

연구 역량 강화를 위한 지능형 지시적 분석 시스템

장윤지·홍승균 (한국과학기술정보연구원), 정한민 (과학기술연합대학원대학교)

- | | |
|----|--|
| 목차 | <ol style="list-style-type: none"> 1. 서론 2. 관련 연구 및 개념 정의 3. 지시적 분석 시스템 4. 관련 연구 동향 5. 결론 |
|----|--|

1. 서론

연구자들은 관련기술이나 연구를 찾는 것으로 자신만의 연구를 시작한다. 오늘날 과학기술 연구 분야에서는 하루가 다르게 새로운 연구 분야가 개척되고 많은 연구성과들이 쏟아지고 있으며, 과학기술의 발달에 따라 참조해야 할 관련문헌의 양도 점점 늘어나고 있다. 이러한 상황에서 기존 연구 분야의 문헌정보를 일일이 찾아보는 것을 넘어서 알고 싶은 과학기술의 흐름을 쉽게 알아볼 수 있는 기술 분석에 대한 사용자 요구가 증가하고 있다. 특히 연구자들의 경우 첨단 연구 분야에서 주도적인 위치를 자리 잡아 영향력을 행사한다면 좋은 연구성과를 낼 수 있는 경쟁력을 갖출 수 있기 때문에 체계적이고 신속한 분석이 요구된다. 한편, 통계에 따르면 연구자들은 전체 연구 활동 시간 중에 관련 선행 연구를 조사하는 데 절반 이상의 시간을 할애하는 것으로 나타났는데, 이는 연구자들이 자신의 연구

활동에 시간을 효율적으로 활용하지 못한다고 볼 수 있다. 그러므로 효율적인 연구 활동을 위해서 연구자들을 위한 기술 분석 서비스의 필요성이 더욱 절실하게 요구된다.

본 논문에서는 연구자의 역량 강화를 위해 과학기술 분석과 이를 기반으로 한 연구지원 지능형 정보시스템을 알아보려고 한다. 특히, 연구자가 효율적으로 연구 활동에 매진할 수 있도록 기술 분석뿐만 아니라 관련 연구자들의 연구 흐름까지 알려주어 최종적으로는 연구자가 동료 집단(Peer Group)에서 선도적인 위치를 차지할 수 있는 방향까지 제시하는 시스템을 알아보려고 한다. 이를 위해, 연구자에게 의사 결정 지원 및 미래의 행동을 추천하는 지시적 분석의 정보시스템으로 InSciTe(Intelligence in Science & Technology)의 발전 방향과 지능형 정보시스템이 나아가야 할 목표를 제시한다.

2. 관련 연구

비즈니스 분석(Business Analytics)은 기존까지 진행해 온 사업의 성과에 대한 분석 또는 시뮬레이션 등의 재귀적인 절차를 통해 현재 직면하고 있는 문제를 파악하고 이를 극복하기 위한 최적의 전략을 도출해내는 일련의 분석기법을 말한다. 비즈니스 분석은 크게 기술적 분석(Descriptive Analytics), 예측 분석(Predictive Analytics), 그리고 지시적 분석(Prescriptive Analytics)등으로 나눌 수 있다^[1].

기술적 분석은 과거에 수집한 데이터를 기반으로 미래를 예측하는 기본적인 분석으로 과거의 행동이 미래의 성과에 미치는 영향을 기계적으로 학습한다. 이를 토대로, 특정 목표의 달성을 위해 고려해야 할 여러 변수들 간의 연결고리를 찾는 경영 보고서가 대표적인 기술적 분석의 예라고 할 수 있다.

예측 분석은 미래에 발생할 수 있는 어떠한 이벤트나 위협 요소를 사전에 탐지하고 이를 해결하기 위한 전략을 분석하는데 사용하는 기법으로 기계 학습, 데이터 마이닝, 모델링, 게임 이론 등의 다양한 방법론이 있다. 과거의 트랜잭션을 나타내는 수치 데이터들이 보통 사용되며 수집된 데이터에서 발견되지 않은 여러 가지 알고리즘, 통계모델, 그리고 패턴 등에서 의미 있는 관계를 찾을 때 사용된다. 이는 빅데이터 시대에 수요가 많은 분석방법 중 하나로, 정확하고 양질의 성과를 달성하기 위해 여러 분석도구와 많은 데이터를 확보하는 것이 중요하며, 금융 서비스에서 개인 신용 평가를 위해 주로 이 분석방법을 사용한다.

