

지시적 분석을 위한 연구자 역량 분석 및 멘토링 모델

김진형*, 황명권**, 김장원***, 송사광****, 정도현*****, 이승우*****, 정한민*****

I. 서론

최근 빅데이터의 등장과 함께 데이터 분석에 대한 기술 및 관련 서비스들이 다시 주목받고 있으며, 이와 관련하여 기존의 비즈니스 인텔리전스에 이어 비즈니스 어널리틱스 또한 각광 받는 분야로 떠오르고 있다. 비즈니스 어널리틱스는 서술적 분석, 예측적 분석, 지시적 분석을 포함하고 있으며, 최근까지 많은 관련 기술이나 서비스들은 서술적 분석 또는 예측적 분석에 집중하여 개발되어 왔다. 하지만 분석에 대한 인사이트의 발견 및 향후 전략 수립은 여전히 사용자의 몫으로 남아 있게 되었다. 따라서 효율적인 빅데이터 분석을 위해서는 서술적 분석 및 예측적 분석 뿐만 아니라 이를 기반으로 사용자에게 다양한 인사이트 및 전략을 제시하고 이에 대한 근거 및 효과까지 제공함으로써 사용자가 향후 전략을 좀 더 쉽게 수립할 수 있게 해주는 지식적 분석이 반드시 필요하다.

학술 분야의 빅데이터, 즉 논문, 특허, 학술 보고서, 관련 웹 기사 등에 대한 분석을 기반으로 하는 기술, 기업, 연구자 분석 서비스도 최근 활발히 개발 및 연구되고 있다. 이 중 특정한 연구 분야에 대한 연구자 및 전문가에 대한 전문적인 역량 분석은 향후 연구 계획 및 전략 수립에 있어 매우 중요한 부분이다. 최근 연구에 의하면 학술 빅데이터 분석 기반의 연구자 역량 분석 서비스를 통해서 특정 연구 분야 연구자의 상대적인 위치를 파악하고 과거 연구 이력 및 연구 성과에 근거하여 향후 적합한 연구 파트너 발굴도 가능하다. 결과적으로 연구 역량 분석 및 평가를 통해서 다음의 질문들에 대한 답변을 획득할 수 있다 [1,2].

- 특정 분야에 얼마나 많은 연구들이 활발하게 진행되고 있는가?
- 특정 연구 분야의 논문과 특허는 증가추세인가 아니면 감소추세인가?
- 특정 연구 분야의 상대적인 영향력은 어느 정도인가?
- 특정 연구 분야의 전문가는 누구이며, 해당 전문가들의 역량은 나에 비해 어느 정도 우수한가?
- 특정 연구 분야에 대한 나의 강점 및 약점은 무엇인가?
- 특정 연구 분야에서 나의 약점을 보완하기 위한 방법은 무엇인가?
- 특정 연구 분야에서 나의 협력 파트너 및 롤모델은 누구인가?

* 김진형, 한국과학기술정보연구원 선임연구원, 042-869-1699, jinhyung@kisti.re.kr

** 황명권, 한국과학기술정보연구원 선임연구원, 042-869-1631, mgh@kisti.re.kr

*** 김장원, 한국과학기술정보연구원 선임연구원, 042-869-1786, jangwon@kisti.re.kr

**** 송사광, 한국과학기술정보연구원 선임연구원, 042-869-0757, esmallj@kisti.re.kr

***** 정도현, 한국과학기술정보연구원 선임연구원, 042-869-1792, heon@kisti.re.kr

***** 이승우, 한국과학기술정보연구원 선임연구원, 042-869-1784, swlee@kisti.re.kr

***** 정한민, 한국과학기술정보연구원 책임연구원, 042-869-1772, jhm@kisti.re.kr

최근까지 많은 컨설팅 및 IT 관련 기업 및 대학들이 연구자 역량 평가와 관련된 서비스 및 도구를 출시하고 이미 일반적으로 활발히 활용되고 있다. 구글이 개발한 Google Scholar [3,4]는 학문적 문헌, 즉 논문과 특허에 기반한 검색 엔진으로 문헌들에 대한 검색과 함께 특정 연구자에 대한 인용정보, h-index 정보를 포함하는 연구자 분석 결과를 제공한다. Elsevier의 Scival Spotlight [5,6]는 연구자의 학술 역량 관리 솔루션으로 과거 학술 액티비티 및 공저자 네트워크에 대한 분석 결과를 제공한다. Microsoft가 개발한 Academic Search [7,8]는 연구자들의 학술 문헌 및 관련 정보 등에 대한 용이한 검색을 지원하는 서비스로 학술 문헌뿐 아니라 관련 연구자, 관련 기관, 관련 주제 및 관련 액티비티에 대한 정보를 포괄적으로 제공한다. 마지막으로 Thomson이 개발한 ResearchID는 학술 데이터를 기반으로 한 학술 정보 제공 서비스로 추세 정보와 협력자 네트워크 정보를 제공한다.

위의 기술한 바와 같이 연구자 역량 분석과 관련된 다양한 서비스 및 도구가 활용되고 있지만, 기존의 서비스 및 도구들은 모두 학술 정보를 단순히 검색하고 보여주는 데 집중하고 있으며, 연구자의 학술 정보 분석 수준에 있어서도 여전히 서술적 분석 또는 예측적 분석 수준에 머무르고 있다. 또한 연구자의 역량을 분석하는 위한 평가 요소는 매우 단순하여, 단순히 연구자가 몇 편의 논문을 썼는지, 공저자가 누구인지 등의 일차적인 정보 외에는 사용자가 얻기가 힘들다. 따라서 사용자가 위와 같은 다양한 분석 서비스를 활용한다고 하더라도 향후 연구 계획을 수립하거나 연구 분야나 관련 연구자에 대한 인사이트를 획득하는 것은 매우 어려운 일이다.

