과학기술 정보분석을 통한 유망연구영역 탐색 방법론

A Methodology for a Emerging Technology Search through Science and Technology Information Analysis

이우형*, 문영호**

* 한국과학기술정보연구원 정보분석센터 미래전략팀 (leewh@kisti.re.kr)
** 한국과학기술정보연구원 정보분석센터장 (yhmoon@kisti.re.kr)

Abstract

본 논문에서는 과학기술 정보분석을 위한 방법론 중의 하나인 지식맵을 기반으로 유망연구영역을 탐색하는 방법론을 다룬다. 이를 위해 정보보안 분야를 대상으로 지식맵을 작성하고, 또한 허브탐색을 통한 정보보안 분야의 유망연구영역을 제안하였다.

1. 서론

세계는 그 동안의 노동 및 자본 중심의 성장전략의 한계로 인해 새로운 패러다임으로의 전환을 요구하였고 이에 기술과 지식이 가장 중요한 핵심요인 지식기반경제로의 전환에 성공하였다. 여기서 말하는 지식기반경제로의 이행을 촉진시키는 촉매제는 다름 아닌 과학기술 지식을 들 수 있다. 예를 들어, 지금까지 인류문명의 2000년 동안 이룩한 과학기술의 중요한 발견과 성과는 지난 100년간에 이루어진 것이 전체의 60%에 이른다고 한다. 그리고 20세기의 100년 중에서도 지난 20년간의 성과가 21세기의 80%를 차지한다고 한다(문영호, 2004).

이러한 세계적 흐름에 대처하기 위해 정부에서는 과학기술기본계획을 수립하고 2002년부터 2006년까지 신기술분야(IT, BT, NT 등 6T)에 약 13조원을 집중투자하는 등 각고의 노력을 기울이고 있다. 이렇듯 신기술에 대한 투자가 증대됨에 따라 투자의 방향을 국가전략목표와 시장요구에 부합하도록 일관성을 가지고 설계해야하며, 특히 효율성을 높일 수 있는 전략을 수립해야 할 것이다. 그러나, 새로운 과학기술이 갖는 수많은 과학기술적 가능성과 불확실성이 존재하는 현실에서, 어떻게 유망한 과학기술을 찾아내고 어떻게 이를 육성·발전시킬 것인가에 대한 심각한 고민이 뒤따라야 할 것이다. 이러한 추세에 따라 요즘 국내외적으로 관련 연구가 활발하게 진행되고 있으며, 그 중에 대표적인 것이 본 연구에서 중점적으로 다루고자하는 과학기술정보분석 방법론이 있다.

과학분야나 사회과학분야에서의 개념, 아이디어, 그리고 문제들 간의 연계성을 도식화할 수 있도록 하는 것은 매우 중요하며, 이러한 도식화를 위해 몇 가지 방법이 시도되어졌다. 과학연구와 과학정책에서 사용되어진 전통적인 방법은, 상대적으로 소수 전문가들의 견해를 구하는 방법이었다(Law and Whittaker, 1992). 과학기술정보분석 연구 또는 서지분석은 정량적 측면에서 이러한 작업을 수행하기 위한 또 다른 방법이다.

혁신주도형 경제와 과학기술정책의 패러다임변화에 따라 우리나라는 과거의 추격적 전략(catch-up strategy)을 포기할 수는 없지만 동시에 선택과 집중(selection & intergration)으로 원천기술을 확보해야 하는 단계에 이르고 있다. 단순한 과학기술의 추적보다는 국가적 아젠다에 맞춘 전략적 지식의 창출과 혁신시스템(innovation system)의 구축이 필요하다(오세정·이달곤, 2004). 또한 연구개발사업 자체는 많은 국가예산을 요하는 것일 뿐만 아니라 과학기술 및 경제사회 전반에 미치는 파급효과가 크기 때문에 어떤 사업을 육성할 것인가를 결정하거나, 사업의 개시여부 등을 결정할 때 최대한 신중하고 객관적인 접근이 필요다. 이러한 요구에 부합하기 위해 많은 학자 및 연구원들이 객관적인 연구방법들을 통해 이를 찾아내려는 시도들이 많았다. 이에 본 연구에서는 그러한 취지에서 조기에 육성사업을 찾아내는 노력이 필요하게 되고 이러한 노력의 일환으로 과학기술 정보분석방법을 통한 신기술 도출 방법론을 연구하게 되었다.

