

과학기술 이슈 발굴을 위한 키워드 경로 탐색 모델 개발 관련 연구

심위* · 권오진** · 김근환*** · 정승호**** · 김도현*****

I. 서론

전문가 기반의 과학기술 이슈 발굴은 전문가의 판단에의 의존성을 보이며, 이는 곧 전문가의 주관이 과도하게 개입될 수 있다는 가능성을 내포하고 있다. 따라서 비편향적 과학기술 이슈 발굴을 위한 학술 및 특허 데이터베이스 기반의 주기적 모니터링 프로세스가 요구된다.

일반적 과학기술 이슈 탐색을 위한 프로세스에는 일부 키워드 분석이 수행되고 있으나, 키워드의 심층적인 분석보다는 빈도의 변화를 파악하는 단순 분석만이 활용되고 있고(Callon, 1990), 키워드의 연결 강도에 따라 군집화한 후 특정 시점의 위상을 파악하는 수준에 머무르고 있다.

과학계량학 분야에서 핵심 키워드를 기반으로 특정 분야의 연구 지형 파악 및 이슈 변화를 추적하는 대표적인 방법은 동시발생 키워드 분석 기법이다. 이는 키워드를 연결 강도에 따라 군집화한 뒤 각 키워드를 군집 내 연결 강도(Density, 밀도), 군집 간 연결 강도(Centrality, 중심성)에 따라 전략 다이어그램에 배치하고, 각 키워드의 위상을 해석하는 방법으로, 키워드의 위상 변화에 대한 해석에는 강점이 있으나 해당 키워드의 변화를 추적하는 데는 한계가 있다. 이와 관련하여 동시발생 키워드 분석에서 군집 내, 군집 간 연결 관계에 대한 구분을 확장하여 네트워크 상의 위상을 세분화하고 위상 변화를 추적하는 방법론이 제시되었다(Guimera & Amaral, 2005). 한편, Schiebel et al.(2010)은 해당 분석 영역에서 키워드의 성격을 해당 영역 신규 등장 키워드, 해당 영역 특화성 확립 키워드, 다영역 횡단 키워드로 구분하고 각 키워드의 해당 특성이 시·공간별로 어떻게 변화하는지를 분석하는 모델(Pathway-3)을 제시하였다. 키워드 동시발생 분석에서는 포착하기 어려웠던 키워드 자체의 진입, 성장, 교체 등 시계열 변화의 측면을 분석하고 이를 바탕으로 해당 학문 영역의 변화를 파악할 수 있는 장점이 있으나, 사후적 분석 성격이 강해 특정 관심 키워드의 동태적 변화 추적과 예측에 활용되기는 어려운 단점이 있다. Cunningham & Kwakkel(2011)은 대부분의 키워드 기반 분석이 전체 키워드 네트워크 내 해당 키워드의 위상 변화에 따른 전체 영역의 지형변화 분석에만 초점을 맞추고 있다는 단점을 극복하고자 개별 키워드 증감 데이터에 기반, 성장 곡선과 속도, 모멘텀 변수를 추정하여 키워드가 가지는 중요성을 구분하는 분석 틀을 제시하였다. 그러나 개별 키워드 각각의 빈도 변화에만 초점을 맞추어, 키워드 간 관계를 파악하기에는 적합지 않은 양상을 보인다.

본 연구에서는 기존 지엽적 키워드 분석의 단점을 극복하고자 키워드의 계층 분석을 통하여 키워드의 종합적 변동 탐색 가능성을 제시하였으며, 이는 과학기술 이슈 탐지 활동의 신뢰도를 높여 이슈 기술의 현황 및 시사점을 분석하는 일련의 과정에 도움을 줄 수 있을 것이다.