따라서 본 논문에서는 다양한 평가요소를 기반으로 연구자의 역량을 다각적인 관점에서 포괄적으로 분석하고 이를 기반으로 연구자가 어떤 연구자를 롤모델로 삼아야 하는지를 지시적 분석 관점에서 제공하는 새로운 연구자 역량 평가 및 롤모델 추천 모델을 제시한다. 연구자 역량 분석 서비스에서 지시적 분석이란 단순히 연구자의 과거 및 현재 역량, 즉 논문이나 특허를 얼마나 출간 및 출원했고, 연도별 추세가 어떻게 되는지, 그리고 지금의 추세라면 향후에는 어떤 추세를 보일지를 단순히 제시하는 수준이 아니라 연구자가 자신의 역량을 개선하기 위한 연관 기술 분야의 롤모델을 추천하고 이러한 롤모델을 분석해 줌으로써 자신이 개선할 분야 및 방법을 찾을 수 있도록 하는 분석의 최상위 단계이다.

II. 관련연구

1. 기존 관련 서비스

1) 구글 Scholar

구글 scholar는 구글에서 2004년 개발한 학술적 문헌 기반의 검색 및 연구자 역량 분석을 위한 서비스이다. 구글 Scholar는 미국과 유럽의 저명한 출판사들의 학술적 문헌 뿐 아니라, 학술적 서적, 그리고 다양한 학술적 보고서들까지 포함하고 있다. 구글 scholar에는 구글의 핵심 기술인 검색 및 인덱스 엔진이 적용되어 다양한 관련 학술 문헌을 빠르고 쉽게 검색할 수 있을 뿐 아니라 이러한 기반 정보를 바탕으로 연구자들의 연구 트렌드 및 연구자들 간의 협력 관계에 대한 정보도 제공한다. 사용자는 키워드를 기반으로 검색이 가능하며 논

문 내 키워드가 포함되는 논문 뿐 아니라 자체 연관성 분석 엔진을 통하여 관련 학술 문헌들도 같이 제공한다. 사용자는 연관성, 인용도, 날짜 등의 다양한 요소들에 의해 결과를 재배열 할 수 있으며, 이를 통해 자신이 원하는 정보를 좀 더 정확히, 그리고 쉽게 검색 할 수 있다. 또한 다양한 학술 문헌 정보를 기반의 인용 정보와 h-index 정보를 제공하며, 이를 기반으로 각종 학술 문헌에 가중치를 부여함으로써 영향력 및 인용도가 높은 학술 문헌들을 쉽게 검색할 수 있다.



그림. 1. 구글 Scholar 서비스

2) Elsevier Scival

SciVal은 Elsevier에 개발한 연구자 및 기관, 조직 역량을 분석하기 위한 패키지 솔루션으로 기관, 조직의 의사 결정에 대한 효율성 및 효과를 극대화 시키는 것을 목적으로 한다. SciVal은 7개의 세부 서비스로 구성되며, 각 서비스들은 내부적으로 상호 연결되어 있다. 7개의 다양한 서비스들 중 SciVal Expert, SciVal Spotlight, 그리고 SciVal Strara가 연구 역량 평가와 관련이 있다.

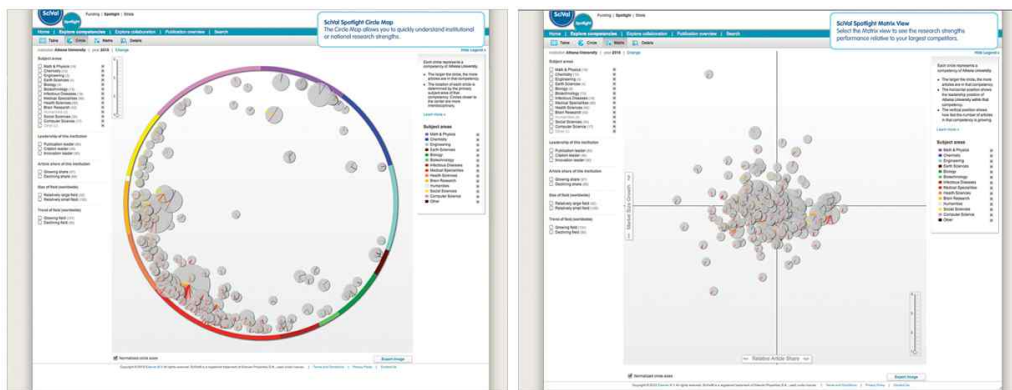


그림 2. Elsevier Scival 서비스

SciVal Expert는 연구자의 개인 프로파일 정보를 기반으로 사용자의 협력 네트워크를 분석하고 이를 기반으로 특정 분야의 협력 가능한 연구자에 대한 정보를 제공한다. 또한 SCOPUS의 학술 문헌들을 기반으로 학술 역량 및 인용 정보를 분석함으로써 특정 기관 내

의 연구자들의 역량을 평가한다. SciVal Spotlight는 연구자들의 개인 프로파일 정보 및 학술 역량을 바탕으로 기관, 대학, 국가에 대한 강점 및 약점을 분석하고 이를 시각화한다. 또한 인용 정보 및 협력 정보를 분석함으로써 향후 유망한 연구분야에 대한 정보도 제공한다. SciVal Strata는 연구자 개인 또는 팀의 역량 학술 문헌 및 인용 정보에 기반하여 분석하고 이를 통해 최적의 팀 구성에 대한 시뮬레이션을 가능하게 해준다.

