https://doi.org/10.5392/JKCA.2017.17.03.100

빅데이터 처리 플랫폼에서 학술 데이터를 사용한 전문가 검색 시스템 설계 및 구현

Design and Implementation of an Expert Search System Using Academic Data in Big Data Processing Platforms

최도진*, 김민수*, 김대윤*, 이서희**, 한진수*, 서인덕**, 임종태*, 복경수*, 유재수*
충북대학교 정보통신공학과*, 충북대학교 빅데이터협동과정**

Dojin Choi(mycdj91@cbnu.ac.kr)*, Minsoo Kim(mskim88@cbnu.ac.kr)*,
Daeyun Kim(kdy0573@cbnu.ac.kr)*, Seohee Lee(shl@cbnu.ac.kr)**,
Jinsu Han(jinsu@cbnu.ac.kr)*, Indeok Seo(duckdeock@cbnu.ac.kr)**,
Jongtae Lim(jtlim@cbnu.ac.kr)*, Kyoungsoo Bok(ksbok@cbnu.ac.kr)*,
Jaesoo Yoo(yjs@cbnu.ac.kr)*

요약

대부분의 연구자들은 새로운 분야의 연구를 수행하기 위해 전문가에게 자문을 받거나 전문가의 논문들을 기반으로 연구 방향을 설정한다. 기존의 학술 검색 서비스에서는 분야별 논문 정보는 제공하지만 각 분야의 전문가를 제공해주지 않기 때문에 사용자들이 검색된 논문을 기반으로 전문가를 직접 판단해야 한다. 본 논문에서는 학회에 발간된 논문 정보를 기반으로 빅 데이터 처리를 이용한 학문 분야별 전문가 검색 시스템을 설계하고 구현한다. 제안하는 시스템은 대량의 논문을 저장하고 관리하기 위해 빅 데이터 분산 저장 기술을 활용하였다. 또한 빅 데이터 분산 처리기술을 활용하여 전문가를 판별하고 전문가와 연관 되는 정보를 분석한다. 분산처리 된 결과는 사용자가 전문가 검색 요청 시 웹페이지를 통해 보여준다. 사용자는 제안하는 시스템을 통해 해당 연구 분야의 전문가를 추천받음으로써 연구를 수행함에 있어 많은 도움을 받을 수 있다.

■ 중심어 : | 학술 데이터 | 전문가 | 빅 데이터 | 검색 | 분산 처리 | 데이터베이스 |

Abstract

Most of the researchers establish research directions to conduct the study of new fields by getting advice from experts or through the papers of experts. The existing academic data search services provide paper information by field but do not provide experts by field. Therefore, users should decide experts by field using the searched papers by themselves. In this paper, we design and implement an expert search system by discipline through big data processing based on papers that have been published in the academic societies. The proposed system utilizes distributed big data storage systems to store and manage large papers. We also discriminate experts and analyze data related to the experts by using distributed big data processing technologies. The processed results are provided through web pages when a user searches for experts. The user can get a lot of helps for the research of a particular field since the proposed system recommends the experts of the corresponding research field.

■ keyword : | Academic Data | Expert | Big Data | Search | Distributed Processing | Database |

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

접수일자 : 2016년 12월 02일
수정일자 : 2016년 12월 23일
심사완료일 : 2016년 12월 23일
교신저자 : 유재수, e-mail : yjs@cbnu.ac.kr

I. 서 론

학술 분야에서 논문은 연구자의 전문성을 판단하기 위한 중요한 지표중 하나이다. 한 분야에 다양한 논문을 출판하고 많은 인용수를 받는다면 해당 논문의 저자는 그 분야의 전문가라고 볼 수 있다. 이러한 전문가 추론으로 학회에서는 전문가가 연구하는 분야의 논문들에 대해 심사를 요청 할 수 있고, 사용자는 전문가의 이전 논문들과 현재 논문 정보를 활용하여 전문가 연구 분야의 이전의 연구 흐름과 현재 연구동향을 쉽게 파악할 수 있다.

연구자들은 새로운 분야의 연구를 수행하기 위해서 기존에 발표되었던 많은 논문들을 참고한다. 연구자들은 참고하고자하는 논문을 찾기 위해서 다양한 학술검색 사이트를 이용한다[1-5]. 기존 학술검색 사이트는 논문과 관련된 다양한 정보를 제공하고 있고, 연구자들은 이를 활용해서 참고하고자 하는 논문들을 찾거나 관련 논문을 추천 받는다. 학술검색 사이트에서 활용 할 수 있는 정보는 논문의 저자 정보, 출판지, 출판 연도, 인용 정보 등이 제공되고 있다[6][7].

대부분의 학술 검색 사이트는 인용이 많이 된 논문이거나 최근에 나온 논문 혹은 검색어에 가장 부합하는 논문들을 추천한다[1-4]. 일부 사이트는 학술 검색 기능에 소셜 네트워크 기능을 추가함으로써 다양한 분야의 연구자들한테 질문을 하거나 무료로 제공되지 않는 논문들을 요청함으로써 연구 활동에 도움을 주는 기능을 제공하고 있다[5][8]. 또한, 최근 연구에서는 소셜 환경에서 게시글과 최신활동 내역을 분석하여 특정 연구 분야의 전문가를 제공하거나 사용자의 평판을 기반으로 논문의 품질을 측정하여 이를 검색 기능에 추가적으로 제공하는 연구도 진행되고 있다[9][10].

학술 검색에서 분야의 전문가를 찾을 수 있다면 연구자는 연구를 수행함에 있어 많은 도움이 될 수 있다. 만약 연구자가 현재 연구하고자 하는 분야의 연구동향을 파악하고 싶을 때 전문가가 누군지 알 수 있다면 해당 분야의 전문가의 논문들만 보더라도 연구의 흐름과 최신 연구동향까지 쉽게 파악할 수 있게 된다. 물론 다양한 논문을 참고할 수 있으면 좋지만 그 분야의 모든 논문을 읽기란 어렵기 때문에 연구동향을 파악하는 것은 매우 힘든 일이다. 기존의 학술검색 사이트에서는 연구자의 생산성이나 소셜 활동 지수만을 제공해주기 때문에 어떤 연구자가 전문가인지는 파악하기 매우 힘들다. 따라서 학술검색에서 전문가를 찾는 것은 연구자들에게 있어서는 유익한 일이고 연구를 수행함에 있어서는 필수적이고 중요한 과제이다.

대부분의 연구자들은 자신만의 연구 분야 혹은 관심 분야를 가지고 있다. 이러한 정보를 전문가 검색에 활용한다면 분야별 전문가를 비교적 쉽게 구별할 수 있을 것이다. 그러나 연구자들마다의 연구 분야를 수집하는 것은 불가능하다. 회원 가입 양식의 학술사이트라면 가입과 동시에 사용자가 직접 기술하는 것으로 수집할 수 있지만 모든 연구자가 가입을 해야 하고 성실하게 답변을 해야만 한다. 혹은 다른 방법으로 논문을 수집하면서 명시된 분야 정보를 활용 할 수 도 있지만 저널마다 양식이 다르고, 저널에 따라 논문이 공개되지 않기도 한다. 또한, 기존 학술사이트에서 제공하는 정보로 논문의 요약정보가 제공되지 않아 논문의 분야를 판단하기 어렵기 때문에 논문의 제목 정보만을 활용하여 분야를 추출하는 기법을 제안한다.