* 심위, 과학기술연합대학원대학교 박사과정, 02-3299-6150, sw@kisti.re.kr

** 권오진, 한국과학기술정보연구원 책임연구원, 02-3299-6097, dbajin@kisti.re.kr

*** 김근환, 한국과학기술정보연구원 선임연구원, 02-3299-6072, khkim75@kisti.re.kr

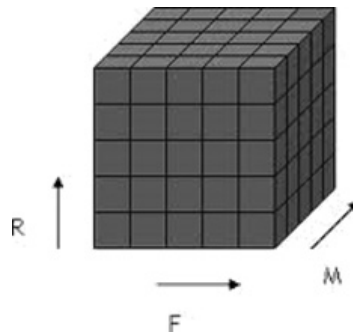
**** 정승호, 한국과학기술정보연구원 위촉연구원, 02-3299-6254, jsh79@kisti.re.kr

***** 김도현, 명지대학교 교수, 031-330-6449, ftgog@mju.ac.kr

II. 연구방법론

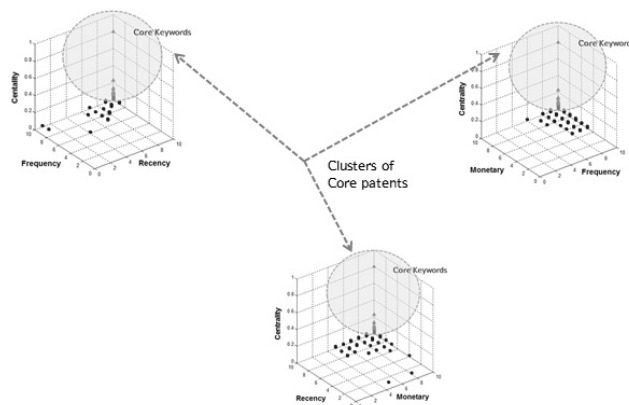
본 연구의 주제인 과학기술 이슈 발굴을 위한 키워드 경로 탐색 방법은 다음과 같다. 먼저 RFM(Recency, Frequency, Monetary) 분석을 통해 핵심 키워드를 선별하고(Kim et al., 2012), 논문의 인용 관계를 활용하여 키워드의 방향성을 부여하였다. 이어서 두 노드 간 최단 경로 기반 Link Betweenness를 계산하여 주 경로 탐색을 수행하였다.

핵심 키워드 선별 방법에는 RFM 분석이 수행되었는데, 이는 마케팅 분야에서 유망고객을 선별하는 변수로 활용되는 구매 최근성(Recency), 구매 빈도(Frequency), 구매 금액(Monetary)을 개별 키워드에 수정 적용한 것이다. Recency는 학술 데이터베이스에서 사용되는 저자 키워드의 연도별 빈도 및 빈도의 기울기를 측정된 것이며, Frequency는 해당 키워드의 총 빈도라 할 수 있다. Monetary는 해당 키워드를 저자 키워드로 가지는 논문의 총 피인용 수를 조사하여, 많은 인용을 받은 논문의 키워드일수록 높은 Monetary 값이 높아지도록 계산한 값이라 할 수 있다. 이러한 세 가지 변수의 값을 Percentile을 이용하여 표준화한 후, 높은 값을 가지는 키워드를 핵심 키워드로 선정하였다.



(그림 1) Recency, Frequency, Monetary를 통한 키워드 분류 도식화

이 과정에서 핵심 키워드 발굴을 위한 RFM 분석의 유용성을 검증하기 위하여 키워드들의 네트워크 연결 중심성(Degree Centrality)를 활용하여 핵심 키워드를 재선정하고, RFM분석을 통해 예측된 핵심 키워드와 비교분석을 수행하였다. 비교 결과 RFM 분석으로 인해 선별된 키워드를 핵심 키워드로 활용할 수 있음을 확인하였다.



(그림 2) RFM 분석의 유용성 검증

다음으로 키워드의 경로 탐색을 위한 흐름을 파악하기 위하여 선별된 키워드들에 방향성을 부여하였다. 키워드에는 방향성이 존재하지 않으므로 해당 키워드를 저자 키워드로 지정한 논문의 인용 관계를 활용하였다. 이어서 Link Betweenness를 통하여 주 경로 탐색을 수행하였다. Link Betweenness란 노드와 노드 사이의 최단거리를 연결하는 Path 위에 위치한 링크일수록 중심성이 높아진다는 개념을 수식화한 지표로서, Link Betweenness가 높을 수록 두 집단을 연결하는 중요한 역할을 수행한다.



(그림 3) 주 경로 탐색 방법

논문의 Main Path 알고리즘에 사용되는 인용 네트워크의 경우, 인용의 흐름이 단방향으로 흘러가는 유향 비순환 네트워크(Directed Acyclic Graph Model, DAGs)임에 반해 키워드 네트워크는 두 키워드 간 쌍방향의 흐름이 존재하게 된다. 이는 곧 역방향 경로, 순환 경로가 존재할 수 있다는 의미이며, 모든 경로를 활용하여 경로를 탐색하는데 난점으로 작용한다. 따라서 본 연구에서는 Main Path 알고리즘 대신 키워드 간 최단경로를 중심으로 경로를 탐색하는 Link Betweenness를 활용하였다.