3) Academic Analytics Portal

Academic Analytics Portal은 영국 및 미국 대학 관리자들을 위해 개발된 역량 분석 서비스로써 각 대학들의 학술역량을 개개인이 학술적 역량에 기반하여 분석하고 이를 통해 각 관리자들에 대학의 강점 및 약점에 대한 정보를 제공한다. 또한 이러한 정보를 기반으로 대학 간의 협력 가능 여부를 분석하고 약점을 보완할 수 있는 방법에 대한 정보도 제공한다. Academic Analytics Portal의 특성은 기반 정보들을 다양한 관점 (비용 관점, 논문 출판 관점 등)에서 분석함으로써 포괄적으로 대학의 강점, 약점을 분석하고 이를 'Flower Chart'에 표현한다.

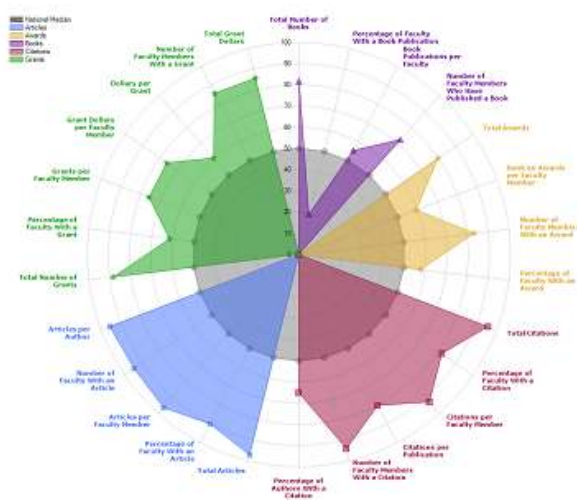


그림 3. Academic Analytics Portal 서비스

4) MS Academic Search

Academic Search는 MS에서 개발한 연구정보 검색 서비스로 연구자, 학생들의 학술 문헌, 연구자, 기관, 액티비티에 대한 검색 서비스를 제공한다. Academic search는 수백만건의 학술 문헌을 다룰 뿐 아니라 연구 주제, 연구자, 저자 등의 상호 연관 네트워크 정보도 제공한다. Academic Search는 5개의 주요 서비스 기능을 가지며, 다음과 같다.

- 공저자 그래프 및 경로 분석: 특정한 연구자의 공저자 정보를 분석하고 두 연구자 간의 공저자 네트워크 구성
- 인용 그래프 분석: 학술 문헌 간 인용 정보를 분석하고 특정 문헌과 다른 문헌과의 인용 네트워크 구성

- 학술 역량 트렌드 분석: 학술 분야에 대한 문헌 출판, 액티비티 정보를 분석하여 일정 기간 내의 학술 역량 트렌드를 분석
- 학술 맵 구성: 학술 정보를 지역적으로 분석하여 빈도수를 기반으로 지도에 색과 크기로 정보를 표현
- 기관 비교 분석: 기관 내 학술 문헌에 대한 인용 정보, 키워드, 상위 수준의 저자 정보를 기반으로 특정 기술 분야에 대한 기관의 역량 분석하고 각 기관의 공통점과 차이점을 비교 분석



그림 4. MS Academic Search 서비스

2. 기존 서비스의 이슈

위의 2.1장에서 기술한 것처럼 연구자 역량 분석과 관련된 서비스는 다양하게 개발되고 있으며, 이외에도 ArnetMiner [9,10], Cyrus [11], ResearchGate [12] 등이 활발히 사용되고 있다. 하지만 기존의 연구자 역량 분석 서비스들은 다음과 같은 한계점을 지니고 있다.

- 빈약한 학술 정보 리소스: 대부분의 연구자 역량 분석 서비스들은 연구자 및 기관, 국가의 역량을 분석하기 위해 논문 정보 또는 일부 소수의 특허 정보 분석에 집중하고 있다. 하지만 정확한 연구자 역량의 분석, 기술의 트렌드 분석, 연구자 협력 정보 분석을 위해서는 관련 정보는 다양하게 필요하다. 따라서 본 논문에서 제안하는 서비스에서는 유수의 출판사 논문 정보, 미국, 일본, EU의 특허 정보, 웹 뉴스, 웹 리포트, 그리고 SNS 데이터 일부도 활용한다.
- 단순한 역량 평가 기준: 기존 대부분의 연구자 역량 분석 서비스들은 논문의 출판 건수 및 일부 인용 정보만을 기준으로 연구자의 역량을 평가한다. 즉, 단순히 논문 출판 건수가 많고 인용 횟수가 많을 수록 역량이 우수한 것으로 평가한다. 하지만 이러한 단순한 역량 평가기준으로는 연구자 각각의 역량을 포괄적으로 세세히 평가 불가하다. 따라서 본 논문에서 제안하는 서비스에서는 10개의 평가 기준 (학술성, 상업성, 영향력, 다양성, 지속성, 기술 유망도, 파트너 트렌드, 시장 트렌드, 수요공급 균형도) 으로 정의하고 이를 통해 연구자를 포괄적으로 분석한다.
- 외부 역량 평가 요소의 부재: 기존 대부분의 연구자 역량 분석 서비스들은 연구자 자체의 역량, 즉 논문 출판 건수, 특허 출원 횟수 등을 기준으로 연구자 역량을 분석하고 있다. 하지만 연구자 자체의 내부 역량 외에도 외부의 요소들도 연구자의 실제 역량에 영

향을 미치는 요소들이 많다. 예를 들어 두 연구자가 동일한 논문 출판 횟수와 인용 횟수를 가진다고 가정할 때, 기존의 서비스들은 두 연구자의 역량을 동일한 수준으로 분석한다. 하지만 한 연구자는 '2G Network'와 같은 이미 쇠퇴기에 접어든 기술을 주로 연구한 반면, 다른 연구자는 'Big Data Management'와 같이 최신 유망 기술을 연구하는 경우 이러한 두 연구자의 역량을 동일하다고는 볼 수 없다. 따라서 이러한 외부 요소, 즉 연구자 자체의 능력과는 연관성이 적은 요소들도 연구자 역량 평가에 반드시 반영을 해야 좀 더 정확한 연구자 역량 평가를 할 수 있다.