2. 이론적 고찰

2.1 서지분석

bibliometrics와 scientometrics라는 용어는 1969년에 Prichard와 Nalimov & Mulchenko에 의해 거의 동시에 도입되었다. Prichard는 bibliometrics를 "책과 다른 의사소통 미디어에 대한 수학적이고 통계적인 방법론의 응용"이라고 설명한 반면에, Nalimov and Mulchenko는 scientometrics를 "정보 프로세스로서의 관점을 가진 과학분석을 처리하는 계량적 방법론의 응용"이라고 정의했다. 그러한 해석에 따르면, 전문적인 scientometrics는 과학 의사소통의 측정으로만 제한하는 반면에, bibliometrics는 좀 더 일반적인 정보 프로세스를 처리하도록 설계되었다. 어쨌든, 지난 30년 동안 두 전공간의 흐릿한 경계는 거의 사라졌고, 오늘날에는 두 용어가 거의 동의어로 사용되고 있다. 대신에, informetrics 분야가 최초의 광범위한 bibliometrics를 대신하고 있다. informetrics라는 용어는 VINITI(Gorkova, 1988)에 의해 채택되었고, 과학의 의사소통 프로세스인 수학적 통계분석을 다루는 좀 더 일반적인 정보과학의 하위분야를 대표하게 되었다.

bibliometrics의 최초 정의와 반대로, informetrics는 또한 전자적 매체를 다루며, 정보생산 프로세스를 위한 모델과 정보검색의 계량적 측면, 전자 도서관에서의 정보 측정, 도서관 대출부수, 텍스트와 하이퍼텍스트 시스템의 통계분석과 같은 주제를 포함하고 있다.

2.2 지식맵

본 연구에서 서술하는 과학기술 분야의 급성장은 당연한 것으로 받아들여지고 있으며 그 속도의 측정은 어느 정도 가능하다. Price(1961)에 따르면 5년 동안 국제전화 통화량은 두 배로, 15년 동안 과학저널의 수가 두 배로, 그리고 20년 동안 대학의 수가 두 배로 증가할 것이라고 추정하였다. 또한 그는 이 법칙이 확산되면 과학 분석의 기본 법칙이 될 것이라고 확신하였고 그는 가스를 이용한 비유에서 분자들이 서로 상호작용하는 방식을 기술하기 위하여 "보이지 않는 대학(invisible college)"이라는 개념을 사용하였다. 그는 특히 과학의 미개척분야로서의 매핑 주제와 관련된 현재 과학학문들의 개념지도를 만들 수 있는 인용 연구를 최초로 제안한 사람으로, 이런 과학학문들의 위상도에서 지도 내 위치에 따른 저널, 저자 혹은 개별 논문들의 중복과 비교우위를 보여줄 것이다.

과학의 역사와 철학에서 이루어진 연구 특히 Kuhn(1962)의 과학진화구조론, Thagard(1992)의 개념적 진화론, Crane(1972)의 "보이지 않는 대학"은 일반 과학자들뿐만 아니라 정보과학자들 세대에 큰 영향을 끼쳤다. 과학 패러다임의 추적(tracking) 개념은 이러한 영향에서 파생되었다. 이러한 수많은 노력 가운데 가치 있는 두 가지 연구가 있다. 하나는 인용분석을 통한 매핑 과학에 있어 ISI의 Small and Garfield(1974)의 연구이다. 또 다른 하나는 유명한 동시단어분석을 이용하여 과학문헌들의 변화를 추적한 Callon et al.(1986)의 연구이다. 사실 그들은 동시단어분석을 더욱 광범위한 과학분야에 적용될 수 있도록 설계하였다. 오늘날 새로운 추세는 과학 프론티어의 원동력을 더욱 강조하고 다음과 같은 질문에 해답을 주기 위해 노력하고 있다. 과학적 논쟁에 있어 주요 이슈는 무엇인가? 유행하고 있는 이론의 진화에 공헌한 것은 무엇인가? 패러다임의 이동 과정을 어떻게 시각화할 것인가? 과학분야에서 어떤 패러다임이 사라지고 어떤 패러다임이 도래하고 있는가? 과학 영역을 시각화하는 가장 적정한 방법은 무엇인가?

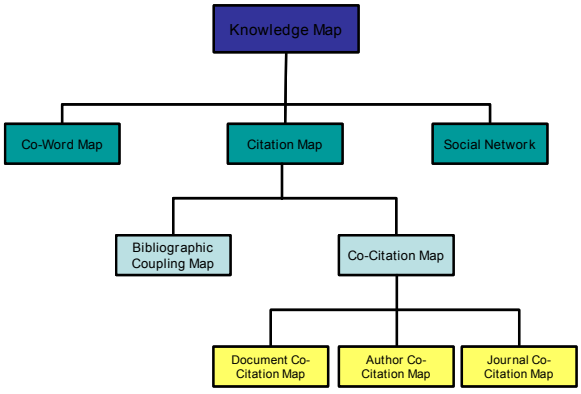
이렇게 과학을 총체적으로 측정하고 연구하는 전통적인 추세 가운데서 과학의 미개척분야로서의 매핑은 이제까지 전례가 없었던 전환을 겪고 있는 중이다. 과학을 과학 그 자체에 적용하기 위해 우리는 과학적 활동의 본질인 과학 철학과 사회학을 이해할 필요가 있다. 과학에 대한 이해는 과학분야의 시각화와 함께 시작되었으며 이는 현대 과학자들이 일상의 연구에서 수행하고 있는 일이 본질적으로 시각화하고, 해석하고, 설명하는 일이라는 것을 보여주고 있다. 우리는 과학영역을 시각화하기 위해 어떠한 은유법을 사용할 수 있을까? 지식영역의 시각화를 위한 우리의 탐색과정은 물리적 세계에서의 육지와 천체 현상의 매핑, 개념지도 작성법과 과학문헌의 지식구조, 과학영역의 원동력을 표현한 통계적 묘사와 지식맵으로부터 출발한다.