<표 1> 논문의 Main Path 알고리즘과 키워드 경로 탐색 알고리즘 비교

구분	논문의 main path 알고리즘	키워드 경로탐색 알고리즘
Node	논문	논문의 저자 키워드
Link	논문의 인용/피인용	해당 키워드를 갖는 논문의 인용/피인용
네트워크 형태	인용/피인용 기반의 방향성이 있는 비순환 네트워크	인용/피인용 기반의 방향성이 있는 순환 네트워크
주경로 탐색 방법	NPPC, SPLC, SPNP	Link Betweenness

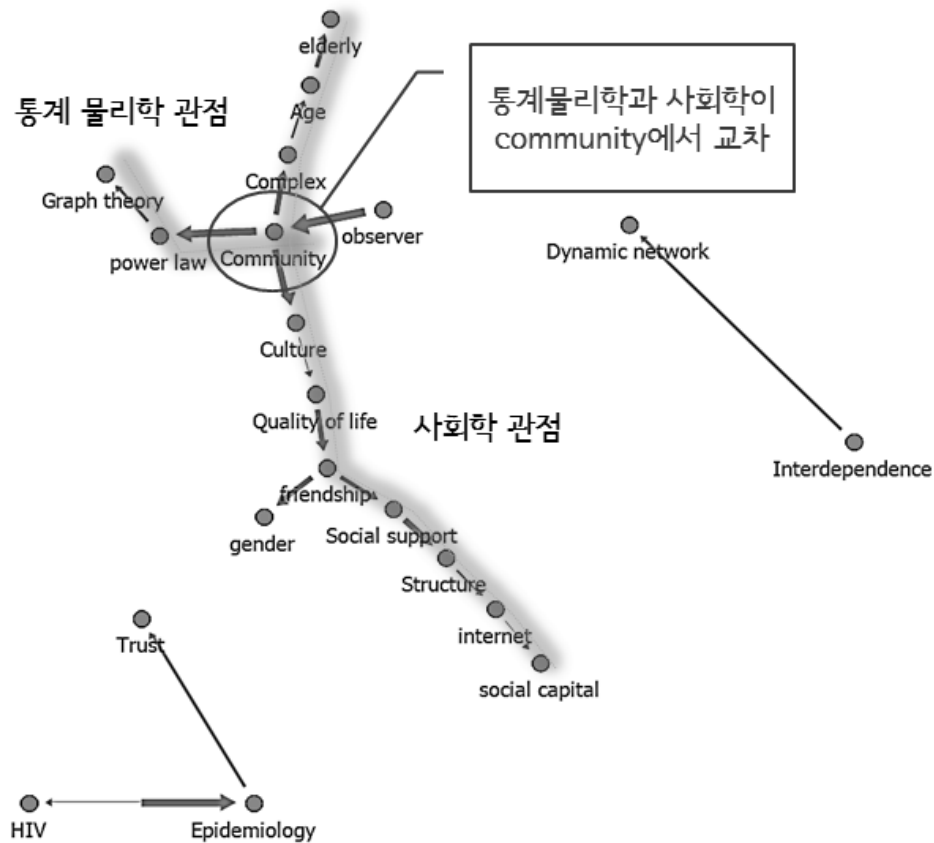
III. 결론

앞서 기술한 방법론을 적용하기 위하여 Web of Knowledge 데이터베이스에서 Complex Network, Stem Cell, Information Science & Library Science 분야에서 도출된 키워드의 흐름을 살펴보았다. 각 검색식과 논문 수는 다음과 같다.

<표 2> 분야 별 검색식과 논문 수

분야	검색식	논문 수
Complex Network	TS=("social networks" OR "social network" OR "random networks" OR "random network" OR "small-world" OR "scale-free" OR "complex networks" OR "complex network")	25,963
Stem Cell	TS=("stem cell") and PY=2012-2013	25,314
Information Science & Library Science	WC=("information science & library science") and PY=2001-2013	25,346

먼저 Complex Network에 키워드 경로 탐색 모델을 적용하였다. 이 때 Complex Network의 세부 이슈 경로를 탐색하기 위하여 검색식에서 사용된 키워드는 분석 대상에서 제외하였다.