III. 연구자 역량 평가 모델 및 롤모델 추천

1. 연구자 역량 평가 모델

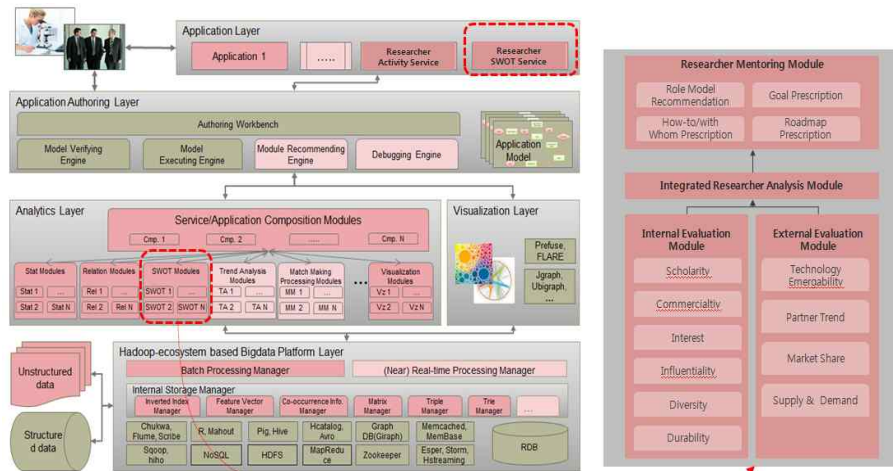


그림 5. MS Academic Search 서비스

그림 5는 연구자 역량 평가 모델의 전체적인 아키텍처를 보여준다. 연구자 역량 평가를 위해서 본 모델에서는 논문, 특허, LOD 데이터와 같은 다양한 구조화된 정보와 함께 크롤링을 통해 비구조화된 정보인 다양한 웹 기사, 웹 리플, SNS 데이터들도 활용한다. 또한 이러한 데이터를 바탕으로 보다 정확하고 의미 있는 정보의 검색을 지원하기 위해 개체들 간의 관계정보를 온톨로지로 구축한다. 이러한 모든 정보들은 매우 다양한 구조와 엄청난 용량을 가지기 때문에 그림 5의 가장 하부 구조인 하둡 에코 시스템 기반의 빅데이터 플랫폼 계층에서 저장 및 관리된다. 그리고 분석 계층에서는 다양한 연구자 역량 평가 서비스를 위한 핵심 분석 모듈들이 위치하고 있다. 본 논문에서는 핵심 분석 모듈들 중 연구자 역량을 분석하는 데 핵심인 연구자 SWOT 분석 모듈에 대해 기술한다.

SWOT 분석 모듈은 크게 내부역량 평가 모듈, 외부역량 평가 모듈, 통합 연구자 분석 모듈, 그리고 연구자 멘토링 모듈로 구성되어 있다. 내부역량 평가 모듈은 연구자의 개인적 역량과 밀접하게 관련되어 있는 논문, 특허와 그와 관련된 역량을 분석하는 모듈이다. 연구자 내부역량 평가를 위해서 본 모델에서는 크게 6가지 평가 요소를 정의한다.

<i>Acronym</i>	<i>Description</i>
$CIT_t^K(PR_n)$	Average number of patent citation of person PR_n , related K topic at t year
$PTA_t^K(PR_n)$	Number of applied patent of person PR_n , related K topic at t year
$PTR_t^K(PR_n)$	Number of registered patent of person PR_n , related K topic at t year
$CPP_t^K(PR_n)$	Patent CPP index of person PR_n , related K topic at t year
$HID_t^K(PR_n)$	H-index of person PR_n , related K topic at t year
$GID_t^K(PR_n)$	G-index of person PR_n , related K topic at t year
$PPC_t^K(PR_n)$	Number of proceeding papers of person PR_n , related K topic at t year
$PPJ_t^K(PR_n)$	Number of Journal papers of person PR_n , related K topic at t year
$PAR_t^K(PR_n)$	Number of positive articles in social data of person PR_n , related K topic at t year
$NAR_t^K(PR_n)$	Number of negative articles in social data of person PR_n , related K topic at t year
$EDC_t^K(PR_n)$	Number of conference editors Number of positive articles in social data of person PR_n , related K topic at t year
$EDJ_t^K(PR_n)$	Number of journal editors Number of positive articles in social data of person PR_n , related K topic at t year
$WN_t^K(PR_n)$	Number of web articles of person PR_n , related K topic at t year
$WR_t^K(PR_n)$	Number of web reports of person PR_n , related K topic at t year
$PC_t^K(PR_n)$	Number of paper citation of person PR_n , related K topic at t year
$PCoA_t^K(PR_n)$	Number of co-authors of person PR_n , related K topic at t year
$PCA_t^K(PR_n)$	Number of correspondence authors of person PR_n , related K topic at t year
$PTCoA_t^K(PR_n)$	Number of co-developer of person PR_n , related K topic at t year
$PCaT_t^K(PR_n)$	Number of paper categories of person PR_n , related K topic at t year
$PK_t^K(PR_n)$	Number of paper keywords of person PR_n , related K topic at t year
$PTIPC_t^K(PR_n)$	Number of IPC class of person PR_n , related K topic at t year
$WCaT_t^K(PR_n)$	Number of web article categories of person PR_n , related K topic at t year
$WK_t^K(PR_n)$	Number of keywords extracted from SNS articles of person PR_n , related K topic at t year