지시적 분석은 실제로 어떤 것을 수행하기 전에 그 결정이 어떤 영향을 미칠 것인지에 대하여 예측하려는 분석으로 단지 언제, 어떤 일이 일어

나는지에 대한 예측뿐만 아니라 왜 그것이 일어나는가를 분석하고 이를 통해 어떤 액션을 취해야 하는지에 대한 추천까지 제공할 수 있다. 이 관점에서 본다면, 연구 경쟁력 향상이라는 목표는 연구자와 관련 기술에 대한 역사적이고 변화 추세 등의 수치 정보를 기반으로 분석이 이루어지므로, 수치 데이터를 기반으로 미래의 행동을 예측해야 하는 비즈니스 분석의 관점으로도 해석할 수 있다.

한편, 연구자의 역량을 알기 위해서는 다양한 경로를 통해 자료를 수집할 수 있는데, 기본적으로 연구자에 대한 기본적인 정보와 현재까지 공개된 학술적 성과와 관련된 기술 지식을 추적하는 것이 객관적이고 편리하다^[2]. 연구자에 대한 정보는 여러 검색 엔진에서 연구자의 이름이 들어간 웹 페이지들을, 연구자의 학술적 성과는 NDSL, DBpia와 같은 학술 검색서비스 사이트로부터 수집할 수 있다.

최근에는 단순히 연구자에 대한 학술 정보뿐만 아니라 연구성과도 자동으로 프로파일링하고 수집된 데이터를 기반으로 여러 가지 지표들로 연구자의 역량을 평가하는 서비스들이 생겨났는데, Microsoft Academic Search나 ArnetMiner 등이 그 예이다. 'Microsoft Academic Search(MAS)'는 특정 연구 분야에서 영향력 있는 저자를 기관, 트렌드 등의 여러 기준으로 추천한다. 또한, 공동저자 간의 관계를 그래프로 쉽게 확인할 수 있는데, 예를 들어 서로 다른 저자가 연구 네트워크에서 어떠한 관계에 있는지, 또한 그 사이엔 어떤 다른 연구자들이 있는지 공동저자의 연결도로 발견할 수 있다. 또한 세부 주제 분야별로 연구자의 영향력을 쉽게 알 수 있는 여러 수치정보인 H-index, G-index, 그리고 인용지수 등이 높은 순으로 저자를 정렬할 수 있어 해당 연구 분야의 영향력 있는 연구자를 찾아내기 수월하다. 이와 더불어, 개인

뿐만 아니라 연구기관 간의 경쟁력 비교도 가능하다. H-index 순으로 대학 순위를 보거나 두 기관(대학) 간의 특정 주제 분야에 대한 비교 검색이 가능하여 연도별 논문수와 피인용수와 대표 연구자 리스트 등을 확인할 수 있어 공동저자나 기관을 찾을 수 있다.

한편, 중국 칭화대에서 개발한 ‘ArmetMiner’는 분산되어 있는 과학 기술 정보들로부터 연구자들을 위한 분석 및 데이터 마이닝 결과를 제공한다. 연구자, 및 간행물 간의 연결 관계를 식별하기 위해서 학술 자료들을 분석한다. 전문가 검색, 지리적 검색, 심사위원 추천, 연관 검색, 학업 성적평가 및 주제 모델링을 제공한다. 또한 소셜 네트워크 순위로 사회적 영향력을 보여줌으로써 연구자가 관련 연구자, 논문 등 여러 정보들을 검색하는데 도움을 준다. 하지만 위에서 언급된 MAS나 ArmetMiner에서는 연구자가 검색할 수 있는 환경과 정보만을 제공하며, 연구자의 역량을 강화시키기 위한 조연과 전략정보는 얻을 수 없다.



(그림 1) InSciTe Advanced 예



(그림 2) InSciTe Adaptive 예

3. 지시적 분석 시스템

InSciTe(Intelligence in Science & Technology)는 다양한 시맨틱 웹 기술과 경영분석 기법들을 기반으로 과학기술 문헌자료에서 텍스트를 수집하고 기술이나 개체간의 관계를 추출하는 서비스이다. (그림 1)은 2011년에 개발된 InSciTe Advanced 화면이고 (그림 2)는 2012년에 개발된 InSciTe Adaptive 화면이다. 이 서비스는 2010년에 한국 과학기술정보연구원(KISTI)에서의 개발을 시작으로, 이후 지능형 의사결정 지원 서비스인 InSciTe Advanced로, 그리고 2012년에 트리플 구조의 데이터를 기반으로 연구자 및 경영 의사결정자를 대상으로 의사결정 지원 및 기술기회 발

굴을 가능하게 하는 InSciTe Adaptive로 발전하였다. 이 시스템은 해외 논문, 특허, 웹 데이터, 소셜 데이터, 링크드 데이터 등 다양한 문헌 빅데이터를 수집/처리/통합/분석하여 연구자의 경쟁력을 분석하고 달성 가능한 목표와 전략을 조언하는 빅데이터 기반 지시적 솔루션을 목표로 현재까지 개발되고 있다^[2-4].

