

## 학술 빅데이터 분석을 통한 전문가 검색 시스템

### Expert Search System Using Analysis of Academic Big-Data

최도진\*, 김민수\*, 김대윤\*\*, 이서희\*\*, 한진수\*,  
서인덕\*\*, 임종태\*, 북경수\*, 이병엽\*\*\*, 유재수\*†

\*충북대학교 정보통신공학과,

\*\*충북대학교 빅데이터학과,

\*\*\*배재대학교 사이버보안학과

Dojin Choi\*, Minsoo Kim\*, Daeyun Kim\*\*, Sehee Lee\*\*,  
Jinsu Han\*, Indeok Seo\*\*, Jongtae Lim\*,

Kyoungsoo Bok\*, Byoungyup Lee\*\*\*, Jaesoo Yoo\*†

\*School of Information and Communication  
Engineering, Chungbuk National University

\*\*Department of Big Data, Chungbuk National  
University

#### 요약

많은 연구 종사자들은 기존 연구를 참고하거나 특정 분야 연구자를 검색하기 위해 학술 검색 사이트를 이용한다. 그러나 학술 검색 사이트는 주제 또는 사용자별 발간 연구 결과물은 검색할 수 있지만 연구 분야별 전문가를 제공하지 않는다. 본 논문에서는 공개된 학술 빅데이터 분석을 통해 분야별 전문가를 검색하는 시스템을 제안한다. 제안하는 기법에서는 특정 분야에 대한 전문가 검색뿐만 아니라 연구자와 관련된 인적 네트워크를 제공한다. 또한, 웹 기반 인터페이스를 제공을 통해 사용자 접근성을 향상시킨다.

## I. 서론

최근 다양한 학문 분야에서 논문, 특허 등을 통해 학술 연구 결과물들이 온라인을 통해 제공되고 있다[1]. 연구자들은 기존 연구들을 참고하기 위해 Google Scholar, DBLP, DBPia 등과 같은 학술 검색 시스템을 이용한다. 많은 연구자들은 새로운 연구 수 또는 공동 연구를 수행하기 위해 특정 연구 분야에 대한 전문가 판별을 요구한다. 그러나 학술 검색 시스템은 사용자 또는 특정 키워드에 해당되는 학술 결과물만을 제공하기 때문에 특정 분야 전문가를 제공하지 못한다.

[2]에서는 전문가 검색 및 논문 검색 기능을 제공한다. 그러나 이 기능은 연구 분야별 전문가를 제공하지 않는다. 본 논문에서는 공개된 학술 빅데이터 분석을 통해 분야별 전문가 검색 시스템을 제안한다. 제안하는 시스템은 특정 분야에 대한 전문가 검색뿐만 아니라 연구자와 관련된 인적 네트워크를 제공한다. 또한 제안하는 기법은 Open API를 통해 논문을 수집하고 온톨로지를 이용하여 논문의 분야를 추출한다. 최종적으로 추출된 분야를 이용하여 분야의 전문가를 추천한다. 제안하는 시스템은 웹 인터페이스를 통해 편리하게 사용이 가능하도록 개발하였다.

II. 제안하는 전문가 검색 시스템

## II. 제안하는 전문가 검색 시스템

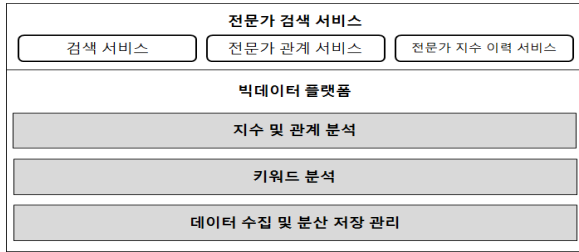
### 1. 시스템 구조

제안하는 시스템은 학술 빅데이터를 수집하고 빅데이터 저장 시스템에 분산 저장한다. 수집된 학술 빅데이터를 통해 논문의 제목을 기반으로 분야를 추출한다. 논문의 제목을 형태소 분석기를 통해 논문 제목의 단어들을 추출한다. 추출된 단어들은 온톨로지를 이용하여 논문의 분야를 구분할 수 있다. 최종적으로 추출된 분야를 기반으로 전문가 지수를 빅데이터 처리 시스템을 통해 계산한다.

그림 1은 제안하는 전문가 검색 시스템의 구조를 나타낸 것이다. 학술 빅데이터를 온라인을 통해 수집하고 분산 저장한다. 키워드 분석은 수집된 논문에 대한 분야를 추출하는 단계이다. 하나의 논문은 여러 개의 분야를 가질 수 있다. 추출된 분야를 이용하여 전문가 지수 단계에서 분야별 전문가를 판별한다. 추가적으로 논문의 저장 정보를 이용하여 전문가 간의 관계 분석을 수행한다. 분석한 관계 정보를 이용하여 사용자는 연구자들의 관계를 쉽게 파악할 수 있고, 이를 검색에 활용할 수 있다. 최종적으로 사용자들은 웹 인터페이스를 통해 키워드나 이름으로 전문가를 검색할 수 있다.