표 1. 함수 및 용어표

- 학술성: 연구자의 학술적인 역량을 평가하기 위한 요소로서 연구자가 출판한 학회/저널 논문, 인용지수, 학회/저널의 우수성 등이 학술성에 영향을 미치는 가장 주요한 요소이다. 즉, 유수의 저널에 인용지수가 높은 논문 출판 횟수가 최근에 많은 경우 학술성은 높게 평가된다.

$$\begin{aligned}
 \text{Scholarity}_t^K(PR_1) &= \frac{HID_t^K(PR_1) * \{PPC_t^K(PR_1) + PPJ_t^K(PR_1)\}}{\sum_{i=1}^n \{PPC_t^{K_i}(PR_1) + PPJ_t^{K_i}(PR_1)\}} \\
 &= \frac{HID_t^K(PR_1) * PPC_t^K(PR_1) + HID_t^K(PR_1) * PPJ_t^K(PR_1)}{\sum_{i=1}^n \{PPC_t^{K_i}(PR_1) + PPJ_t^{K_i}(PR_1)\}}
 \end{aligned}$$

- 상업성: 연구자의 상업적/실무적 역량을 평가하기 위한 요소로서, 연구자가 출원/등록한 특허, 특허의 CPP지수 등이 가장 주요한 요소이다. 따라서 인용이 많이 되는 특허를 최근에 많이 등록한 경우 상업성은 높게 평가된다.

$$\begin{aligned}
 \text{Commerciality}_i^K(PR_1) &= \frac{CIT_i^K(PR_1) * \{PTA_i^K(PR_1) + PTR_i^K(PR_1)\}}{PTR_i^K(PR_1) * \sum_{l=1}^n \{PTA_l^{K_i}(PR_1) + PTR_l^{K_i}(PR_1)\}} \\
 &= \frac{\frac{CIT_i^K(PR_1) * PTA_i^K(PR_1)}{PTR_i^K(PR_1)} + \frac{CIT_i^K(PR_1) * PTR_i^K(PR_1)}{PTR_i^K(PR_1)}}{\sum_{l=1}^n \{PTA_l^{K_i}(PR_1) + PTR_l^{K_i}(PR_1)\}} \\
 &= \frac{\frac{CIT_i^K(PR_1) * PTA_i^K(PR_1)}{PTR_i^K(PR_1)} + CIT_i^K(PR_1)}{\sum_{l=1}^n \{PTA_l^{K_i}(PR_1) + PTR_l^{K_i}(PR_1)\}} \\
 &= \frac{CIT_i^K(PR_1) * \{1 + \frac{PTA_i^K(PR_1)}{PTR_i^K(PR_1)}\}}{\sum_{l=1}^n \{PTA_l^{K_i}(PR_1) + PTR_l^{K_i}(PR_1)\}}
 \end{aligned}$$

- 영향력: 연구자가 타 연구자에게 미치는 과급력 또는 다른 연구자의 관심도를 평가하기 위한 요소로서, 연구자와 관련된 인용도, 공저자, 교신저자, 공동개발, CPP지수 등이 가장 주요한 요소이다. 따라서 연구자의 논문이나 특허의 공저자가 많고, 인용횟수가 높은 경우 영향력은 높게 평가된다.

$$\begin{aligned}
 \text{Influentiality}_i^K(PR_1) &= \frac{\sum_{t=2001}^{2013} \{PC_i^K(PR_1) + PCoA_i^K(PR_1) + PCA_i^K(PR_1) + CPP_i^K(PR_1) + PTCOA_i^K(PR_1)\}}{\sum_{l=1}^n \{ \sum_{t=2001}^{2013} PC_l^K(PR_1) + CPP_l^K(PR_1) + 3l \}}
 \end{aligned}$$

- 다양성: 연구자가 얼마나 다양한 분야에 기여했는지를 평가하기 위한 요소로서, 논문 카테고리, 논문 키워드, 특허 IPC, 문헌별 추출 기술 정보 등이 가장 주요한 요소이다. 따라서 연구자가 다양한 연구 분야에서 논문이나 특허를 출판 또는 출원 했을 경우 다양성은 높게 평가된다.

$$\begin{aligned}
 \text{Diversity}_i^K(PR_1) &= \frac{\sum_{t=2001}^{2013} \{PCaI_t^K(PR_1) + PK_t^K(PR_1) + PTIP_t^K(PR_1) + WCaI_t^K(PR_1) + SK_t^K(PR_1)\}}{\sum_{l=1}^n [\sum_{t=2001}^{2013} \{PCaI_t^K(PR_1) + PK_t^K(PR_1) + PTIP_t^K(PR_1) + WCaI_t^K(PR_1) + SK_t^K(PR_1)\}]}
 \end{aligned}$$

- 지속성: 연구자가 얼마나 특정 연구분야에 지속적으로 기여했는지를 평가하기 위한 요소로서 다양성과는 동일한 요소들이 주를 이루나 의미적으로는 정반대의 평가 요소이다. 따라서 연구자가 특정 분야에서 오랜 기간동안 논문이나 특허를 출판 또는 출원 했을 경우 지속성을 높게 평가된다.