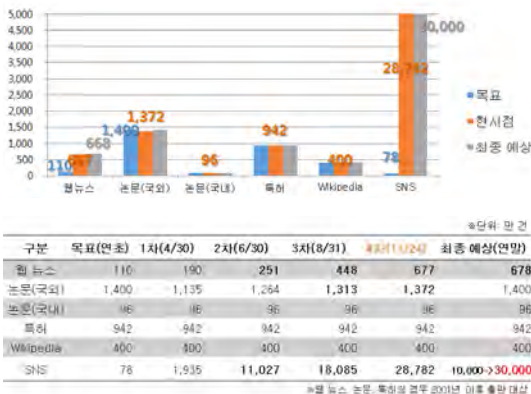
InSciTe가 과학기술의 연구개발 흐름을 분석 및 예측하기 위해서는 기존에 발견된 기술 지식을 전적으로 사용한다. 이 기술은 연구자의 현재 역량을 전략적으로 향상시키기 위해 과거부터 현재까지 달성한 학술적 및 대외적인 성과 등의 연구자 활동 히스토리와 관심분야에서 경쟁력이 있다고 판단되는 역량들을 추적하고 분석을 우선적

으로 진행한다.

연구자 활동 히스토리 정보는 연구자들의 분야별 활동 정보, 연구자의 연구 행동별 활동 정보를 포함하는데, 이는 연구 개발 동향을 분석하고 이와 관련된 통계를 수집하기 위해 사용된다. 연구자 활동 정보는 주로 웹, SNS, 논문, 특허 등에서 수집한다. 수집된 연구자 활동 정보들은 연구자 별로 연구자의 연구영역, 연구 활동 히스토리, 연구 관련 이벤트 등을 조직화해 분석을 수행할 수 있다. InSciTe를 위해 수집된 데이터를 살펴보면, 2014년 11월 24일을 기준으로 웹 뉴스 677만 건, 국외논문 1,372만 건, 국내 논문 96만 건, 특허 942만 건, Wikipedia 400만 건, SNS에서 28,782만 건 등 총 3억 건 이상의 데이터가 수집되었다.

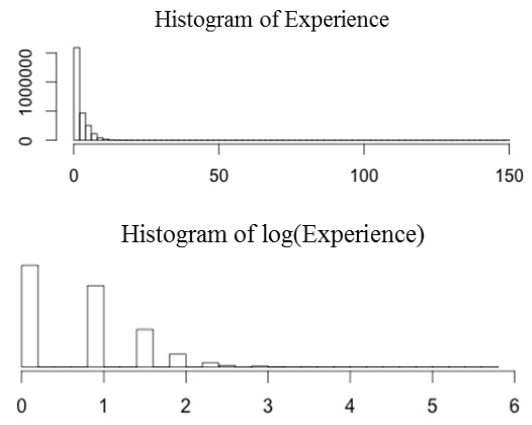
그 다음, 기술 통계 분석을 이용하여 현재 연구자 수준을 진단할 수 있으며 이를 기반으로 향후 연구자가 수행해야 하는 활동들을 예측할 수 있으며 연구자가 특정 수준에 도달하기 위해 필요한 활동 등을 분석하여 제공한다. 또한, 연구자에게 연구자의 롤모델 그룹 또는 멘토를 추천해 줄 수 있으며 R&D 트렌드 기반의 연구자에 대한 지시적 분석을 수행할 수 있다.

〈표 1〉 InSciTe에 사용된 다중 과학기술 텍스트 빅데이터 수집 현황



(그림 3)은 연구자 히스토리 정보 추출 과정에서 식별된 연구자들의 경력 활동에 관한 히스토그램이다. 식별된 연구자는 250여만 명으로 특정 연구자들의 경우 최대 150 여건 이상의 경력을 가지고 있지만 대다수 연구자들은 0 ~ 3건의 경력을 가지는 것으로 나타났다.