기존 학술검색 사이트들은 과학자의 영향력을 표현하는 h-index (Hirsch index) 지표나 소셜 활동 기반의 수치를 일부 제공하고 있다. 그러나 소셜 활동의 수치는 저자의 전문성 판별에 적합하지 않은 수치이고 h-index는 분야의 전문가를 판단하는 수치로는 사용하기 어렵다. 제안 하는 기법은 분야별 전문가를 판단하기 위해서 연구자의 전문성을 판단하는 수치가 필요하다. 연구자의 전문성을 판단하기 위해 논문의 인용 지수, 저자 지수, 논문의 최신성, 분야의 희소성을 사용한다. 계산된 논문의 전문성을 총합 하면 연구자의 전문성 지수가 도출되고, 최종적으로 연구자의 분야별 전문성 수치는 분야별 전문가를 판단하는데 사용한다.

국내에 연구재단 등재지는 약 2천개 이상의 학회가 존재하고 매달 10건 이상의 논문이 학회에서 발표된다고 가정하면 매년 20만 건 이상의 논문이 발표된다[2]. 국내뿐만 아니라 국외 저널까지 감안한다면 수많은 논문이 매해 출판되는 셈이다. 그리고 매년 발표되는 논

문에 과거에 발표된 논문까지 본다면 엄청난 양의 논문이 존재하게 된다. 이러한 대용량의 논문 데이터에서 전문가를 판단하기 위해서는 효율적인 처리 시스템을 사용해야만 한다. 기존의 학술 검색사이트에서는 논문의 메타데이터만을 가지고 검색기능을 제공하는 것이 목적이므로 논문 정보를 영속적으로 저장하고 빠른 탐색을 위한 시스템을 설계하였다. 그러나 만약 대용량으로 수집된 논문 정보에서 논문의 분야를 판단해야 하고 추가적으로 분야별 전문가를 판단하는 기법까지 요구가 된다면 단순하게 저장하고 관리하는 것만으로는 기능적 한계에 봉착할 수밖에 없다. 본 논문에서는 이러한 문제점을 해결하기 위하여 빅 데이터 처리 시스템을 사용하여 이를 해결하였다. 제안하는 시스템은 단순 검색 데이터는 응답성 높은 저장소에 저장하고 전문가를 판단하는 기법을 위해서는 빅 데이터 시스템을 사용하여 처리와 검색 모두 다 만족할 수 있는 효율적인 시스템을 설계하였다.

본 논문에서는 다양한 학술 검색 사이트에서 제공되는 논문 정보를 활용하여 분야별 전문가를 검색하는 시스템을 설계하고 구현한다. 제안하는 시스템은 논문 제목 정보를 활용하여 논문의 분야를 판단한다. 또한, 논문의 인용 지수와 더불어 최신 논문의 인용지수가 낮은 것을 보완하는 최신성과 희소한 분야에 연구를 한 논문에 가중치를 더하는 희소성을 고려하여 전문가를 판별한다. 제안하는 기법은 대량의 논문데이터를 저장하고 이를 기반으로 전문가를 판별해야하므로 빅 데이터 처리 시스템을 기반으로 한다. 연구자들은 사용자 인터페이스를 통해 특정 분야에 대한 검색을 요청하면 서버는 분야에 대한 전문가를 검색한다. 뿐만 아니라 전문가와 다른 전문가 간의 관계, 검색된 전문가가 작성한 논문과 같이 연계되는 정보를 쉽게 제공받을 수 있다. 사용자는 전문가가 작성한 논문, 분야에서 가장 이슈가 되는 논문, 분야를 대표하는 학회, 분야에서 가장 많이 언급되는 토픽과 같은 기존 학술검색에서는 제공하지 못하였던 의미 있는 정보를 받아 볼 수 있다. 사용자는 이를 활용하여 질 높은 연구를 수행할 수 있게 된다.

본 논문의 구성은 다음과 같다. I장에서는 본 논문의 연구배경을 설명하고 II장에서는 기존에 연구와 학술검색 사이트에 대해 분석한다. 그리고 III장에서는 본 논문에서 제안하는 전문가 검색 서비스의 구성과 시스템 구조에 대해 설명한다. IV장에서는 제안한 시스템을 이용한 다양한 서비스의 기능에 대해 설명한다. V장에서는 제안하는 시스템의 성능을 평가하고 마지막으로 VI장에서는 본 논문에 대한 결론을 제시한다.

II. 관련 연구

구글 스칼라(2004)는 세계적으로 가장 유명한 검색 사이트인 구글에서 제공해주는 학술검색 사이트이다[1]. 구글 스칼라는 기본적인 논문 정보와 더불어 논문의 인용수와 다양한 정렬 알고리즘을 제공하고 있다. 키워드 일치 순으로 정렬을 하거나 년도 순으로 검색결과를 정렬 시킬 수 있다. 또한, 사용자가 직접 저자정보를 등록하면 과학자의 생산성과 영향력 지표를 표현하는 h-Index 지표를 제공한다[11].

한국학술지 인용색인(2004)은 국내 학술지 정보와 참고문헌을 DB화하여 논문간의 인용관계를 분석하는 시스템이다[2]. 게재논문과 국낸 학술지에 대한 정보를 제공하고 있고 인용빈도에 따른 학술지의 영향력을 평가하고 있다. 이를 활용하여 학술지의 영향력을 계산하여 사용자들에게 제공하고 있다. 최근에는 논문의 유사도 검사 서비스를 제공하여 논문의 표절 판단에 활용되기도 한다.

DBpia(2000)는 자체적으로 이용 수라는 데이터를 제공하고 있다[3]. 이용 수는 사용자가 논문을 파일로 다운로드 받는 수를 의미한다. 사용자는 이용 수라는 정보를 활용하여 개인적인 논문의 선호도에 반영할 수 있다. 최근에는 연관검색 논문이라는 기능이 추가됨으로써 검색한 논문과 관련된 논문들을 제공해준다.

DBLP(2005)는 300만 건 이상의 국제 논문들을 검색할 수 있는 유용한 사이트이다[4]. 그러나 사용자가 입력한 검색어와 완전히 일치하는 부분만 검색이 되며 연구자들이 검색한 논문의 전문성/연관성에 대한 판단을 쉽게 할 수 없는 단점이 존재하고 있지만 다른 검색사이트로의 연결을 통해 이를 보완하고 있다.

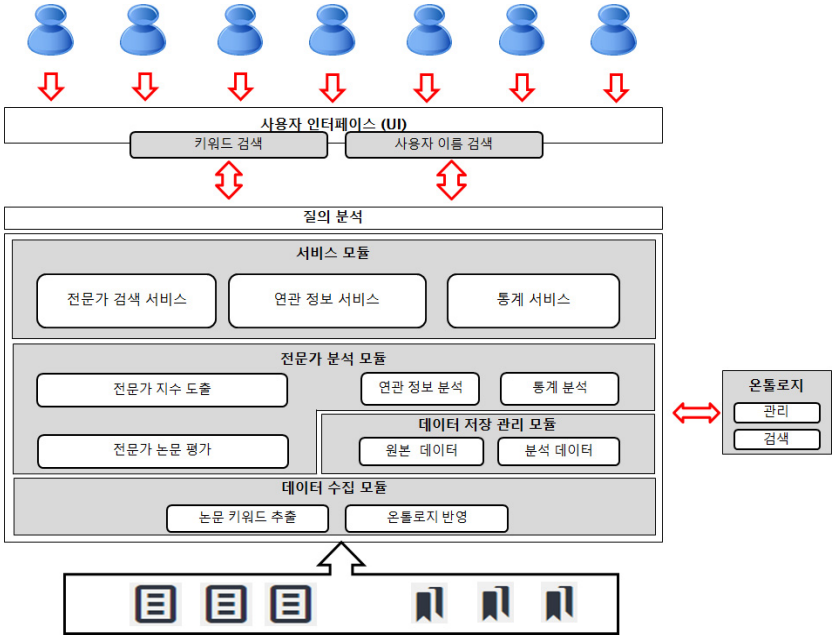

그림 1. 전문가 검색 시스템 구성도

Research Gate(2008)는 학술검색 기능에 소셜 네트워크 기능을 추가한 검색 사이트이다[5]. 사용자들은 회원가입을 통해 소셜 네트워크에 참여할 수 있고 다양한 사용자와 관계를 맺고 질문을 할 수 있다. 사용자는 사이트에 가입함과 동시에 자신의 연구 분야를 등록한다. Research Gate는 사용자의 연구 분야를 기반으로 다른 사용자가 질문한 게시물에 대해 연결시켜주는 기능이 있다. 또한, 구글 스칼라와 같이 h-Index와 더불어 사용자 평판 기반의 RG Score (Research Gate Score)를 제공한다. RG Score는 논문이 읽힌 수, 팔로우 수 Research Gate에 기여한 수(답변 수)를 기반으로 측정한 수치이다.