$$\begin{aligned}
 & \text{if}\{(PTA_t^K(PR_1)+PTR_t^K(PR_1))=0\} \\
 & DPT_t^K = 0 / \text{else} / DPT_t^K = 1 \\
 & \text{if}\{(PPC_t^K(PR_1)+PPJ_t^K(PR_1))=0\} \\
 & DPT_t^K = 0 / \text{else} / DPT_t^K = 1 \\
 & \text{if}\{(WN_t^K(PR_1)+WR_t^K(PR_1))=0\} \\
 & DW_t^K = 0 / \text{else} / DW_t^K = 1 \\
 \\
 & Durability_t^K(PR_1) = \frac{\sum_{t=2001}^{2013} (DPT_t^K + DPP_t^K + DW_t^K)}{39}
 \end{aligned}$$

외부역량 평가 모듈은 연구자와 직접 관련은 없지만 역량에 영향을 미칠 수 있는 외부적인 환경에 대한 평가를 수행하는 모듈이다. 예를 들어, 두 연구자가 동일한 학술적인 역량을 가지고 있지만, 한 연구자는 매우 유망한 기술에 집중하는 반면, 다른 연구자는 오래된 진부한 기술을 연구하는 경우가 발생 가능하다. 하지만 두 연구자의 학술적 역량이 동일하다고 해서 전체적인 역량이 동일하다고 결론지을 수는 없다. 이러한 경우 외부적인 역량을 분석하여 두 연구자 중 어떤 연구자가 전체적인 역량 측면에서 더 우수한지 구분이 가능하다. 연구자의 외부역량 평가를 위해서 본 모델에서는 크게 4가지 평가 요소를 정의한다.

- 기술 유망도: 기술의 유망 정도를 상대적으로 파악하기 위한 요소로서 특정 기술에 대한 논문, 특허의 증감추세, 특정 기술의 발전 정도, 특정 기술 분야에 대한 학회/저널의 활성화도 등이 가장 주요한 요소이다. 기술 유망도를 파악하기 위해서 신규유망기술, 연관유망기술, 성숙유망기술, 참조유망기술을 정의하고 각각의 기준에 따라 기술의 유망도를 평가한다.

- ✓ 신규유망기술: 최근 2~3년내에 등장하여 타 기술들에 비해 빠른 성장세를 보이고 있는 기술을 신규유망기술로 정의한다.

$$\begin{aligned}
 & \sum_{t=t_{cur}-2}^{t_{cur}} \left\{ \sum_{n=1}^m (PTA_t^K(PR_n) + PTR_t^K(PR_n) + \right. \\
 & \left. PPJ_t^K(PR_n) + PPC_t^K(PR_n)) \right\} \\
 & > \sum_{t=t_{ini}}^{t_{cur}-3} \left\{ \sum_{n=1}^m (PTA_t^K(PR_n) + PTR_t^K(PR_n) + \right. \\
 & \left. PPJ_t^K(PR_n) + PPC_t^K(PR_n)) \right\} \\
 & \sum_{t=t_{cur}-2}^{t_{cur}} \left\{ RGR_t^K(PTA) + RGR_t^K(PTR) \right. \\
 & \left. RGR_t^K(PPJ) + RGR_t^K(PPC) \right\} \\
 & > \sum_{t=t_{ini}}^{t_{cur}-3} \left\{ RGR_t^K(PTA) + RGR_t^K(PTR) \right. \\
 & \left. RGR_t^K(PPJ) + RGR_t^K(PPC) \right\}
 \end{aligned}$$

- ✓ 연관유망기술: 최근 활발히 연구되고 있는 최고 기술들의 요소 기술들 중 증가세를 보이는 기술들을 연관유망기술로 정의한다.

$$\begin{aligned}
 ET(K), & \text{ iff } \sum_{n=1}^{\infty} \{PPJ_t^K(PR_n) + PPC_t^K(PR_n)\} \\
 & \geq \forall_{K=Tech} \{PPJ_t^K(PR_n) + PPC_t^K(PR_n)\} \\
 ET(K), & \text{ iff } \sum_{n=1}^m \{PTA_t^K(PR_n) + PTR_t^K(PR_n)\} \\
 & \geq \forall_{K=Tech} \{PTA_t^K(PR_n) + PTR_t^K(PR_n)\}
 \end{aligned}$$

- ✓ 성숙유망기술: 이미 거의 성숙단계에 이르러 조만간 안정화 단계에 이를 수 있는 기술들을 성숙유망기술로 정의한다.

$$\begin{aligned}
 & \sum_{t=t_{cur}-2}^{t_{cur}} \{TDS_t^K(PTA) + TDS_t^K(PTR)\} \\
 & TDS_t^K(PPJ) + TDS_t^K(PPC) \\
 & > \sum_{t=t_{pre}}^{t_{cur}-3} \{TDS_t^K(PTA) + TDS_t^K(PTR)\} \\
 & TDS_t^K(PPJ) + TDS_t^K(PPC)
 \end{aligned}$$

- ✓ 연관유망기술: 최근 활발히 연구되고 있는 최고 기술들의 요소 기술들 중 증가세를 보이는 기술들을 연관유망기술로 정의한다

$$\begin{aligned}
 CT(K), & \text{ iff } \sum_{n=1} \{PPJ_t^K(PR_n) + PPC_t^K(PR_n)\} \\
 & \geq \forall_{K=Tech} \{PPJ_t^K(PR_n) + PPC_t^K(PR_n)\} \\
 CT(K), & \text{ iff } \sum_{n=1}^m \{PTA_t^K(PR_n) + PTR_t^K(PR_n)\} \\
 & \geq \forall_{K=Tech} \{PTA_t^K(PR_n) + PTR_t^K(PR_n)\}
 \end{aligned}$$