경력에 대한 상세 활동의 분포를 기반으로 전체 연구자에 대한 등급을 계산하고 각 연구자별 개별 등급을 도출한다. 도출된 각각의 상세 활동 내역을 종합하여 연구자들의 전체 활동 지수가 도출된다. 한편, 연구자 역량 분석은 연구자의 역량을 내/외부적인 평가 요소를 기반으로 다양한 측면에서 포괄적으로 평가하기 위한 서비스인데, 내부역량 평가 요소 6개, 외부역량 평가 요소 4개를 기준으로 정의한다. 내부 역량 평가 요소는 연구자의 개인 역량과 밀접하게 관련된 요소로 학술성, 상업성, 영향력, 관심도, 다양성, 지속성이 포함되며, 외부 역량 평가 요소는 연구자의 주변 환경과 관련된 요소로 기술의 유망도, 시장 규모 트렌드, 파트너 연구 추세, 수요-공급의 균형도가 포함된다. 각 평가 요소들은 기반 데이터인 논문, 특허, 웹 보고서, 웹 기사, LOD 데이터 등의 메타



(그림 3) 연구자들의 연구 경력 히스토그램과 Log 히스토그램

데이터의 결합으로 구성 및 계산한다.

또한, 연구자 역량 분석 결과를 기반으로 연구 분야에서 선도하는 위치를 차지하기 위한 연구 롤모델을 추천한다. 이는 동일 연구 분야에서 연구자의 상위로 평가되는 연구자들을 그룹별로 분류하여 레벨을 구분 짓고 롤모델을 제시한다. 기본적으로는 연구자의 레벨보다 한 단계 높은 연구자들을 추천하지만 개개인이 스스로 더 높은 레벨의 연구자 그룹을 선택 할 수 있다. 롤모델을 선정하기 위해 연구자 별로 활동을 많이 한 순으로 기술을 도출하고, 도출된 기술들을 공유하는 연구자들을 선별한다. 선별된 연구자들은 연구자 역량 평가 결과를 바탕으로 순위가 정해져 최종적으로 롤모델의 전체 수에 따라 5~20개 그룹으로 구분된다. 이와 같이 선정된 롤모델 구성원들의 리스트와 각 구성원들의 평균 역량 평가치, 그리고 롤모델 연구자들이 공통적으로 가지고 있는 기술들의 목록을 나열한다. 또한, 롤모델 그룹에 속해있는 개별 연구자의 최근 활동 정보를 제공

한다.

마지막으로 이전 단계에서 수행된 모든 역량과 롤모델 추천 결과를 기반으로 최종적으로 연구자에게 지식적 분석을 수행하고 보고서를 생성한다. (그림 4)는 연구자에 대한 5WIH 분석 결과 화면이다. 연구자 지식적 분석 서비스에서 5WIH에 맞춰 정보를 제공한다. 먼저 ‘How-When’에서는 연구자가 롤모델 그룹의 연구자들처럼 역량을 향상시키기 위한 세부 전략과 액션의 시점에 대한 정보를 제공한다. 세부 전략에는 평가 요소별로 논문 투고지, 논문의 양, 심지어는 논문 투고의 시기 등의 지식적 분석 정보들을 제공한다. 즉, 연구자가 롤모델에 속한 연구자들의 수준으로 발전되기 위해 어떤 전략을 세워야 하는지에 대해 구체적인 정보를 제공한다. 다음으로 ‘What to do’에서는 연구자가 롤모델 그룹에 도달하기 위해 수행해야 하는 활동에 대한 전술을 나타낸다. 현재 연구자의 위치가 3등급에 위치함으로 기본적으로 추천되는 롤모델 그룹인 2 등급에 도



(그림 4) 연구자에 대한 5WIH 분석 결과

달하기 위하여 해당 연구자에게 톨모델 그룹에서 주로 연구하는 기술을 추천해 준다. 또한 톨모델 그룹과 연구자의 학술성, 상업성, 다형성 등을 비교하여 해당 항목 중에서 전략적으로 보강해야할 영역을 추천한다. 이를 통해, 연구자는 톨모델 그룹에서 주로 수행하는 기술을 기반으로 연구를 계속하거나 다른 기술 분야를 선택하여 다수의 논문을 출판하기를 추천받는다.

4. 시스템 평가 방안

역동적으로 변화하는 정보 서비스 환경에 능동적으로 대처할 수 있는 기술 개발이 요구되고 있다. 정보를 이용하는 사용자는 신뢰할 수 있는 정보 서비스를 받고 싶은 욕구를 만족시킬 수 있는 체계적인 시스템을 필요로 하게 되었다. 앞서 소개한 지시적 분석 시스템의 결과물이 공신력 있는 데이터로 인정받기 위해서는 개발된 시스템의 비즈니스 또는 공공역역에서의 실제 활용 가능성 높이기 위해 여러 분석 전문 기업과의 공동협력으로 활용성을 점검하고 적극 개선해야 하는 것이 바람직하다. 전문가들과 사용자들의 피드백을 통해 시스템에 사용되고 있는 데이터들 품질 검증도 가능하다.