최근에는 소셜 네트워크에서 소셜 활동기반의 전문가 검색 기법이 연구되고 있다[9][10]. 일부 연구에서는 페이스북과 트위터와 같은 대중적인 소셜 네트워크 서비스에서 사용자가 게시한 글을 분석하여 검색한 질의와의 유사성을 기반으로 전문가 검색 기법을 제안하고 있다[9]. 마지막으로 온라인 과학 커뮤니티에서 논문의 품질과 사용자의 평판, 저자의 신용도를 기반으로 전문가를 판별하는 연구도 진행되고 있다[10].

이러한 연구자의 영향력을 표현하는 수치인 h-Index와 RG Score는 분야의 전문가를 찾아내는데 다음과 같은 한계가 있다. 먼저, h-Index는 연구자가 작성한 모든 논문을 기반으로 전체적인 영향력을 계산한 수치이기 때문에 특정 분야의 영향력에 대해서는 분별할 수 없는 단점이 존재한다. RG Score 또한 사용자 평판과 사용자의 소셜 활동 수를 기반으로 측정되는 수치이기 때문에 소셜 활동이 적은 사용자는 낮은 점수를 받을 확률이 높고 이는 좋은 논문을 쓴 사용자를 전문가로 판단하지 못할 확률이 높다. 따라서 본 논문에서는 두 지표가 가지는 한계점을 해결하기 위해 논문과 분야 정보를 기반으로 분야별 전문가 지수를 측정하는 새로운 지표를 제안한다. 이를 바탕으로 분야별 전문가를 판단 할 수 있게 된다.

III. 제안하는 전문가 검색 시스템 설계

1. 전문가 검색 서비스 구조

기존의 대부분 학술검색 사이트는 메타데이터 기반

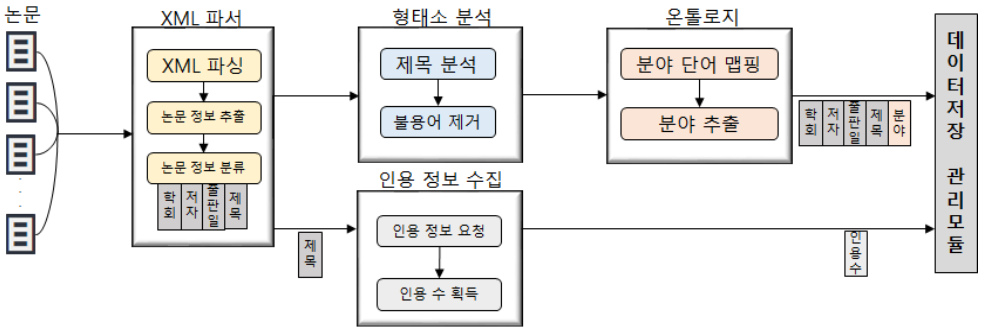

그림 2. 데이터 수집 모듈 수행 과정

의 논문 검색을 목적으로 하고 있기 때문에 저자의 분야와 분야의 전문가를 판별해주지 않고 있다. 분야의 전문가를 제공해주기 위해서는 분야와 전문가를 판별하는 기법이 필요하고, 더불어 대용량의 논문 데이터에서 분석을 해야 하므로 대용량 처리를 위해 설계되어야 한다. 제안하는 기법은 논문의 제목을 사용하여 분야를 자동으로 추출하고, 논문의 품질을 측정하여 분야의 전문가를 검색하는 기법을 제공한다. 대용량 처리를 효율적으로 수행하기 위해서 빅 데이터 처리 시스템을 기반으로 설계하였다. 처리 된 결과들은 웹 페이지와 같은 사용자 인터페이스를 통해서 분야의 전문가를 제공받거나 검색한 분야 혹은 전문가의 연관정보를 서비스 받을 수 있다. 제안하는 전문가 검색 시스템 구조는 다음 [그림 1]과 같다. 데이터 수집, 온톨로지, 데이터 저장관리, 전문가 분석, 그리고 서비스 모듈로 이루어진다. 데이터 수집은 전문가 검색하기 위한 기반 데이터를 수집하고, 온톨로지는 분야를 자동으로 추출하기 위해 사용한다. 데이터 저장관리는 수집/처리된 데이터를 저장하는 데 사용하고 서비스는 사용자에게 다양한 서비스를 제공하기 위해 사용된다. 마지막으로 사용자는 사용자 인터페이스를 통해 서비스를 제공받을 수 있다.

2. 데이터 수집

전문가를 판단하기 위해서는 전문가 별로 전문성을 나타내는 지표 혹은 수치가 필요하다. 본 논문에서는 전문가 수치를 전문가 지수라고 표현한다. 전문가 지수를 측정하기 위해서 논문의 정보를 활용하게 되는데 이러한 기반 데이터를 수집하는 기능을 하는 것이 데이터 수집 모듈이다. 데이터 수집 모듈은 논문의 기본 정보와 논문의 인용 정보를 수집한다. 전문가 지수는 수집된 논문의 정보를 활용하여 분야별로 지수를 측정하게 된다. 논문의 분야를 특정 지을 수 있다면 분야별 전문가 지수를 쉽게 측정할 수 있다. 그러나 대부분의 논문은 분야를 논문 내에 표시하고 있고, 데이터 수집 모듈에서는 논문의 내용 혹은 요약 정보를 수집할 수 없어 기본적인 정보만으로 논문의 분야를 판단해야만 한다.

본 논문에서는 논문의 분야를 판단하기 위해서 온톨로지와 형태소 분석기를 활용한다. 온톨로지에는 분야별로 대표 할 수 있는 단어들을 내포한다. 논문이 수집되면 논문의 제목을 형태소 분석기를 통해 단어를 추출하게 되고 단어가 온톨로지 에서 정의된 분야의 대표적인 단어와 일치하는지 확인한다. 만약 일치한다면 해당 논문은 그 분야를 내포한다고 판단한다. 하나의 논문에는 여러 가지의 분야를 포함 할 수 있기 때문에 모든 단어에 대해서 온톨로지 검색을 수행한다. [그림 2]는 데이터 수집 모듈의 전체적인 수행과정이다. 논문 데이터가 수집되면 XML (Extensible Markup Language) 파서를 통해 논문제목과 기타 논문정보를 추출한다. 기타 정보는 데이터를 분류한 후에 데이터 저장관리 모듈에 전달한다. 논문 제목정보는 이전에 언급했던 분야 추출 기법을 통해 분야를 추출한 후에 데이터 저장 관리 모듈에 전달한다. 그리고 추가적으로 논문의 제목을 이용하여 인용 정보를 웹을 통해 획득한 후 인용 정보를 저장한다.