- 파트너 연구추세: 공저자들의 연구추세를 파악하여 특정 기술 분야에 대한 관심도를 파악하기 위한 요소로서, 기존 동일 분야의 공저자들이 여전히 동일 분야에 집중하는지, 아니면 다른 분야로 연구 분야를 변경했는지에 따라 기술 분야의 관심도 평가 가능하다.
- 기술분야 마켓 추세: 기술 분야의 시장에 대한 투자 증감을 분석함으로써 특정 기술 분야에 대한 관심도 및 집중도를 파악하기 위한 요소로서, 특정 기술 분야에 대한 투자증감 추세, 참여기업의 증감 추세 등이 가장 주요한 요소이다.
- 수요 및 공급 추세: 기술 분야의 수요 및 공급의 균형 정도를 분석함으로써 특정 기술 분야에 대한 관심도를 파악하기 위한 요소로서, 특정 기술 분야에 대한 관련 제품의 증감 추세, 제품에 대한 아류 제품들의 증감 추세 등이 가장 주요한 요소이다. 수요, 공급에 대한 정보를 분석하기 위해서 본 논문에서 제안하는 서비스 모델에서는 그림 6과 같은 온톨로지 정보를 활용한다. 온톨로지 내에 정의되어 있는 기술 개체와 제품 개체간의 관계를 활용하여 해당 관계 정보의 증가, 감소 추세를 분석하여 수요, 공급 트렌드 분석한다. 기술과 제품의 관계를 분석하여 특정 기술에 대한 연관 제품이 증가하면 이는 공급이 증가하는 것으로 파악한다. 반대로 제품과 제품의 관계를 분석하여 특정 제품과 관련된 제품이 증가하면 이는 수요의 증가로 인해 기존 제품이 시장을 요구를 충분히 소화

하지 못하고 있기 때문에 아류 제품들이 빠르게 생산되고 있는 것으로 파악하여 수요가 증가하는 것으로 파악한다.

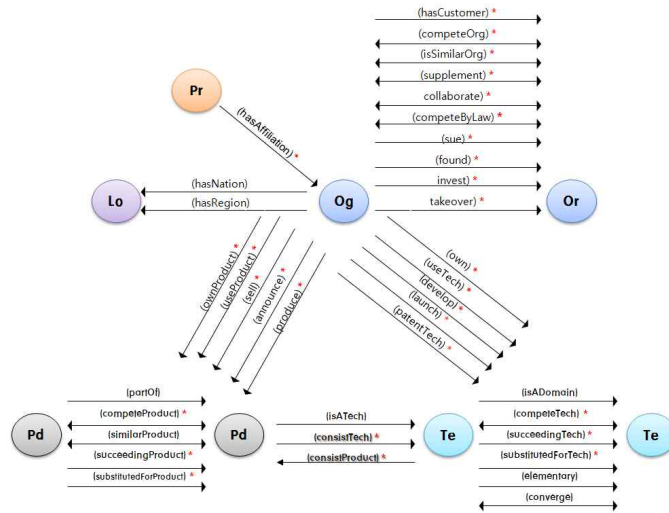


그림 6. 온톨로지 스키마

위에 기술한 6개의 내부역량 평가 요소와 4개의 외부역량 평가 요소를 기반으로 연구자의 역량을 포괄적으로 분석이 가능하다. 그림 7는 10개의 평가요소를 기반으로 연구자의 역량을 분석한 연구자 SWOT 분석 서비스를 보여준다.

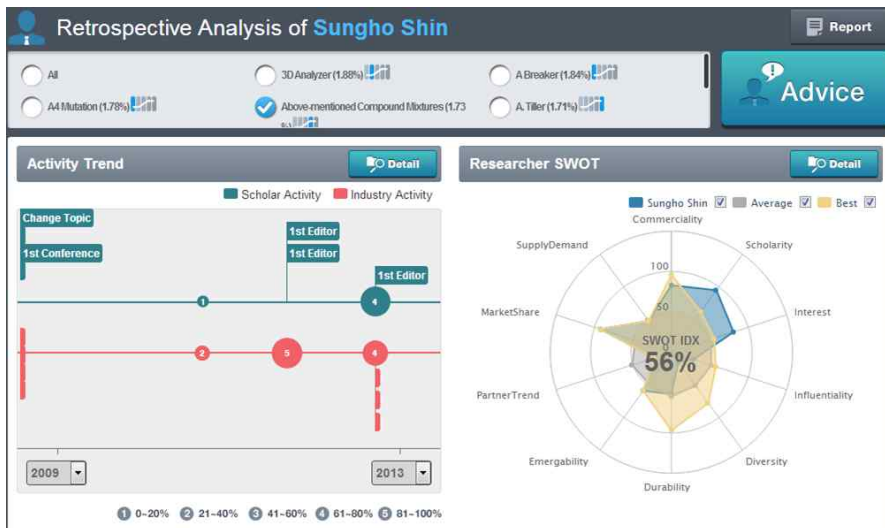


그림 7. 연구자 역량 분석 서비스

2. 연구자 롤모델 추천

롤모델 추천은 2장에서 기술한 10개의 평가요소에 의해 연구자의 역량을 포괄적으로 분석하고 연구자의 중점 연구 분야를 파악한다. 이를 기반으로 연구자의 중점 분야와 관련된 유망한 기술들을 우선 추천하고 해당 유망 기술 분야의 롤모델 연구자를 추천한다. 기술 추천

