*, Vol.7, No.2, pp.21-27, 2015.
 [11] A. Bozzon, M. Brambilla, S. Ceri, M. Silvestri, and G. Vesci, "Choosing the Right Crowd: Expert Finding in Social Networks," *Proc. International Conference on Extending Database Technology*, pp.637-648, 2013.
 [12] 전인배, 임종태, 복경수, 유재수, "소셜 네트워크 환경의 신뢰성 향상을 위한 전문가 검색 기법," *정보과학회논문지: 컴퓨팅의 실제 및 테더*, 제20권, 제5호, pp.311-315, 2014.

- [13] X. Li and T. Watanabe, "Automatic Paper-to-reviewer Assignment, Based on the Matching Degree of the Reviewers," *Procedia Computer Science*, No.22, pp.633-642, 2013.
- [14] D. Liu, W. Xu, W. Du, and F. Wang, "How to Choose Appropriate Experts for Peer Review: An Intelligent Recommendation Method in a Big Data Context," *Data Science Journal*, Vol.14, p.16, 2015.
- [15] A. Kale, R. Kharat, S. Bodkhe, and P. Apte "Automated Fair Paper Reviewer Assignment for Conference Management System," *Proc. International Conference on Computing Communication Control and Automation*, pp.408-411, 2015.
- [16] 이서희, 박윤정, 한진수, 최도진, 임종태, 복경수, 유재수, "논문의 중요성 및 품질을 이용한 전문가 검색 기법," 한국 콘텐츠학회 2016 춘계종합학술대회, pp.17-18, 2016.

저 자 소 개

이 서 희(Seo-Hee Lee) 준회원



- 2015년 2월 : 청주대학교 통계학과(이학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터협동과정 석사과정

<관심분야> : 소셜네트워크, 빅데이터, 전문가 검색

박 윤 정(Yun-Jeong Park) 준회원



- 2015년 2월 : 청주대학교 통계학과(이학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터협동과정 석사과정

<관심분야> : 소셜네트워크, 빅데이터

한 진 수(Jin-Su Han) 준회원



- 2016년 2월 : 충북대학교 정보통신공학과(공학사)
- 2016년 3월 ~ 현재 : 충북대학교 정보통신공학과 석사과정

<관심분야> : 그래프 분산처리, 빅데이터

최 도 진(Do-Jin Choi) 정회원



- 2014년 2월 : 한국교통대학교 컴퓨터공학과(공학사)
- 2016년 2월 : 한국교통대학교 컴퓨터공학과(공학석사)
- 2016년 3월 ~ 현재 : 충북대학교 정보통신공학과 박사과정

<관심분야> : 전문가 검색, 연속 질의 처리, 그래프 스트림, 데이터베이스, 빅데이터

임 종 태(Jong-Tae Lim) 정회원



- 2009년 2월 : 충북대학교 정보통신공학과(공학사)
- 2011년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2015년 8월 : 충북대학교 정보통신공학과(공학박사)

▪ 2015년 9월 ~ 현재 : 충북대학교 정보통신공학과 박사후연구원(Post.doc)

<관심분야> : 시공간 데이터베이스 시스템, 이동 객체 질의 처리, 위치기반 서비스, P2P 네트워크, 빅데이터

북 경 수(Kyoung-Soo Bok)

종신회원



- 1998년 2월 : 충북대학교 수학과 (이학사)
- 2000년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 8월 : 충북대학교 정보통신공학과(공학박사)

- 2005년 3월 ~ 2008년 2월 : 한국과학기술원 정보전자연구소 Postdoc
- 2008년 3월 ~ 2011년 2월 : 가인정보기술 연구소 연구원
- 2011년 3월 ~ 현재 : 충북대학교 전자정보대학 정보통신공학부 초빙교수

<관심분야> : 데이터베이스 시스템, 이동 객체 데이터베이스, 이동 P2P 네트워크, 소셜 네트워크 서비스, 빅데이터

유 재 수(Jaesoo Yoo)

종신회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 한국과학기술원 전산학과(공학석사)
- 1995년 2월 : 한국과학기술원 전산학과(공학박사)

- 1995년 2월 ~ 1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월 ~ 현재 : 충북대학교 전자정보대학 정교수

<관심분야> : 데이터베이스 시스템, 멀티미디어 데이터베이스, 센서 네트워크, 바이오 인포메틱스, 빅데이터