3. 데이터 저장 관리

수집된 데이터와 분석 모듈에서 분석 한 결과 데이터를 사용자에게 서비스하기 위해서는 안정적인 저장 모듈이 필요하다. 데이터 저장 관리 모듈은 영속성과 분산 저장을 지원하기 위한 모듈이다. 3.2절에서 수집한 데이터는 원본 데이터에 저장되고 3.4절에서 분석한 데이터는 분석 데이터에 저장된다. 이를 구분하는 이유는 데이터에 따라 저장되는 특성이 다르기 때문이다. 원본 데이터는 영속성을 제공하는 것이 목표이기 때문에 복제 본을 여러 개 유지하게 된다. 분석 데이터는 원본 데이터보다 사용자가 접근 요청이 많은 데이터이기 때문에 영속성 보다는 응답성이 더 높이 요구된다고 볼 수 있다. 그래서 분석 데이터는 응답성을 최대한 제공해줄 수 있는 저장소에 저장하여 관리 한다. [그림 3]은 데이터 저장관리 모듈의 전체적인 동작과정이다. 원본 데이터로 키워드와 인용 데이터, 학회, 저자, 출판 정보를 저장한다. 그런 다음 전문가 분석 모듈은 원본데이터를 활용하여 분석을 수행한다. 최종적으로 분석 결과는 다시 데이터 저장관리 모듈에 분석 데이터로 저장한다. 분석 정보는 전문가 지수, 연관 정보, 통계 정보로 구성되며 서비스 요청에 따라 사용자에게 제공한다. 전문가 지수는 전문가를 검색할 때와 전문가의 지수 이력을 제공하는데 사용된다. 연관 정보는 전문가간의 관계 정보를 제공하고 분야와 관련된 논문을 제공하는 연관 논문 서비스에 사용된다. 마지막으로 통계 정보는 분야와 관련된 학회의 출판 통계와 분야의 핫 토픽 서비스를 제공하는데 사용된다.

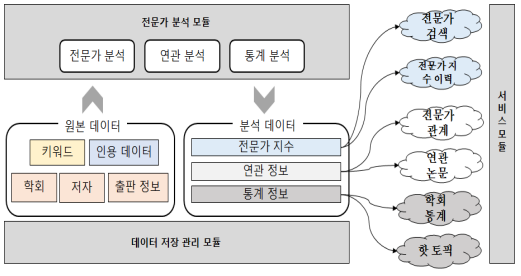

그림 3. 데이터 저장 관리 모듈 수행 과정

4. 전문가 분석

본 논문에서는 계산한 전문가 지수가 높은 저자를 전문가라고 한다. 전문가 분석 모듈은 수집된 데이터의 모든 저자의 분야별 전문가 지수를 계산하는 기능이다. 뿐만 아니라 전문가 간의 관계 등 6가지 서비스를 제공하기 위해 다양한 분석을 수행한다. 분석 모듈은 데이터 수집 모듈에서 수집한 데이터를 바탕으로 다양한 분석을 수행한다. 전문가 분석은 다음 [그림 4]와 같이 이루어진다.

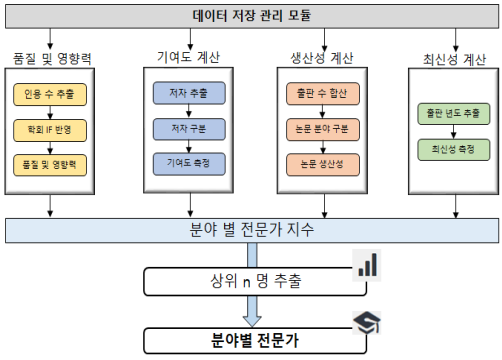

그림 4. 전문가 분석 수행 과정

먼저, 데이터저장 관리 모듈에서 분야, 인용 수, 학회 IF (Impact Factor), 저자, 출판 정보를 추출한다. 각 데이터는 지수를 측정하는 데에 중요한 의미를 가지는 데이터이다. 인용 수와 학회의 IF는 논문의 전체적인 품질과 영향력을 판단하기 위해 사용하고, 저자정보는 저자구분과 저자 수에 따라 논문의 기여도를 판단하는데 사용한다. 논문의 출판 수는 연구자의 생산성과 일맥상통하고 논문의 출판 연도는 논문의 최신성, 즉 얼마나 최근에 게재되었는지를 확인할 수 있다.

수식 (1)은 위에서 설명한 전문가 분석 수행 과정을 하나의 계산식으로 표현한 것이다. 모든 저자는 분야별로 전문가 지수를 측정한다. 수식에서 n은 해당분야에서 저자로 포함된 논문의 총 수이다. 저자가 포함된 총 논문의 수는 생산성($\sum$)으로 표현될 수 있다. 그리고 모든 논문별로 품질(Quality), 영향력(Leverage), 기여도(Contribution), 최신성(Freshness)을 계산하여 최종

적으로 합산하게 되면 저자의 분야별 전문가 지수를 계산 할 수 있다.

$$Expert\_Score_{keyword} = \sum_{i=1}^{n}(Quality_i + Leverage_i + Contribution_i + Freshness_i) \quad (1)$$

논문의 품질은 논문의 인용 횟수와 학회의 IF를 활용하여 계산한다. 논문의 인용 횟수는 다른 연구자들이 논문을 얼마나 많이 참조하는가에 대한 수치이므로 인용 수가 높으면 논문의 품질/영향력이 높다고 볼 수 있다. 또한, 출판 되는 학회의 IF가 높다 면 마찬가지로 논문의 품질과 영향력이 높다고 볼 수 있다.

논문에 대한 기여도는 각 논문의 저자 정보를 활용한다. 논문은 주 저자, 공저자, 교신저자로 이루어져있는데 주 저자와 교신저자는 논문의 기여한바가 공저자보다 많기 때문에 각 저자구분에 따라 논문의 기여율을 차등하게 측정한다. 제안하는 기법은 주 저자와 교신저자를 구분하지는 못한다. 따라서 주 저자와 공저자 두 개의 구분과 저자의 수를 이용하여 논문의 기여도를 측정한다.

논문의 생산성은 앞서 기술 한 것과 같이 해당 분야에서 출판된 논문의 양을 합산하면 그 연구자의 생산성이라고 볼 수 있다. 마지막으로, 논문의 최신성은 출판 연도를 기반으로 측정한다. 대부분의 논문은 오래 된 논문이 최신에 나온 논문보다 인용수가 높은 경향이 있다. 당연히 최신에 나온 논문은 아직 다른 연구자가 아직 읽지 못한 경우가 많기 때문이다. 그래서 만약 인용수를 이용하여 논문의 품질만 측정한다면 오래된 논문을 작성한 저자가 전문가가 될 확률이 크다. 이를 보완하기 위해서 출판 연도 정보를 활용하여 최신성이라는 가중치를 부여한다. 이는 최신에 출판된 논문의 지수측정을 보완해주는 역할을 한다. 최종적으로 이렇게 측정된 각 지표와 논문의 분야 정보를 활용하여 분야별 전문가 지수를 측정할 수 있다. 본 논문에서는 측정된 전문가 지수에서 상위 n명을 뽑고 추출된 n명을 전문가라고 판단한다.