2.3 지식맵의 종류

과학기술정보 분석방법론으로 일컫는 bibliometric은 일반적으로 논문, 책, 보고서 등이 포함된 명시화된 지식을 대상으로 수학적, 통계적 방법을 적용하는 것으로, 분석의 결과는 주로 맵(정확하게 표현하면 "위치도"라고 표현하는 것이 더 나을 듯 하다)의 형태를 취한다. 그리고 이러한 맵을 여러 학자들은 지식맵이라고 명명하여 사용하고 있다.

지금까지 연구된 결과를 기반으로 지식맵의 종류를 살펴보면 크게 두 가지 즉, 동시단어분석맵과 인용분석맵으로 나눌 수 있다. 그러나 그 출발점은 조금 상이하지만 사회학과 물리학에서 활발히 연구되고 있는 네크워크분석도 그 결과는 맵의 형태를 취하고 있고, 이 방법론에서 사용하고 있는 기본 알고리즘을 일반 지식맵 분석에서 사용하는 알고리즘에 많이 이용하고 있는 상황이기 때문에 이제는 네트워크 분석을 별도의 또 다른 분석방법이라고 말하는 것은 무리가 있어 보인다.

실제로 본 연구에서 주장하고 있는 신기술 도출 방법론의 상당부분은 네트워크 분석에서 사용하고 있는 지표들을 사용하고 있다. 따라서 본 연구에서는 지식맵을 동시단어분석맵, 인용분석맵, 그리고 네트워크 분석맵으로 분류하고자 한다.

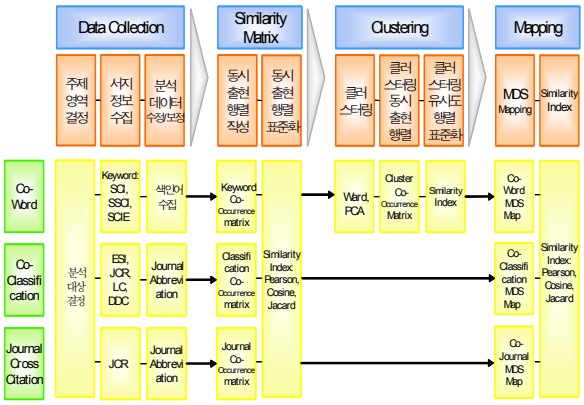


<그림 1> 지식맵의 종류

2.3 지식맵 구축 프로세스

연구분야 혹은 학문분야의 발전상태를 통찰하는 방법 중 하나로 그 분야 연구자들에 의해 발행된 문헌들을 고찰하는 방법이 있다. 문헌들을 분석하고 계산하는 수많은 방법 중 아마도 데이터의 본질을 밝히는 가장 좋은 방법은 그 문헌 내에서 연구자들이 사용한 키워드를 관찰하는 방법일 것이다. 이전 문헌들의 키워드를 통해 우리는 현재 사용되고 있는 아이템뿐만 아니라 이를 처음 사용한 시기까지 알 수 있다.

Science Citation Index(SCI)와 Social Sciences Citation Index(SSCI) 등과 같은 문헌 DB에 키워드들이 대량으로 쌓이게 되면, 그 집합패턴을 분석하여 논문에 의해 상징된 연구분야의 중대관심사를 밝혀낼 수 있다. 이것이 이전 시기의 문헌들을 통해 현재 개념적 프레임워크를 파악할 수 있는 방법으로, 지식맵 구축을 위한 프로세스는 다음과 같다.



<그림 2> 지식맵 구축 프로세스

3. 과학기술 정보를 통한 신기술 도출 방법론

3.1 개요

과학기술 정보를 분석함에 있어 가장 먼저 고려해야 할 사항은 신기술의 특징이다. 여기서 말하는 신기술의 특징은 다른말로 표현하면 신기술의 불확실성을 의미한다. 과학기술정보 분석을 통한 신기술 도출을 위한 사전 작업으로 신기술의 불확실성에 따라 상이한 분석방법을 사용해야 함을 주장한 바 있다(<표 1> 참조). 본 연구에서는 이를 기반으로 불확실성 수준이 3-4 이상의 신기술을 분석하는 방법론 개발에 초점을 맞추었다.