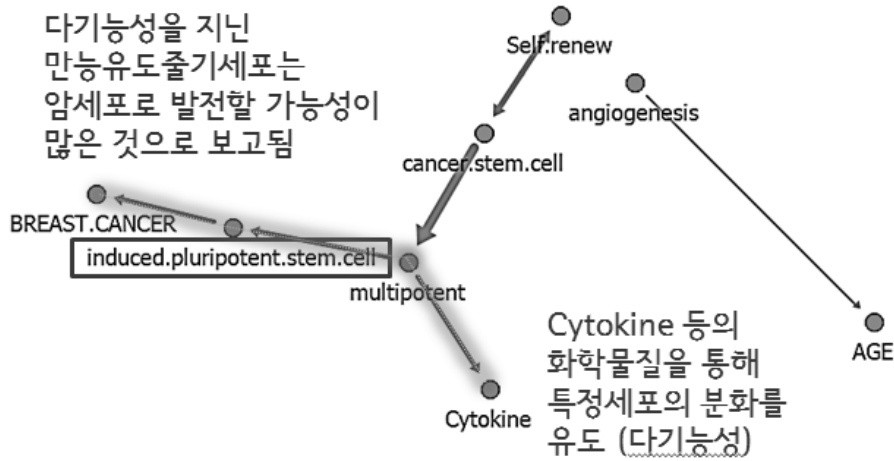


(그림 4) Complex Network의 이슈 흐름 파악

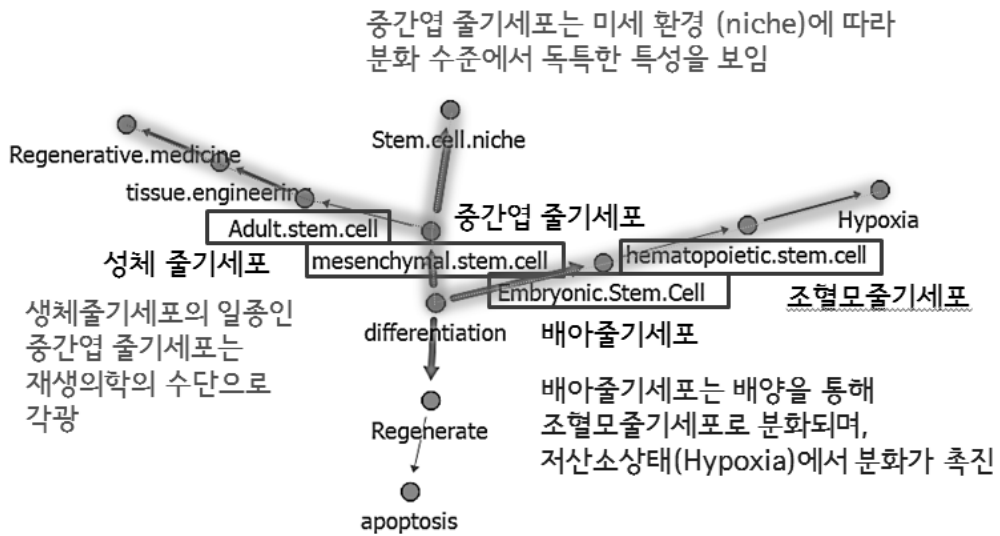
Complex Network 분야의 키워드 경로 탐색 모델 적용 결과, 사회학 관점 접근과 통계물리학 관점의 주 경로를 확인할 수 있었다. Social capital, Quality of life 등 사회학 관점의 키워드가 다수 도출되었으며, Graph theory, Power law 등 통계물리학 관점이 키워드 역시 일부 나타났다. 이 때 두 학문의 교차점은 Community 키워드인 것으로 확인되었다.

다음으로 Stem Cell 분야에 해당 모델을 적용하였다. 줄기세포 관련 키워드는 유도줄기세포와 배아줄기세포/성체줄기세포 키워드로 분리되어 나타났다. 유도줄기세포 관련 연구로는 다기능성 만능유도줄기세포가 압

세포로 발전할 가능성이 많다는 점, Cytokine 등을 통한 특정 세포의 분화유도 관련 연구가 주를 이루고 있다는 점 등을 확인하였으며, 성체줄기세포의 경우 재생의학의 수단으로 생체줄기세포의 일종인 중간엽줄기세포에 대한 연구가 다수 이루어지고 있는 것으로 나타났다.

