

## 논문의 중요성 및 품질을 이용한 전문가 검색 기법

### Expert Search Method Using Importance and Quality of Papers

이 서 희, 박 윤 정, 한 진 수\*, 최 도 진\*, 임 종 태\*,  
 복 경 수\*, 유 재 수\*<sup>†</sup>  
 충북대학교 빅데이터학과,  
 충북대학교 정보통신공학부\*

Seohee Lee, Yunjeong Park, Jinsu Han\*, Dojin Choi\*,  
 Jongtae Lim\*, Kyoungsoo Bok\*, Jaesoo Yoo\*<sup>†</sup>  
 Department of big data, Chungbuk National University,  
 School of Information and Communication Engineering,  
 Chungbuk National University\*

#### 요약

본 논문에서는 논문의 중요성 및 품질을 고려한 분야별 전문가 검색 기법을 제안한다. 논문의 중요성은 연구 분야의 희소성 및 최신 토픽을 고려하여 계산하며 논문의 품질은 연구 결과물의 인용수, IF, 최신성, 저자 관계를 고려하여 연구 결과물의 전문성을 계산한다.

### I. 서론

새로운 지식과 기술의 발전과 함께 많은 양의 정보가 끊임없이 발생하고 있다. 많은 양의 정보 중 사용자들은 자신이 원하는 질의에 대한 신뢰할만한 정보를 얻기 위해 관련된 분야에 상당한 지식과 경험을 가진 전문가를 검색하기를 원한다. 전문가 검색을 통해 사용자는 원하는 질의에 대해 알맞은 정보를 제공해 줄 전문가를 선별할 수 있다.

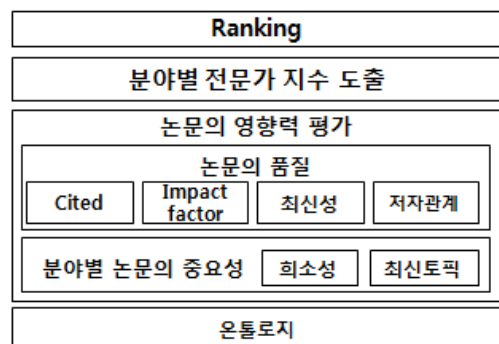
기존 전문가 검색 기법은 특정한 키워드로 질의가 주어지면 질의에 대해 가장 가치 있는 정보를 제공할 수 있는 전문가를 검색한다. 하지만 기존 전문가 검색 기법은 분야를 고려하지 않기 때문에 해당 분야에 적합한 전문가를 검색하기 어렵다. 따라서 각 분야별 전문가를 판별하기 위해 논문의 중요성과 논문의 품질을 고려한 기법이 필요하다.

제안하는 기법에서는 논문의 중요성 및 품질을 이용한 전문가 검색 기법을 제안한다. 제안하는 기법은 논문의 중요성을 계산할 때 논문의 희소성과 최근 이슈가 되는 토픽을 고려한다. 사용자가 원하는 분야의 전문가를 검색하기 위해 유사한 키워드를 하나의 키워드로 판단하는 온톨로지를 구성한다. 또한 연구 결과물의 중요성 및 품질을 기반으로 각 분야별 전문가 지수를 계산하고 사용자에게 적합한 전문가를 검색한다.

### II. 제안하는 전문가 검색 기법

#### 1. 제안 기법 구조

본 논문에서는 사용자의 검색 요청에 따라 분야별 전문가검색 기법을 제안한다. 그림 1은 제안하는 기법의 시스템구조를 나타낸다. 유사한 키워드를 하나의 키워드로 판단하는 온톨로지를 구성하고 논문의 점수는 분야별 논문의 중요성과 논문의 품질을 이용해 논문의 영향력을 평가한다. 평가된 논문의 영향력을 이용해 분야별 전문가 지수를 도출한 후, 랭킹 결과를 얻는다.



▶▶ 그림 1. 분야별 전문가 검색을 위한 시스템 구조

#### 2. 분야별 논문의 중요성

논문의 중요성은 각 분야별로 희소성이 있는 연구와 최신토픽을 연구하는 논문이 중요성이 높다고 판단한다. 식 (1)은 키워드의 희소성을 계산하는 식이다. 키워드의 희소성은 논문이 발행(발행/출판 통일)된 년도의 총 키워

<sup>†</sup> 교신저자 : yjs@chungbuk.ac.kr

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학 ICT연구센터육성 지원사업의 연구결과 (IIP-2016-H8501-16-1013)이며, 2015년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 에너지인력양성사업으로 지원받아 수행한 인력양성 성과 (No. 20144030200450)입니다.

드 수와 논문이 출판된 년도의 총 논문 수를 이용해 계산한다.

$$KS_i = \log\left(\frac{\text{Number Of Paper}_{Y_i}}{\text{Number Of Keyword}_{Y_i}}\right) \quad (1)$$

식(2)는 최신 토픽을 이용하여 키워드 별 가중치를 계산하는 식이다. 최신에 떠오르는 분야에 가중치를 부여하기 위해 최신 토픽은 키워드별로 년도 당 키워드 수 즉 토픽의 변화량을 계산하여 판단한다. 현재년도는 출판된 논문이 많지 않기 때문에 현재년도 이전의 논문부터 계산한다. 토픽의 변화량은 년도 별로 출판된 논문의 수에 따라 달라진다. 논문이 적게 출판된 해에 상대적으로 낮은 토픽 변화량이 무시되지 않도록 년도 별로 총 논문수를 고려해 토픽 변화량을 보정해준다. 논문 수를 고려해 보정한 토픽 변화량을 최대 토픽 변화량으로 나누어서 0에서 1사이로 정규화 해준다.