<표 1> 신기술 불확실성 수준에 따른 분석기법

	범위설정	탐색	평가	몰입	분석기법
불확실성 수준 4	전략적 의지설정 국가역량 분석	과학기술 정보원천탐색	신기술 출현 패턴 분석	체계적인 신호 감지를 위한 투자	KDD, Text Mining, Data Mining, Bibliometric, TRM, 기술예측, 시나리오 기법
불확실성 수준 3	기술개발 범위도출	약한 신호 감지 기술 및 잠재시장 예측	후보기술군 도출 시장개발에 관한 시나리오 도출	시나리오별 예상 스토리 작성/기획	
불확실성 수준 2	기술 대안도출	강한 신호 감지 경쟁 기술 파악 경쟁 시장 파악	평가기준 설정 기술선택	기술개발 계획 작성 및 실행	Real Option
불확실성 수준 1				상용화 계획 작성 및 실행	전통적 경제성 분석기법 (ROI, NPV)

불확실성 수준: 높음 ↑ / 낮음 ↓

자료원 : (윤문섭 외, 2004)

따라서 본 연구에서 개발하고자 하는 신기술 도출방법은 불확실성 수준이 가장 높은 기술에 대해 적용 가능하다. 그리고 현재 이러한 신기술의 불확실성을 극복할 수 있는 분석 방법론으로 현재 많은 논의가 진행되고 있는 것이 네트워크를 통한 분석 방법이다.

이러한 네트워크 분석의 목적은 구조나 연결망 형태의 특징을 도출하고, 관계성으로 체계의 특성을 설명하거나 체계를 구성하는 단위의 행위를 설명하는 것이다. 이러한 네트워크를 통해 연구활동에 필요한 지식자원의 상호연관성을 이해하는 것은 미래의 유망분야를 예측하거나 제한하는데 도움이 될 것이다. 네트워크적 사고는 유망분야가 확산되는 경로를 예측할 수 있도록 하며, 노드를 강화하는 방법을 알려 줌으로써 연구자들에게 연구방향을 제시해주고, 또 국가나 기업차원에서도 연구자들의 연구과제의 방향과 연구목표가 올바르게 진행되고 있는지 등에 대한 좌표를 제시해 준다.

과학 기술활동에서 일어나는 네트워크의 다양성은 이루 다 열거 할 수 없을 정도이다. 정책 네트워크, 협력체계의 네트워크, 조직에 의한 네트워크 등 그 예는 무궁무진 하다. 이렇게 다양한 네트워크를 총체적으로 통합하여 하나의 원칙을 이끌어낸다는 것 자체가 불가능 할지도 모른다. 그러나 다행인 점은 어떠한 단계의 네트워크를 살피든 하나의 보편적 법칙이 있다는 것이다. 다만 당면한 문제는 이러한 법칙을 과학기술 네트워크에 얼마나 적절히 활용하느냐에 있을 것이다.

따라서 네트워크 분석에 신기술 도출방법론을 적용하게 되었다는 것은 네트워크와 다른 연구활동에 필요한 지식자원들의 자료적 현상들을 연결할 수 있게 되었으며, 또한 네트워크를 복잡한 시스템 일반에 대한 연구의 전면에 부각시켰다는 점에서 중요한 의미가 있다.

3.2 연구 프레임워크

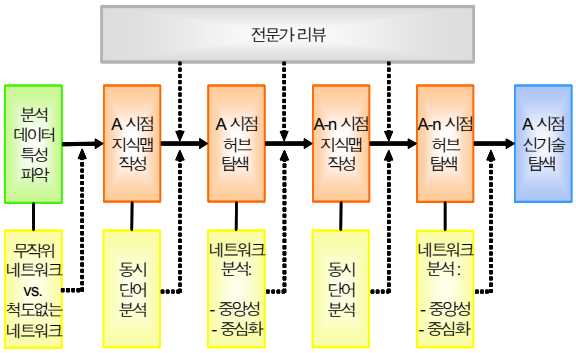
본 연구에서 구축하고자 하는 과학기술정보분석을 통한 신기술 도출 방법론은 네트워크적 시각에서 구축된다.

네트워크적 분석방법에 의한 구축을 하기 위해서 첫째, 먼저 분석자료의 성격을 규명하는 작업이 선행되어야 한다. 데이터의 성격은 무작위 분포(Random Distribution)를 따르는가 멱함수 분포(Scale-free; Power law Distribution)를 따르는가로 구분하게 된다.