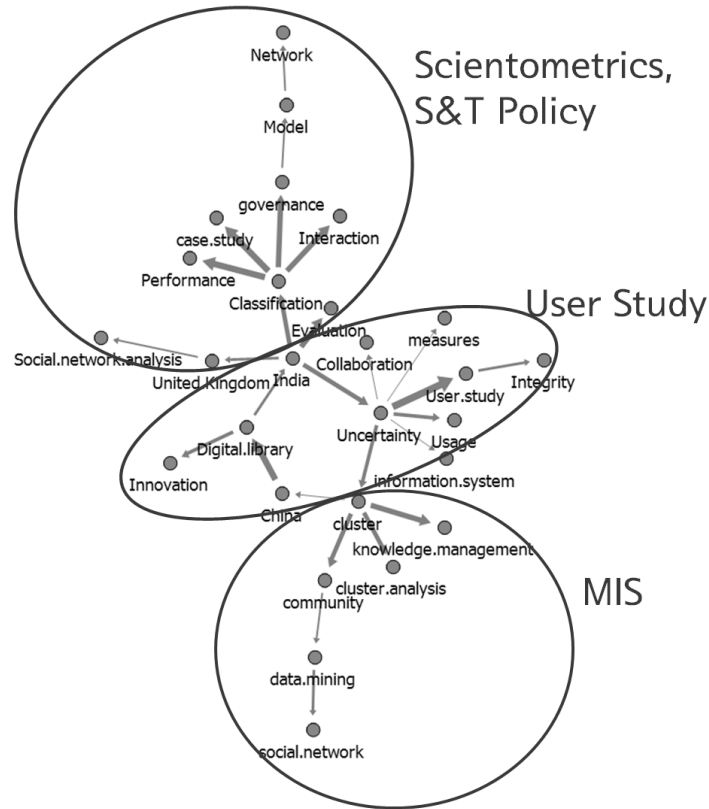


(그림 5) 유도줄기세포 관련 연구의 이슈 흐름 파악



(그림 6) 배아줄기세포/성체줄기세포 관련 연구의 이슈 흐름 파악

Information Science & Library Science 분야는 Scientometrics, S&T Policy, User Study, MIS 등의 키워드가 주요 관심 키워드인 것으로 나타났으며, 개별 관심 키워드를 중심으로 키워드가 분화된 것을 확인하였다.



(그림 7) Information Science & Library Science 관련 연구의 흐름 파악

본 연구에서는 마케팅 분야에서의 RFM 분석 기법을 통한 키워드 계층 분석으로 키워드의 종합적 변동 탐색 가능성을 제시하였으며, Complex Network, Stem Cell, Information Science & Library Science 분야에 대한 사례 분석을 수행하였다. 또한 본 연구는 기존 문헌의 흐름을 파악하는 Main Path 알고리즘을 변형하여 키워드 중심 과학기술 이슈 흐름을 파악할 수 있는 방법론을 개발함으로써 과학기술 이슈의 계량적 분석을 위한 일련의 과정에 도움을 줄 수 있을 것이다.

참고문헌

- Callon, M. (1990). "Techno-economic Networks and Irreversibility", *Sociological Review*, London, Routledge. 38: 132-161.
- Cunningham, S.W., Kwakkel, J.H.(2011), "A complex network perspective on the world science system", *Science and Innovation Policy*, 2011 Atlanta Conference on, 1-17
- Guimera, R., Amaral, L.(2005), "Functional cartography of complex metabolic networks", *Nature*, 433(7028): 895-900.
- Kim, D., Lee, J. Y., Ahn, S., Moon, Y. and Kwon, O. J.(2012), "RFM Analysis for Detecting Future Core Technology". *Proceedings of ACM RACS 2012*, 55-59.
- Schiebel E., Hörlesberger M., Holste, D. Roche I., François C. & Besagni D.(2010), "Identification of emerging technologies through pathway-III analysis", *Proceedings of the 19th International Conference on Management of Technology*. Cairo, Egypt, March 8th-12th, 2010