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

### 3. 논문의 품질

제안하는 기법에서는 논문의 품질을 파악하기 위하여 인용수와 출판된 학회의 IF, 연구의 최신성, 저자관계를 고려했다. 본 논문에서는 사용자들이 다수의 논문들 중 품질이 높은 논문에 관심을 갖고 인용하기 때문에 인용수를 논문의 품질에 반영하는 것이 의미 있다고 판단한다. 하지만 인용지수만을 이용해 논문의 품질을 평가할 경우, 최근 논문들이 상대적으로 낮은 품질로 평가될 수 있기 때문에, 최신성을 고려하고, 품질을 평가하기 위한 속성으로 논문이 게재된 학회의 평판을 고려하는 IF를 이용해 논문 품질을 계산한다. 식(3)은 논문의 인용지수와 IF를 이용해 논문의 품질을 계산하는 식이다.[1]

$$Quality_i = (Cited_i + 1) \times IF / IF_{max} \quad (3)$$

식 (4)는 전문가의 전문성을 판단하기 위한 식이다. 주저자 또는 공저자의 관계를 파악해 저자점수를 계산한다.

$$AS_i = \begin{cases} \frac{2}{n+1}, & \text{주저자인 경우} \\ \frac{1}{n+1}, & \text{공저자인 경우} \end{cases} \quad (4)$$

식 (5)은 최종적으로 식(3)과 식(4)의 결과를 이용해 논문에 대한 품질을 계산하는 식이다. 최신 출판되는 논문들의 품질이 낮게 평가되는 점을 고려해 최신 출판년도에 가중치를 부여한다. 논문에 대한 품질을 계산 후에 각 분야별로 전문가 점수를 계산한다.

$$Paper_i = \frac{1}{\log(2 + Y_c - Y_i)} \times Quality_i \times AS_i \quad (5)$$

### 4. 전문가 검색

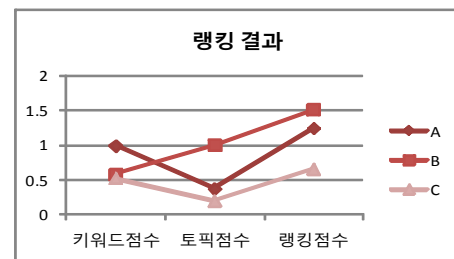
사용자의 검색 요청에 따라 적합한 전문가 랭킹 결과를 제공하기 위해, 전문가 검색 랭킹 계산은 다음 식(6)과 같이 계산된다. 이때, KS<sub>i</sub>는 키워드 점수이고 TS<sub>i</sub>는 최신토픽 가중치이다. 각 논문마다 점수를 계산하기 때문에 분야별로 저자의 논문 편수도 함께 고려할 수 있다. 이 식을 통해 분야별 전문가 랭킹 점수를 계산한다.

$$\text{Ranking} = \sum_{i=1}^n Paper_i \times KS_i + TS_i \quad (6)$$

### III. 성능평가

본 논문에서는 DBpia에서 제공하는 Open API를 이용해 논문을 수집하고 논문에 대한 인용 수는 Google 학술 검색을 이용했다. 분야별 전문가를 검색해 주기 위해 회소성 있는 분야와 최근 이슈가 되는 분야에 가중치를 부여한다.

그림 2는 논문 지수가 같은 저자들을 대상으로 제안하는 기법을 이용하여 랭킹 결과이다. A저자는 회소성이 있는 논문을 작성하여 키워드점수가 높지만 토픽점수는 낮게 측정되었고, B저자는 최근 이슈가 되는 분야의 논문을 작성하여 높은 토픽점수로 가장 높은 랭킹 점수를 보인다. C저자는 키워드점수와 토픽점수가 높지 않아서 가장 낮은 랭킹 결과를 보인다. 실험을 통해 제안하는 랭킹 방법이 분야별로 논문의 중요성과 논문 품질을 고려해 적절한 랭킹을 부여하는 것을 확인 할 수 있었다.



▶▶ 그림 2. 랭킹 결과

### IV. 결론

본 논문에서는 논문의 중요성과 품질을 고려하여 전문가를 검색하는 기법을 제안하였다. 논문의 중요성은 회소성이 있는 연구분야와 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 논문 점수를 받더라도 중요성과 논문 품질의 가중치에 의해 전문가가 기존 기법과 다르게 판단될 수 있음을 확인하였다.

### ■ 참고 문헌 ■

- [1] Xinlian Li and Toyohide Watanabe, "Automatic Paper-to-reviewer Assignment, Based on the Matching Degree of the Reviewers", *Procedia Computer Science* 22 pp.633 - 642, 2013