둘째, 동시단어분석을 통해 분석시점의 지식맵을 작성하게 되는데 이를 통해 분석대상의 특징을 살펴볼 수 있다. 그리고 네트워크 분석 방법을 이용하여 분석 데이터의 허브를 탐색하는데 이때 사용되는 것은 네트워크 분석 지표이다. 네트워크 분석 지표는 연결성, 중앙성, 중심화 등의 지표를 이용하였다.

셋째, 이러한 방법에 의해 허브를 찾았으면 이제 이 허브가 처음 등장했던 시기로 거슬러가서 그 당시 도메인 맵 분석을 실시하였다. 이때 그 당시의 허브를 중심으로 어린 허브(즉, 현재의 허브)와의 관련성을 앞서 고찰한 지표들을 이용하여 분석하였다.

이제 이 분야의 전체적인 특성이 찾아졌다면 다시 현재로 와서 현재의 허브를 중심으로 향후 허브가 될 가능성이 있는 도메인을 찾는다. 즉, 현재의 도메인 맵에서 허브를 중심으로 이 분야에서 향후 허브가 될 가능성이 있는 지표들의 특성과 가장 많이 유사한 도메인을 선정하는 것이다.



주) A 시점 : 분석 기준 시점
A-n 시점 : A 시점에서 허브로 결정된 도메인이 처음 등장한 시점

<그림 3> 신기술 도출을 위한 Framework

3.3 허브의 탐색

Gradwel(2000)l은 어떤 개인이 얼마나 사교적인가를 테스트해 본 결과를 통해 "모든 계층을 막론하고 친구나 아는 사람을 만드는 데에 있어서 극히 예외적인 솜씨를 가진 소수의 사람들이 있고, 그들은 허브(또는 커넥터)라 할 수 있다"로 결론 내렸다. 또한 Barabasi(2002)는 허브들(비정상적으로 많은 링크를 갖고 있는 노드들)은 경제에서 세포에 이르기까지 매우 다양하고 복잡한 시스템에서 발견된다는 사실을 실증분석을 통해 증명하였다. 그리고 그것들은 대부분의 네트워크의 근본적 속성이라 할 수 있으며, 생물학, 컴퓨터 과학, 생태학 등 다양한 분야의 과학자들의 호기심을 자극하고 있는 것이라고 하였다.

Gradwell(2000)이나 Barabasi(2002)가 말한 것처럼 이러한 허브들은 우리의 사회 네트워크에서 극히 중요한 부분으로, 그들은 경향과 유행을 만들며, 중요한 거래를 성사시키고, 유행을 전파하며, 식당을 개업하는 것을 도와주기도 한다. 그들은 서로 다른 인종, 교육 수준, 가문들을 부드럽게 이어주는 사회의 실과 같은 존재이며, 따라서 본 연구의 주제인 조기경보시스템을 통해 분석하고자 하는 분야에서도 이러한 허브들이 전혀 다른 학문분야, 국가, 연구분야간을 연결하여 새로운 학문을 전파시켜 패러다임을 형성할 수 있을 것이다.

허브는 분명 주목할 만한 가치가 있고, 특별한 것으로, 허브는 전체 네트워크의 구조를 지배하며, 그것은 좁은 세상으로 만드는 역할을 한다. 즉, 허브는 엄청나게 많은 수의 노드들과 링크를 가짐으로써 시스템 내의 두 노드 간의 경로를 짧게 만드는 것으로, 노드의 연결선 수가 멱함수 법칙적 분포를 따르는 네트워크를 척도없는(Scale-Free) 네트워크라고 한다. 따라서 이러한 척도없는 네트워크는 몇 개의 거대한 허브를 가질 것이며 이들이 전체 네트워크의 위상구조를 규정하게 될 것이라고 예측할 수 있다.

허브를 찾기 위한 노력은 네트워크에 관심이 있는 학자들 사이에 최대 이슈 중 하나일 것이다. 그러나 지금까지 연구결과를 살펴보면 네트워크 분석에서 일반적으로 통용되는 방법은 나오지 않고 있다. 그중에서 가장 두드러진 연구결과를 소개하면 다음과 같다. 먼저, Chen(2004)은 네트워크 분석을 통해 지식 전환점을 찾기 위한 방법으로 Turning point, Hub, Pivot node 등과 같은 세 가지 개념의 노드를 소개했다. 그러나 각 노드들을 정확하게 측정하는 방법을 제시하지는 못했다.

본 연구에서는 네트워크분석에서 사용되는 지수들로 구성되어 있다. 그리고 Map Indicators에는 크게 결속, 중앙성, 구조적 등위성 등이 있다.