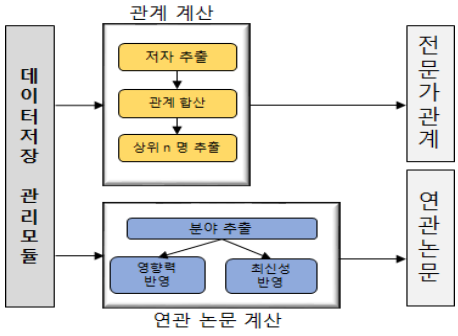

그림 5. 연관 정보 분석 수행 과정

그림 5는 연관 정보 분석 절차이다. 연관 정보 분석은 전문가 관계와 연관 논문 분석으로 이루어져있다. 전문가 관계는 논문의 저자정보를 활용하여 저자간의 관계를 가시화하기 위한 기능이다. 저자간의 관계는 저자 정보만으로도 표현 할 수 있지만 분야별로 저자관계가 다를 수 있기 때문에 분야정보를 추가적으로 사용하였다. 본 논문에서는 저자의 관계를 전문가 측정과 유사하게 자주 관계가 형성되는 상위 n명만을 표시한다. 연관 논문은 사용자가 검색한 분야에 관련 있는 논문들을 제공하기 위한 기능이다. 연구자들은 검색한 분야에 대한 전문가가 누군지를 알고 싶기도 하지만 분야에 연관 있으면서 영향력 있는 논문을 찾길 원한다. 본 논문에서는 논문의 분야를 특정 지을 수 있기 때문에 분야별로 가장 연관성 있는 논문을 쉽게 알 수 있다. 논문의 출판 연도와 인용 수, 분야 정보를 이용하여 연관 논문을 분석한다. 인용 수 는 연관성 있는 논문 중에서 영향력이 높은 논문을 찾고 싶을 때 사용되고, 출판 연도는 최신 논문 순으로 찾고 싶을 때 사용된다.

[그림 6]은 통계 분석 절차이다. 통계 정보는 학회 통계와 핫 토픽 분석으로 이루어져 있다. 학회 통계는 분야와 학회 명을 이용하여 학회에서 출판된 논문수를 합산하여 통계를 내린다. 핫 토픽 분석은 분야별로 어떤 단어가 의미 있는지 분석하는 기능이다. 논문의 제목 정보를 수집단계에서 했던 것과 동일하게 형태소 분석기를 통해 다시 단어 단위로 분할하고 이를 TF-IDF (Term Frequency - Inverse Document Frequency) [12-14] 기법을 통해 의미 있는 단어를 추출한다. 마지

막으로 분야 정보를 취합하여 분야별로 의미 있는 단어 즉 핫 토픽을 최종적으로 계산한다. 핫 토픽의 자세한 기법은 4.2절에서 설명한다.

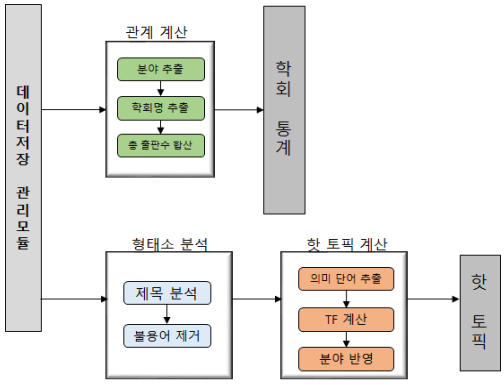

그림 6. 통계 분석 수행 과정

IV. 제안하는 전문가 검색 시스템 구현

1. 구현 환경

제안하는 전문가 서비스 구현 환경은 다음 [표 1]과 같다. 제안하는 시스템은 분산 처리를 위해서 3개의 노드로 구성하였다. 웹 서버는 Apache 웹 서버와 웹 서비스를 위해 Glassfish를 사용하였다. 사용자와의 통신은 REST (Representational State Transfer) 방식으로 연동하였으며 메시지는 JSON (JavaScript Object Notation) 방식으로 통신을 한다. 온톨로지는 Jena 플랫폼을 사용하여 구성하였다. 데이터 저장을 위한 저장소는 먼저 분산 저장을 위한 HBase와 다음으로 사용자 응답성을 위한 PostgreSQL 두 가지를 사용하였다. 그리고 빅데이터 분석 처리를 위해서 Spark 플랫폼을 사용하였다. 데이터 수집은 DBpia에서 제공하는 OpenAPI를 이용하여 수집하였다. 추가적으로 인용 정보를 수집하기 위해 Google Scholar API (Application Programming Interface)를 사용하여 수집하였다.

표 1. 구현 환경

구성	사용 요소
프로세서	Intel i5-3570k, 3.4GHz, 4Core
메모리	4GB DDR3
노드 수	3
서버 (웹, 온톨로지)	Apache, Glassfish, Jena
데이터베이스 (NoSQL, RDBMS)	HBase 1.1.3[15] PostgreSQL 9.4
분산 처리 시스템	Spark 1.6.1[16]
데이터 수집	JAVA, DBpia API, Google Scholar API

[그림 7]은 제안하는 시스템의 물리적 구성도이다. Node1은 Hadoop, HBase, Spark의 마스터 역할을 한다. Node2, 3은 슬레이브 역할을 수행한다. Node1은 추가적으로 웹 서버와 웹 서비스를 제공하기 위해서 Apache와 Glassfish를 추가적으로 탑재하였다. 또한, 온톨로지를 사용하기 위해서 온톨로지 저장소를 추가적으로 가지고 있다. PostgreSQL은 Node1의 부하를 줄이기 위해서 Node3에 설치하였다.

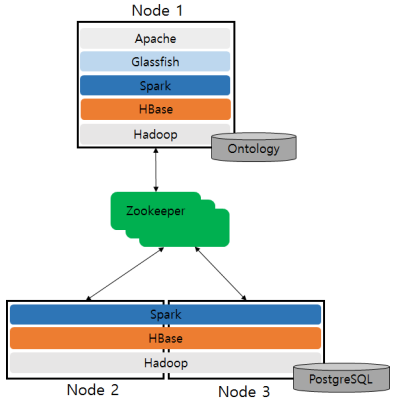

그림 7. 물리적 시스템 구성도

2. 서비스 내용

2.1 검색

분야별 전문가 검색을 하기 위해서는 분야별로 검색을 하는 기능이 필요하다. 전문가 검색 서비스는 두 가지의 검색 기능을 제공하는데 첫 번째는 키워드를 이용한 검색 기능이다. 키워드 검색은 사용자가 입력한 키워드를 이용하여 분야별 전문가를 검색한다. 두 번째는

이름을 이용한 검색으로써 해당 전문가의 상세한 정보를 확인할 수 있다. 두 가지의 검색은 검색 창 옆의 스위치 형태의 UI (User Interface)를 구성하였다. 다음 [그림 8]에서 좌측은 키워드를 이용한 검색이고 우측은 이름을 이용한 검색이다.

그림 8. 검색 기능 (좌 : 키워드 검색, 우 : 이름 검색)

아래 [그림 9]는 키워드 검색을 수행했을 때의 결과 화면이다. 검색어는 '센서'를 입력하였고 사용자는 센서 분야의 전문가 10명과 선택된 전문가의 관계 서비스와 전문가의 지수이력 서비스를 받을 수 있다. 추가적으로 제공되는 연관 정보 서비스와 통계 서비스는 이후 절에서 설명한다.

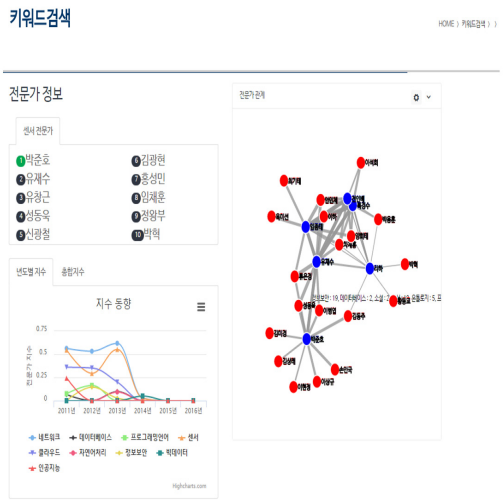

그림 9. 키워드 검색 결과 화면

[그림 10]은 이름 검색을 수행했을 때의 결과화면이다. 특정 전문가의 이름을 입력하면 키워드 검색과 마찬가지로 지수이력서비스와 전문가 관계서비스를 받을 수 있다. 이와 더불어 해당 전문가가 작성한 논문을 상단에 표시하여 전문가가 작성한 논문을 바로 찾아 볼 수 있게 제공하였다.