을 위해서 우선 연구자의 핵심 연구분야와 연관된 기술 정보와 기술 간 공기 정보를 활용한다. 즉, 논문과 특허, 웹 기사 등을 기반으로 구축된 온톨로지에서 관계를 가지는 연관 기술과 논문의 키워드, 논문의 카테고리, 특허 IPC 정보 등을 바탕으로 공통적으로 출현이 빈번한 기술들에 대한 정보를 기반으로 해당 기술들에 대한 유망도를 판단한다. 기술 유망도의 판단 방법은 2장에서 기술한 외부역량 요소 중 하나인 기술 유망도 요소와 동일하다. 유망 기술이 추천되면 해당 기술 내의 관련 연구자들을 10개의 역량 요소에 근거하여 역량을 포괄적으로 평가하고 연구자들 중 20개~100개의 그룹으로 분류한다. 여기서 무조건 연구자에게 최상위 그룹의 연구자를 롤모델로 추천하는 것이 아니라, 연구자의 현재 역량 및 역량의 발전속도와 롤모델의 역량 발전속도를 고려해 보았을 때 3~5년의 단기간 내에 개선이 가능한 수준의 연구자를 롤모델로 추천한다.

T_s: starting point of role model

T_C: current point, T_N: n-year future point

Y_M: current capability of me

Y_R: current capability of role model

Y_{R'}: n-year future capability of role model

<Role Model>

$$y = \{Y_R / (T_C - T_S)\}x = \{(Y_{R'} - Y_R) / (T_N - T_C)\}x + Y_{R'}$$

$$= \{Y_R(T_S - T_N) + Y_{R'}(T_C - T_S) / (T_N - T_C)(T_C - T_S)\}x + Y_{R'}$$

<Mine>

$$y = \{(Y_R - Y_M) / (T_N - T_C)\}x + Y_M$$

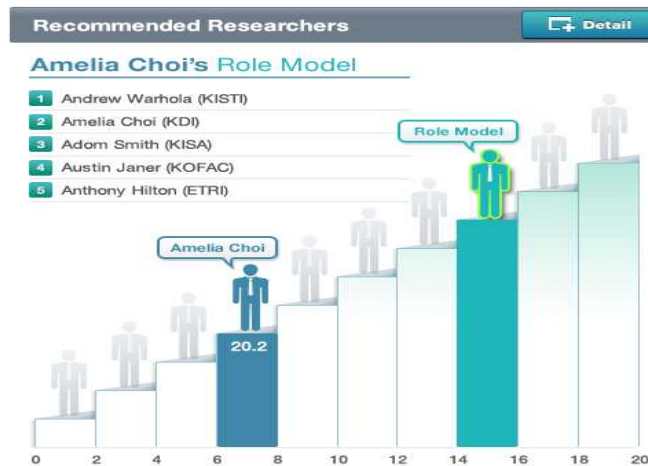


그림 8. 연구자 롤모델 추천 서비스

IV. 결론

본 논문에서는 연구자 역량을 포괄적으로 상세히 분석하기 위한 새로운 연구자 역량 분석 모델을 제시하고 이를 위한 10개의 평가 요소를 정의하였다. 또한 연구자의 분석뿐 아니라 평가에 기반하여 연구자에게 가장 적합한 연구자를 추천해주는 연구자 롤모델 추천 모듈도 제시하였다. 제시하는 모델은 다각화된 평가요소를 통해 기존 서비스들에 비해 좀 더 정확

하고 포괄적인 연구자 역량 평가 결과를 제시하며 룰모델 추천을 통해 지식적 분석 단계 수준에 이르고 있다.

[참고문헌]

- [1] Leitner, Karl-Heinz (2005), “Managing and Reporting Intangibles Assets in Research Technology Organizations”, *R&D Management*, 35(2): 125-136.
- [2] P. Goyal, Z. hahman, A. Kazmi (2013), “Corporate sustainability performance and firm performance research: Literature review and future research agenda,” *Management Decision*, 51(2): 361-379.
- [3] S. Wolter (2013), “The Use of Bibliometrics to Measure Research Performance in Education Sciences,” *Research in Higher Education*, 54(1): 86-114.
- [4] J. Beel, B. Gipp (2009), “Google Scholar’s Ranking Algorithm: An Introductory Overview,” *The proceeding of Scientometrics and Informetrics*, 1: 230-241.
- [5] J. Beel, B. Gipp (2009), “Google Scholar’s Ranking Algorithm: The Impact of Citation Counts,” *The proceeding of Research Challenges in Information Science*, 439-446.
- [6] E. Vardell, B. Feddern, M. Moore (2011), “SciVal Experts: A Collaborative Tool,” *Medical Reference Services Quarterly*, 30(3): 283-294.
- [7] Microsoft (2011). Academic Categories in Microsoft Academic Search. <http://social.microsoft.com/Forums/en-US/mas/thread/bf20d54a-ed2-48a9-8bbb-f6c1c1f30429>
- [8] Microsoft (2012), Microsoft Academic Search Data Coverage <http://social.microsoft.com/Forums/en-US/mas/thread/7ed5d49d-65b7-4a2f-9adf-4de9e23ee70e>
- [9] J. Li, J. Tang, J. Zhang, L. Hong, Q. Luo, Y. Liu (2008), “Arnetminer: Expertise Oriented Search using Social Networks,” *Frontiers of Computer Science in China*, 2(1): 94-105.
- [10] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, Z. Su (2008), “ArnetMiner: Extraction and Mining of Academic Social Networks,” *Proceeding of International Conference on Knowledge Discovery and Data Mining*, 990-998.
- [11] N. Honarmand, N. Dautenhahn, J. Torrellas, S. king, G. Pokam, C. Pereira (2013), “Cyrus: Unintrusive application-level record-replay for replay parallelism,” *Proceeding of International Conference on Architectural Support for Programming Languages and Operating Systems*, 193-206.
- [12] ResearchGate (2013), “<http://www.researchgate.net/>”, 2013.