4. 실증분석

4.1 자료수집 프로세스

정보보안기술에 대한 연구주제 선정방법론을 실제 적용하기 위해 제일 먼저 할 작업은 분석 자료를 수집하는 것으로, 분석 자료는 논문과 특허가 대표적이며, 본 연구에서는 논문만을 분석대상으로 선택하였다.

논문분석을 위해 1994년부터 2003년까지의 논문이 수록되어 있는 SCI-CD 버전을 이용하였고 정보보안기술과 관련된 키워드를 보스톤 대학교 정보시스템 학과의 관련 교수들과 협의하여 범위를 정하였다. 이렇게 만들어진 분석범위를 기초로 다음과 같이 28개의 키워드를 선정하였다.

<표 2> 1차 분석 키워드

Security, network, cryptography, encryption, authentication, secure, Intrusion, antivirus, detection, spam, vulnerability, digital signature , Verification, firewall, traffic, recovery, monitoring, hacking, attack, penetration, incident response, risk analysis, filtering, buffer overflow, Denial Service, DoS, audit, scan

분석 결과, 추출 문헌(8,464편)으로부터 저자 키워드 50,392(중복 없이 294, 394)개를 선정하였다. 그리고 이 키워드의 클린징 작업을 수행하였는데 수행결과 부적절한 키워드들이 상당부분 포함되 있다고 판단하여 아래 표와 같이 다시 한 번 키워드 조정을 실시하였다.

<표 3> 2차 분석 키워드

(security or secure or crypto*) and (encryption or authentication or intrusion or verification or firewall or recovery or hacking or audit or scan or algorithm*)

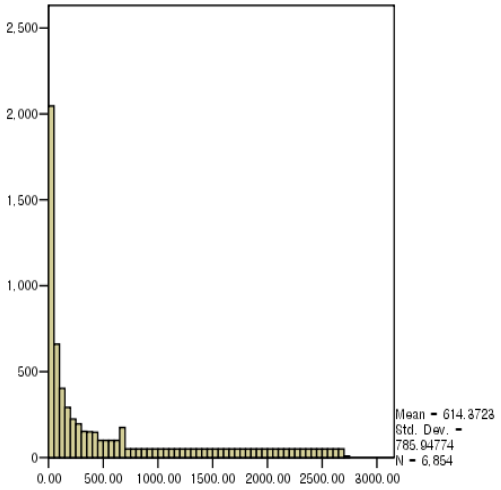
분석 결과, 추출 문헌 (976편)으로부터 저자 키워드 2,880(중복 없이 6,857)개 선정하였다. 이러한 각각의 논문들로부터, SCI와 SSCI의 색인가들에 의해 추가되어진 키워드들뿐만 아니라, 제목과 초록에서 중요한 키워드들을 직접 추출해 내었다. 색인가들에 의해 추가되어지거나 주제나 초록으로부터 선택되어진 모든 키워드들을 일관성(단수/복수)있고, 통일성(동의어)을 가지며, 명백(동음이의어)하게 만들기 위해 LISA 시소러스, LCSH와 Information Technology Terms의 시소러스(컴퓨터에 기억된 정보의 색인)를 사용하여 표준화 하였다.

단어빈도가 한 번이나, 두 번 정도인 단어들은 그들의 넓은 의미의 용어로 병합되어졌다. 단어의 빈도가 한 번이나, 두 번인 단어들 중, 넓은 의미나 비슷한 용어가 리스트에 없는 경우에는 그 단어를 무시하였고, 최종적으로, 빈도가 5 번 이상인 223개

의 키워드들이 동시단어분석의 연구샘플로 선택되어졌다.

4.2 분석 데이터 특성 파악

신기술 탐색을 위해 첫 번째 수행할 일은 원천 데이터의 성격을 규명하는 것으로, 분석자료의 성격이 무작위 네트워크(the random network) 성격을 가지는지 혹은 척도없는 네트워크(scale-free network) 성격을 가지는지를 규명하는 것이다. 자료의 성격에 따라서 Hub를 찾을 수고 있고 그렇지 못할 수도 있기 때문에 이 작업은 상당히 중요하며, 본 연구의 분석대상인 정보보안 분야의 데이터의 특성을 분석한 결과 척도없는 네트워크의 성격을 가지고 있는 것으로 나타났다.