또한, 최적의 방법을 찾기 위해서 추천 모델의 시뮬레이션을 통해 추천모델 실현 가능성 여부를 확인할 수 있다. 앞서 설명한 지시적 분석 시스템에서는 5W1H에 기반을 두어 사용자에게 조언을 해준다. 분석된 결과인 5W1H 조언의 모든 경우를 시뮬레이션을 진행하면서 추천 방법에 따른 보상들에 대해 점수를 부여하면 추천 모델 중에서도 최적의 방법을 선별할 수 있다. 시뮬레이션 과정에서도 사용자의 피드백을 받아 사용자 유형별로 각 개인에 최적화된 조언을 얻을 수 있다.

5. 결론

본 논문에서는 과학기술 문헌에서 텍스트 마이닝 기법으로 개체 및 관계를 인식하여 기계적으로 생성된 기술 지식을 바탕으로 연구자가 역량 강화를 위해 학술적으로 어떠한 전략 액션을 취해야 하는지 시나리오별 지시적 분석을 수행하는 지능형 정보시스템의 개념과 구조를 알아보았다. 다양한 과학기술 문헌에서 추출된 연구자 정보 및 기술지식 등을 연결해 숨겨진 지식을 발견하는 것은 연구자로 하여금 제한된 환경 내에서 최상의 실적을 낼 수 있도록 연구 행위에 대해 분석 및 제안을 가능하게 해 준다. 향후에는 연구자에게 정보시스템의 신뢰를 줄 수 있도록 기계적으로 발견한 연구자 데이터 및 기술 지식관의 연관성이 현실적으로 얼마나 타당성을 가지는 지 시나리오 추천 기법에 대한 권위성을 확보해야 한다. 또한, 위와 같은 기술 지식 정보체계를 기반으로 경영 의사결정자와 같은 다른 분야에서도 사용자에게 기술 및 비즈니스 역량 강화를 위한 시나리오 추천 기법을 설계할 필요가 있다.

참고 문헌

- [1] IBM. "Descriptive, Predictive, Prescriptive: Transforming Asset and Facilities Management with Analytics", <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=TIW14162USEN>, 26th Dec. 2014.
- [2] 정한민 등. "연구 역량 강화를 위한 지시적 분석." 한국정보과학회 학술발표논문집, pp. 486-488, 2014.
- [3] Sa-kwang Song, Do-Heon Jeong, Jinhyung Kim, Myunggwon Hwang, Jangwon Gim, Hanmin Jung, "Research Advising System

Based on Prescriptive Analytics", In Proceedings of the Future Information Technology, pp. 569-574, 2014.

- [4] Mikyoung Lee, Minhee Cho, Jangwon Gim, Do-Heon Jeong, Hanmin Jung, "Prescriptive Analytics System for Scholar Research Performance Enhancement", In Proceedings of the HCI International 2014-Posters' Extended Abstracts, pp. 186-190, 2014.

홍 승 균

이메일 : xo@kisti.re.kr

- 2014년 계명대학교 Microsoft IT (학사)
- 2014년~현재 과학기술연합대학원대학교 빅데이터과학 (통합)
- 2014년~현재 한국과학기술정보연구원 학생연구원
- 관심분야: 자연어처리, 가상화, 시각화

저 자 약 력

장 윤 지

이메일 : yunji@kisti.re.kr

- 2010년 계명대학교 컴퓨터공학과 (학사)
- 2014년~현재 한국과학기술연합대학원대학교 석사
- 2014년~현재 한국과학기술정보연구원 학생연구원
- 관심분야: LOD, 데이터마이닝, 데이터분석

정 한 민

이메일 : jhm@kisti.re.kr

- 1992년 POSTECH 전자계산학과 (학사)
- 1994년 POSTECH 전자계산학과 (석사)
- 2003년 POSTECH 컴퓨터공학과 (박사)
- 1994년~2000년 ETRI 선임연구원
- 2000년~2004년 다이렉스트 CTO
- 2004년~현재 한국과학기술정보연구원 실장/책임연구원
- 2005년~현재 과학기술연합대학원대학교 교수
- 관심분야: 시맨틱웹, 빅데이터, 자연어처리, HCI