그림 10. 이름 검색 결과 화면

2.2 전문가 관계

전문가간의 관계 서비스는 논문의 저자정보를 이용하여 저자간의 관계성을 네트워크 형태로 표현하는 기능이다. 전문가 관계를 파악하면 전문가와 관련된 의미있는 저자를 파악 할 수 있다. 다음 [그림 11]은 전문가 관계 서비스 결과이다.

전문가 관계는 파란색/빨간색의 두 가지 형태의 노드(Node)와 회색 실선의 엣지(Edge)로 구성된다. 노드는 한명의 전문가를 의미하며 파란색은 이미 표현이 완료된 노드이고 빨간색은 표현이 좀 더 가능한 노드를 의미한다. 빨간색 노드를 선택을 하게 되면 선택된 노드를 중심으로 다시 상위 5명의 추가적인 저자가 표현 된다.

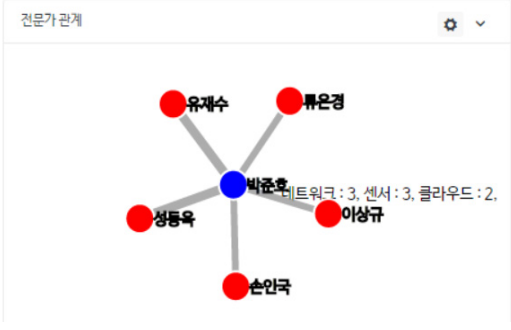

그림 11. 전문가 관계 서비스

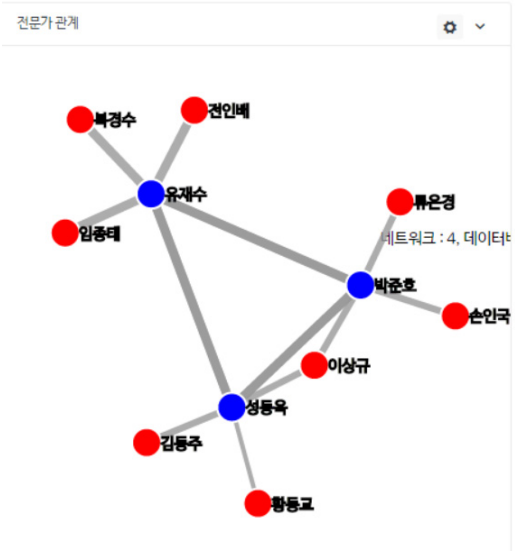

그림 12. 확장된 전문가 관계 그래프

[그림 12]는 추가적으로 확장된 형태의 관계 그래프이다. 그래프에서 엣지는 전문가간의 연관성과 가중치를 의미한다. 관계가 많이 형성될수록 엣지는 더욱 두꺼워지고 관계가 적을 경우에는 엣지는 얇아진다. 추가적으로 관계를 자세하게 알고 싶을 때는 엣지에 마우스 오버를 할 경우 관계에 대한 상세 정보가 나타난다. 예를 들어 [그림 11]에서 '박준호'와 '이상규' 는 '네트워크' 키워드 논문을 3편 '센서' 키워드 논문을 3편 '클라우드' 키워드 논문을 2편 작성한 사이라는 의미이다.

2.3 전문가 지수 이력

전문가 지수 이력 서비스는 본 논문에서 제안하는 전문가 지수 알고리즘을 통해 계산된 수치를 분야별로 이력을 제공해주는 서비스이다. 이력 서비스를 통해 전문가의 연구동향을 쉽게 파악 할 수 있다. [그림 13]은 전문가 지수 이력 서비스 결과이다. 검색된 전문가의 지수를 분야별, 년도 별로 그래프 형태로 표현한다.

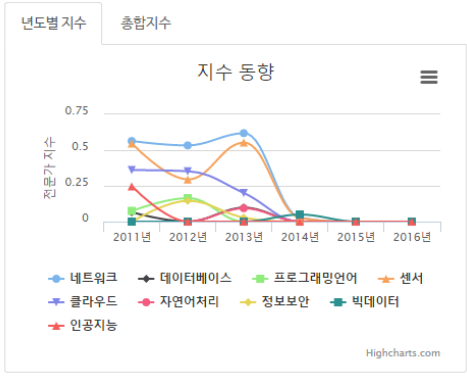

그림 13. 전문가 지수 이력 서비스 (년도별)

전문가 지수 이력 그래프는 HighCharts [17]에서 제공하는 API를 이용하여 구현 했다. 해당 API는 다양한 형태의 그래프를 제공해주고 표현된 그래프를 이미지나 PDF (Portable Document Format) 형태로 쉽게 다운로드 받을 수 있다. 전문가 지수 이력은 어느 분야를 더욱 치중했는지 파악 할 수 있게 총합 지수 그래프도 제공한다. 아래 [그림 14]는 분야별 총합 지수 그래프이다. 분야별 치중도를 쉽게 파악할 수 있게 하기 위하여 파이그래프 형태로 표현하였다.

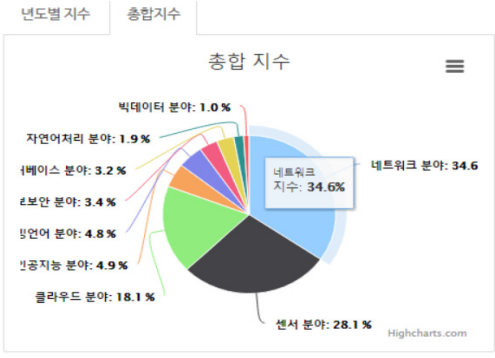

그림 14. 전문가 지수 이력 서비스 (총합)

2.4 전문가 상세 논문

사용자가 검색 한 전문가에 대해 전문가 지수 이력 서비스뿐만 아니라 해당 전문가가 작성한 논문을 제공해주는 것도 중요한 기능 중 하나이다. [그림 15]는 검

색한 전문가에 대한 전체 논문이다. 사용자가 선택한 혹은 검색한 전문가의 상세 논문이 나타난다. 전문가 상세 논문은 테이블 형태로 제공되고, 최신 논문 순으로 확인 할 수 있다. 더불어 제목/저자/학회 명/발행번호/발행연도 등 특정 행을 지정하여 간단하게 검색을 수행 할 수 있다.

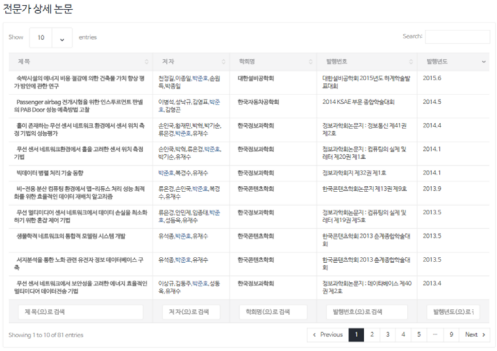

그림 15. 전문가 상세 논문

2.5 핫 토픽

핫 토픽 서비스는 수집 한 논문을 기반으로 논문의 제목에 많이 포함되는 키워드를 분석하여 사용자에게 가시화하는 서비스이다. 핫 토픽 분석을 통해 현재 연구동향의 트렌드를 파악할 수 있다. 핫 토픽을 분석하는 기법은 대중적으로 알려진 TF-IDF기법이 있다. DBpia OpenAPI 에서는 논문의 키워드 정보와 요약정보를 제공하지 않기 때문에 TF-IDF 기법을 적용하기는 쉽지 않다. 그래서 본 논문에서는 논문의 제목을 기반으로 한 핫 토픽 추출 기법을 사용하였다. 그러나 제목을 기반으로 TF-IDF를 적용 하더라도 의미 있는 단어를 추출하기는 힘들다. 그 이유는 논문 제목에는 자주 사용되는 단어들이 많이 모여 있고 매우 요약이나 본문에 비해 매우 짧은 내용이라 의미 있는 단어를 추출하기가 매우 어렵다. 그래서 본 논문에서는 해당 분야에서 미리 의미 있다고 판단되는 단어들에서 TF만을 적용하여 핫 토픽을 추출 하였다.