<그림 4> 분석 데이터의 특성

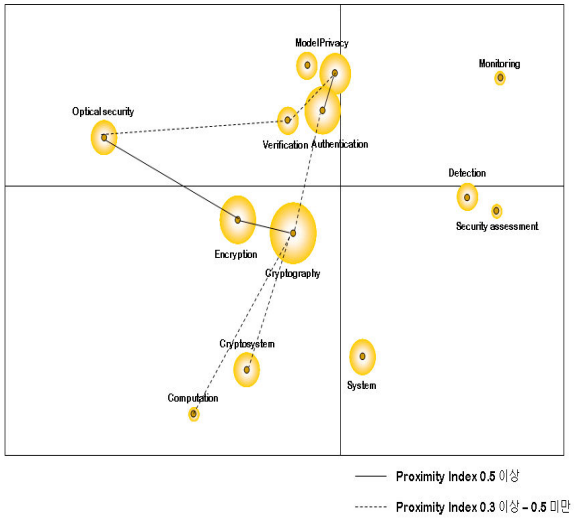
멱함수 분포란 <그림 4>에서 보는 바와 같이, 우하향하는 곡선의 특징을 보이는 분포이다. 노드의 연결선 수가 멱함수 법칙적 분포에 따르는 척도없는 네트워크에서는 대부분의 노드들이 단지 소수의 링크만을 가지고 그들이 연결선 수가 매우 많은 소수의 허브들에 의해 연결되어 있는 모양을 예상케 해준다. 멱함수 법칙은 시각적으로 항공노선 네트워크와 매우 유사한데, 여기서 다수의 작은 공항들이 소수의 주요 허브들을 통해 서로 연결되어 있는 모양을 하고 있다.

완만하게 감소하는 멱함수 분포는 변칙적으로 링크가 많은 허브를 자연스럽게 포용할 수 있다. 그것은 모든 척도없는 네트워크는 몇 개의 거대한 허브를 가질 것이며 이들이 전체 네트워크의 위상구조를 규정하게 될 것이라고 예측하게 된다. 많은 실증분석을 통해 이론적으로 중요한 대부분의 네트워크들이 척도없는 네트워크라는 사실이 밝혀지면서 허브는 정당성을 인정 받았다.

4.3 현 시점의 지식맵 작성

각 클러스터에서 높은 빈도를 갖는 키워드들이 그 클러스터를 대표하기 위해 선택되어졌고, 전반적 도식은 정보보안 영역의 13개 클러스터들의 전반적인 위치를 표현하기 위해 MDS (MultiDimensional Scaling) 기술을 이용하여서 완성되어졌다.

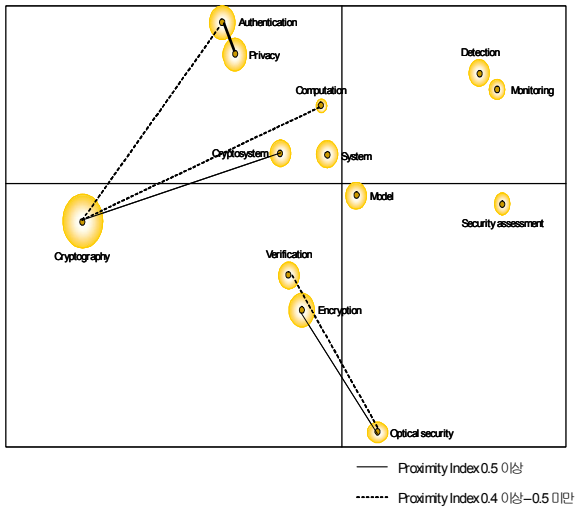
지도에서 보여지는(<그림 5>) 두 단어간의 실선은 Proximity 지수가 0.5 이상의 값을 가지는 높은 상관성을 나타내며, 두 단어간의 점선은 Proximity 지수가 0.3 이상 0.5 미만의 값을 가지는 낮은 상관성을 나타낸다.



<그림 5> 동시단어 분석을 통한 도메인 맵(1994-2003년)

4.4 특정 시점의 지식맵 작성

지금까지 정보보안 분야의 전체적인 모습을 지식맵과 네트워크 분석 지표들을 이용하여 분석하였다. 분석결과를 기반으로 이 분야의 전공 교수들과 협의한 결과 현 시점의 허브를 Encryption으로 선정하였고, Encryption 분야가 처음 등장한 시점이 언제인지를 탐색하여 1994년부터 1999년까지를 한 시점으로 선정하여 이를 분석하였다.



<그림 6> 동시단어 분석을 통한 도메인 맵(1994-1999년)

4.5 현 시점의 지식맵 작성

지금까지 현 시점(A시점; 2003년)에서의 지식맵과 허브를 탐색하였고, 현 시점의 허브가 출현했던 특정 시점(A-n 시점; 1994년)에서의 지식맵과 허브를 탐색하였다. 특정시점(1994년)에서의 허브와 어

린 허브와의 관계를 살펴보면 다음과 같다.