† 교신 저자 : yjs@chungbuk.ac.kr

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업(ITP-2016-H8501-16-1013)과 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 한국연구재단에서 부여한 과제번호 : 2014R1A1A2055778).



▶▶ 그림 1. 전문가 검색 서비스 아키텍처

2. 세부 기능 및 구현

학술 데이터를 수집하기 위해서 DBpia의 OpenAPI를 이용하여 논문을 수집하였다. 논문에 대한 인용 수 수집은 Google scholar의 자료를 참고한다. 수집된 논문은 영속 저장을 위해 Hadoop 기반의 NoSQL 엔진인 HBase에 저장한다. 수집된 논문들은 제목을 기반으로 온톨로지를 통해 분야별로 나누고 분야별 전문가 추천 알고리즘을 수행한다. 제안하는 시스템에서는 분야별 전문가 지수를 계산하기 위해서 분산 인-메모리 플랫폼인 Spark을 이용하여 계산하였다. Spark에서 계산된 전문가 지수의 결과는 HBase에 저장한다.

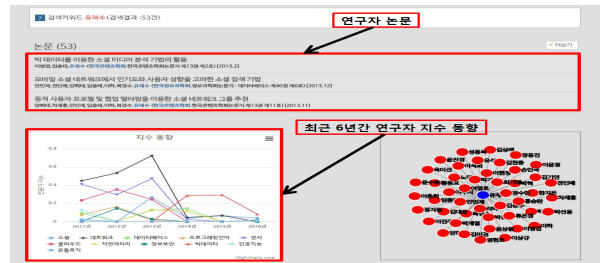
그림 2는 학술 빅데이터의 저장 및 처리 구성도를 보여준다. 학술 빅데이터를 저장하기 위해 빅데이터 저장 시스템인 HBase를 이용하여 분산 저장을 한다. 저장된 논문 정보와 저자정보를 빅데이터 처리 시스템인 Spark을 통해 키워드를 고려하여 계산한다. 처리된 전문가 지수와 전문가 관계로 다시 HBase에 저장된다.



▶▶ 그림 2. 저장 및 처리 구성

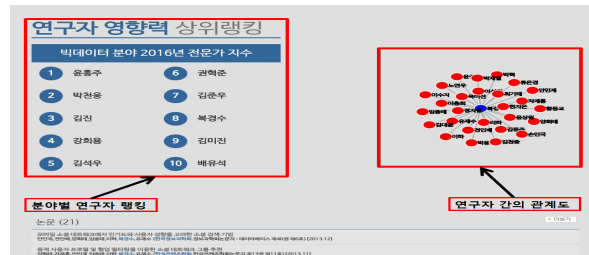
전문가 검색 서비스에서는 연구자 이름 기반 검색과 키워드 기반 검색 기능을 제공한다. 이름 기반 검색 기능은 연구자의 논문, 그 연구자의 최근 6년간 연구자 지수 동향을 그래프로 보여주고 그 연구자와 다른 연구자들의 관계성을 시각화 한다. 키워드 기반 검색 기능은 키워드 별 연구자 영향력 랭킹과 연구자 간의 관계성을 시각화하고 영향력 랭킹은 연도별로 제공한다.

그림 3은 검색 결과로 연구자가 참여한 논문의 목록과 전문가의 지수 동향을 보여준다. 전문가의 지수 동향은 연구자가 연도별로 연구하였던 분야들을 그래프 형태로 시각화한 기능이다. 우측은 연구자의 관계성을 시각화한 것이다.



▶▶ 그림 3. 연구자 검색 서비스

그림 4는 키워드 기반 검색 기능을 보여준다. 좌측은 검색한 분야의 상위 전문가들을 나타낸다. 우측 그림은 분야별 상위 전문가에서 사용자가 선택한 전문가의 관계를 그래프로 시각화 하였다.



▶▶ 그림 4. 분야 검색 서비스

III. 결론

본 논문에서는 논문의 분야를 추출하여 분야별 전문가 추천이 가능한 시스템을 설계하고 구현하였다. 제안하는 시스템은 전문가가 출판한 논문의 제목에서 단어를 추출하여 분야를 판별하였고, 판별된 분야별로 전문가 지수를 분석할 수 있었다. 제안하는 시스템은 논문의 저자 정보를 이용하여 연구자 간의 관계 정보를 시각화하였다. 또한 사용자들이 웹 인터페이스를 통해 편리하게 전문가 검색 서비스를 사용할 수 있게 하였다.

■ 참고 문헌 ■

[1] 조석민, 정은희, 강동환, 조동욱, 전세림, 김소영, 임소현, 최윤정, 차은지, 홍성표, 황덕수, 최태진, “2015 교육부 학술연구지원사업 성과분석보고서”, 한국연구재단, pp.1-11, 2015  
 [2] http://www.dbpia.co.kr/  
 [3] 김성숙, 백창원, 박기진, “스파크 (Spark) 인메모리 처리를 이용한 스마트 다바이스 빅데이터 실시간 분석.” 한국정보과학회 학술발표논문집, pp.317-319, 2015