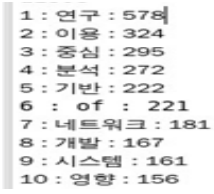

그림 16. TF 중심의 핫 토픽 결과

[그림 16]은 전처리 과정 없이 TF만을 적용을 할 때의 결과이다. '연구', '이용', '중심', '분석' 과 같이 논문에서 대중적으로 많이 사용되는 단어들을 핫 토픽으로 판별하게 된다. 그래서 '네트워크' 와 같은 의미 있다고 판단되는 단어들을 온톨로지에 수집한 후 의미 있는 단어 기준으로 핫 토픽을 검출 하였다. 다음 [그림 17]은 '센서' 분야의 핫 토픽 결과이다. 핫 토픽 결과 '센서'라는 단어가 가장 많이 출현하여 가장 크게 나타났다. 추후에는 요약정보 기반의 핫 토픽을 적용하여 보다 의미 있는 단어를 추출해나갈 예정이다.

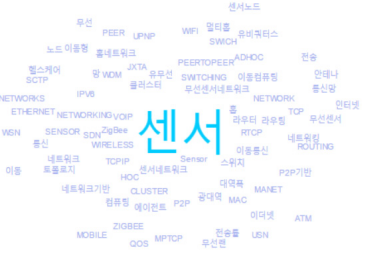

그림 17. 핫 토픽 서비스

2.6 연관 논문

사용자가 검색한 키워드와 관련된 논문을 알 수 있다면 논문 검색에 많은 시간을 절약할 수 있다. [그림 18]은 '소셜' 분야와 관련된 연관 논문들을 테이블 형태로 제공하는 연관 논문 서비스 결과이다. 연관 논문 서비스는 좀 더 의미 있는 정보를 제공하기 위해서 구글 스칼라의 인용 정보를 기반으로 인용수가 높은 순으로 결과를 보여준다. 연관 논문 서비스도 전문가 상세 논문 서비스와 마찬가지로 행에 대한 검색이 가능하다.

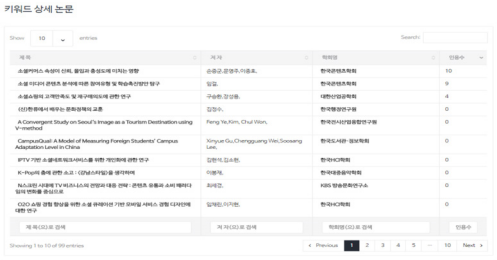

그림 18. 연관 논문 서비스

2.7 학회 통계

학회 통계 서비스는 어떤 학회가 분야별로 논문이 많이 접수되고 있는지 통계 정보를 확인하는 서비스이다. 분야는 본 논문에서 제안하는 기법으로 판단하였기 때문에 정확한 분야라고 특정 짓기는 힘들지만 추세가 어떠한지는 대략적으로 알 수 있는 정보이다. 다음 [그림 19]는 '소셜' 분야의 학회 통계 서비스이다. 가로축은 학회를 의미하고 세로축은 발표된 논문의 수이다. 8개의 상위 학회를 보여주며 사용자는 이러한 정보를 활용하면 어떠한 학회가 분야의 전문 학회일지 유추 해볼 수도 있다.

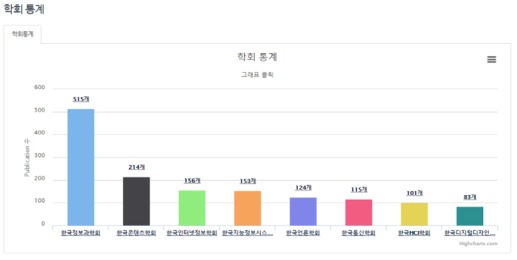

그림 19. 학회 통계 서비스 (종합)

학회 통계 서비스는 이력 정보를 제공함으로써 학회의 현재 출판 추세를 파악할 수 있다. [그림 20]은 학회별 이력정보이다. 이력 정보는 년도 별로 발표 된 논문의 수를 막대그래프 형태로 표현한다. 가로축은 년도이고 세로축은 발표된 논문의 수이다. 약 10 여 년 전에는 많이 출판 되지 않던 논문이 최근에 100개 이상의 논문을 출판했다는 것을 확인함으로써 최근 이 분야의 연구가 많이 되거나 학회가 주로 출판한다고 볼 수 있다.

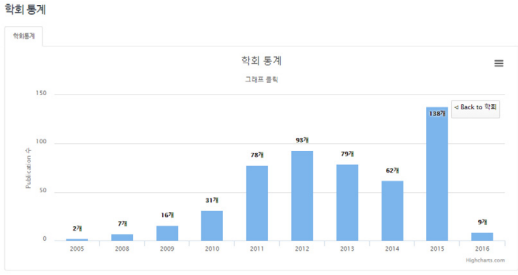

그림 20. 학회 통계 서비스 (이력)

V. 성능 평가

제안하는 전문가 판별 기법의 우수성을 검증하기 위해서 실제 전문가가 전문가로 검색되는지 정확성을 측정하였다. 기존 전문가 판별 기법은 저널의 IF와 최신성, 인용수를 기반으로 전문가를 판별한다[18]. 제안하는 전문가 판별기법은 저자의 논문 생산성과 논문의 기여도를 추가적으로 고려하였다. 성능평가는 전체 논문 수 106,095건에서 소셜 분야의 논문인 1,791건을 기반으로 측정하였다. 논문 저자는 총 147,326 명 중에 해당 분야의 저자인 2,363명을 기반으로 실험을 진행하였다.

본 논문에서는 기존 기법과 제안하는 기법의 F-Score를 기반으로 정확도를 비교하였다. F-Score는 재현율(Recall)과 정밀도(Precision)의 조화평균 값으로 검색 시스템에서 검색 결과의 정확도를 측정하는 대중적인 방법이다. 재현율은 실제 전문가가 검색된 비율을 의미하고 정밀도는 검색된 결과에서 전문가가 얼마나 포함되어있는가를 의미한다.

실험에서 나오는 정답 데이터 셋은 88명으로 이루어져있다. 정답 데이터 셋은 다음과 같이 3개의 결과값에서 모두 포함된 결과 값을 기반으로 만들었다. 세 개의 랭킹결과는 첫 번째로 인용 수가 높은 순으로한 랭킹 결과와 두 번째는 인용 수와 저널의 IF를 같이 고려한 랭킹 결과, 마지막으로 세 번째는 저널의 IF, 최신성, 인용수를 모두 고려한 랭킹결과이다.

[그림 21]은 전문가 추천 수에 따른 F-Score 지수 이다. 검색되는 전문가의 수는 60명부터 80명까지 늘려가며 수행하였다. 성능평가 결과 제안하는 전문가 판별

기법은 기존 기법에 대비하여 전문가 추천 수에 따라 각 11%, 13%, 21%만큼 정확도가 향상된 것을 확인 할 수 있었다. 기존 기법과 제안하는 기법 모두 전문가 추천 수 에 따라 정확도가 향상되는 것을 확인 할 수 있었다. 그러나 기존 기법에 생산성과 기여도가 반영된다면 더욱 높은 정확도를 보일 수 있다는 것을 확인 할 수 있었다. 따라서 논문의 생산성과 논문의 기여도는 전문가 판별에 있어서 의미 있는 인자임을 확인할 수 있었다.

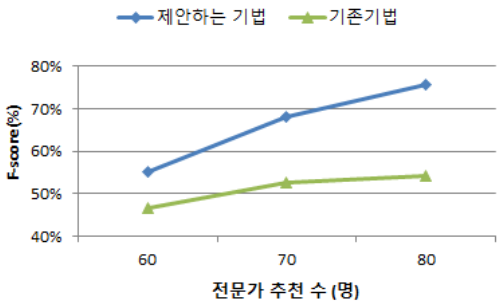

그림 21. 성능 평가 결과

VI. 결 론

본 논문에서는 기존 학술검색 사이트에서 제공하지 못하였던 분야별 전문가를 검색하는 전문가 검색 시스템을 설계하고 구현하였다. 제안하는 시스템은 대용량의 데이터 처리를 위해 빅 데이터 시스템을 기반으로 한다. 사용자는 전문가 검색 시스템을 웹 페이지를 통해 쉽게 접근하여 서비스를 받아볼 수 있다. 제안하는 시스템은 기존의 소셜 활동이 필요하였던 전문가 판별 기법들과 달리 논문 정보만을 이용했기 때문에 소셜 활동이 적은 연구자도 전문가로 고려할 수 있다. 제안하는 시스템은 전문가에 대한 검색뿐만 아니라 전문가와 자주 연관된 전문가, 분야별 핫 토픽과 같은 가공된 정보를 제공하여 사용자가 논문을 작성하는데 더욱 많은 도움을 줄 수 있을 것으로 기대한다. 제공하는 서비스는 그래프나 표와 같이 사용자가 한눈에 알아 볼 수 있는 UI들을 사용하여 사용자의 편의성을 향상시켰다.

본 연구의 한계점은 다음과 같다. 먼저, 제안하는 기법은 논문의 제목정보만으로 분야를 구별하고 수동적으로 온톨로지를 구성해야한다. 만약 특정 분야에 새로운 연구 중심어가 추가되면 개발자가 이 연구 중심어를 수동으로 온톨로지에 반영해야만 한다. 더불어서 새로운 연구 분야가 생성되었다면 위와 같은 작업을 다시 반복해야만 한다. 즉, 새로운 연구 분야를 자동적으로 판단할 수 없는 한계점이 존재한다. 두 번째로 논문의 메타데이터 수집의 한계로 인해 제목만으로 분야를 구별해야하고 아무리 수동으로 중심어를 선정한다고 하더라도 이 중심어가 해당 분야에 포함된다고 확신할 수 없는 문제점이 존재한다. 그러나 이는 추후에 다른 학술사이트에 적용이 된다면 요약정보나 중심어 정보를 획득할 수 있을 것이라 판단하고, 이러한 정보는 온톨로지 구성에 많은 도움이 될 것으로 기대한다. 마지막으로, 제안하는 시스템은 전문가 지수를 계산할 때 주저자와 공저자를 기반으로 논문의 기여도를 판단하는데 공저자가 많으면 많을수록 논문의 점수가 떨어지고, 어떠한 분야에서는 저자 구분에 따른 기여도가 큰 차이가 없는 분야도 존재한다.

현재 연구 분야에 관련된 중심어 온톨로지는 수동으로 구성해야만 한다. 향후에는 논문의 요약정보와 논문 자체의 중심어 정보를 수집하여 의미 있는 중심어가 자동적으로 온톨로지에 반영할 수 있게 개선해나갈 것이다. 그리고 저자 구분과 저자 수에 따라 계산되는 전문가 지수의 편차가 크기 때문에 조금 더 합리적인 수식으로 개선해 나갈 필요성이 있다. 제안하는 시스템은 국내 논문 정보만으로 전문가를 판단하고 있기 때문에 국외 논문 정보를 추가적으로 수집한다면 국내외 전문가를 판별할 수 있는 시스템으로 향상 시킬 수 있을 것이다. 마지막으로, 사용자의 사용 이력 정보와 피드백 정보를 활용하여 전문가 검색에 대한 정확성에 검증을 해나가고 이를 다시 검색결과에 반영할 것이다.

참 고 문 헌

[1] https://scholar.google.co.kr
[2] https://www.kci.go.kr

[3] http://www.dbpia.co.kr
[4] http://dblp.uni-trier.de
[5] https://www.researchgate.net
[6] 한희준, 예용희, 류범종, “학술정보서비스에서 인명검색 고도화 방법,” 한국콘텐츠학회논문지, 제10권, 제2호, pp.490-498, 2010.
[7] 이민호, 이원구, 윤화묵, 신성호, 류재철, “해외 과학기술 학술논문 메타데이터의 비교 분석,” 한국콘텐츠학회논문지, 제11권, 제9호, pp.515-523, 2011.
[8] www.academia.edu
[9] J. Zhang, J. Tang, and J. Li, “Expert finding in a social network,” Proc. International Conference on Database Systems for Advanced Applications, pp.1066-1069, 2007.
[10] H. Liao, R. Xiao, G. Cimini, and M. Medo, “Ranking users, papers and authors in online scientific communities,” arXiv preprint arXiv:1311.3064, 2013.
[11] J. E. Hirsch, “An index to quantify an individual's scientific research output,” Proceedings of the National academy of Sciences of the United States of America, Vol.102, No.46, pp.16569-16572, 2005.
[12] S. E. Robertson, “Term specificity,” Journal of Documentation, Vol.28, pp.164-165, 1972.
[13] S. E. Robertson, “Specificity and weighted retrieval,” Journal of Documentation, Vol.30, No.1, pp.41-46, 1974.
[14] S. E. Robertson, “The probability ranking principle in information retrieval,” Journal of Documentation, Vol.33, No.4, pp.294-304, 1977.
[15] http://hbase.apache.org
[16] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, “Spark: cluster computing with working sets,” Proc. USENIX Workshop on Hot Topics in Cloud Computing, pp.10-16, 2010.
[17] http://www.highcharts.com
[18] X. Li and T. Watanabe, “Automatic Paper-to-reviewer Assignment, Based on the Matching Degree of the Reviewers,” Proc. International Conference in Knowledge Based and Intelligent Information and Engineering Systems, pp.633-642, 2013.

저 자 소 개

최 도 진(Dojin Choi) 준회원

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

<관심분야> : 연속 질의 처리, 그래프 스트림

김 민 수(Minsoo Kim) 준회원

- 2013년 2월 : 충북대학교 정보통신공학부(공학사)
- 2014년 9월 : (주) 매크로 임팩트 연구원
- 2015년 2월 ~ 현재 : 충북대학교 정보통신공학과 석사과정

<관심분야> : 빅데이터, RDF, 그래프, 고차원 인덱스

김 대 윤(Daeyun Kim) 준회원

- 2015년 2월 : 청주대학교 경영학과/컴퓨터공학과(공학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터협동과정 석사과정

<관심분야> : 데이터베이스 시스템, 복합이벤트 처리, 빅데이터

이 서 희(Seohee Lee) 준회원

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

<관심분야> : 소셜 네트워크 서비스, 빅데이터

한 진 수(Jinsu Han) 준회원

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

<관심분야> : 그래프 분산처리, 빅데이터

서 인 덕(Indeok Seo) 준회원

- 2016년 2월 : 청주대학교 통계학과(이학사)
- 2016년 3월 ~ 현재 : 충북대학교 빅데이터협동과정 석사과정

<관심분야> : 소셜 네트워크, 빅데이터

임 종 태(Jongtae 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

복 경 수(Kyoungsoo 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 초빙부교수

<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