<표 4> 정보보안 분야의 허브의 특성

분석 지표	결과	의미
degree 정도	low	·연결된 결점의 수를 의미하는 것으로 중앙성을 나타냄 ☞ 어린허브는 아직까지 연결된 결점의 수가 적은 것으로 그 이유는 다른 연구들과에 비해 주체적인 연구가 아직까지는 어렵다는 의미로 판단됨
betweenness 정도	high	·한 결점이 연결망 내의 다른 점들 '사이에' 위치하는 정도를 측정함 ·사이 중앙성은 다른 결점들 사이에서 브로커 역할을 하는 정도를 측정함 ☞ 어린허브는 아직까지 주체적인 연구는 어렵지만 다른 연구를 활성화 시킬 수 있는 연구주제라는 의미로 판단됨
closeness 정도	low	·한 점의 전체 중앙성을 표현하는 대표적인 지표 ·두 점들간의 거리는 두 점들을 연결하는 최단 거리. 즉 경로거리를 말함 ·경로거리의 합이 가장 작은 결점이 전체 중앙성이 가장 높은 전체적으로 중심을 차지하는 점 ☞ 어린 허브는 정보보안 분야의 중심에 있지 못한 분야라는 의미로 해석됨

이러한 분석결과를 바탕으로 현 시점에서 향후 몇 년안에 정보보안 분야의 허브로 등장할 가능성이 높은 연구분야를 선정했다. 연구분야는 크게 3가지 정도로 압축되었는데, security assessment, computation, cryptosystem 등의 분야이다. 최종적으로 관련 교수들과의 협의를 통해 cryptosystem 분야가 향후 몇 년안에 이 분야에서 가장 활발한 연구주제가 될 가능성이 높을 것으로 나타났다. 따라서 현재의 허브에 연구연량을 집중하면서 앞으로 활성화 될 것으로 예상되는 crytosystem 연구분야에 관한 연구도 병행하는 것이 현 우리나라의 과학기술 정책의 기조인 "선택과 집중"을 가장 효과적으로 실현하는 길이라 할 수 있다.

5. 결론

본 연구는 목적은 과학기술정보를 이용하여 어떻게 신기술을 탐색할 것인가에 대한 방법론을 제시하는 것으로 이번 연구를 통해 의미있는 신기술 탐색 방법론을 제시하였으며, 본 연구에서 제시한 신기술 탐색 방법론은 그 어느 때보다도 정보분석 체제의 구축에 대한 필요성이 시급한 현 실정에서 의미있는 첫 시도라고 판단된다.

또한 본 연구에서 제시한 신기술 탐색 방법론은 다음과 같이 활용될 수 있다. 첫째, 기존에 연구된 바 있는 text mining, bibliometrics, technometrics, information visualization, KDD(knowledge discovery in database) 등의 기법들과 통합하여 국가연구개발 프로젝트의 연구기획 사전 타당성 분석에 활용될 수 있을 것이다. 둘째, 본 연구에서 개발된 방법론을 활용하여 도출된 결과를 기반으로 기획예산처 및 국가과학기술위원회가 신규 대형연구개발사업의 기획보고서 타당성을 평가함에 있어 합리적이고 객관적인 의사결정을 하는데 도움이 될 수 있을 것이다. 셋째, 향후 지속적인 연구를 통해 과학기술정보를 이용한 다양한 정량지표를 도출하고 이를 시스템화할 수 있는 시점에서는 이들을 통합하여 궁극적으로는 국가연구개발사업 타당성 분석 시스템을 구축할 수 있을 것이다.

본 연구와 관련된 향후 연구를 위한 몇 가지 제언을 하면 다음과 같다. 첫째, 향후 연구는 본 연구에서 도출된 결론이 이론적·실증적 연구들에서 예상된 결과와 모순되는 부분을 조사하는 것에서 시작할 수 있다. 둘째, 본 연구에서 제안된 방법론을 보다 확실히 검증하기 위해서는 실제 사례를 통한 추가적인 실증연구가 필요하다.

참고문헌

문영호 (2004), 기술진보의 모니터와 조기경보, *연구기획평가연구회 1차 워크샵*, KISTI

오세정·이달곤 (2004), 과학기술혁신체제의 재구축과 전개방향, *국가과학기술혁신체제 개편에 관한 공청회*

윤문섭·오해영·이우형·박각로·박상진 (2004), *국가연구개발의 전략기획을 위한 새로운 연구기획 방법론 개발*, 과학기술정책연구원

Barabasi, A. L. (2002), Statistical Mechanics of Complex Networks, *Reviews of Modern Physics*, **74**, 47-97

Callon. M., Courtial, J. P., & Turner W. (1986), *Future developments*, In M. Callon, J. Law, & A. Rip(Eds.), Mapping the dynamics of science and technology : Sociology of science in the real world, London : The Macmillan Press Ltd

Crane, D. (1972), *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, Chicago, IL: University of Chicago Press

Gorkova, V.I. (1988), Informetrics, *Informatics*, VINITI, Moscow

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press

Prichard, A. (1969), Statistical bibliography or bibliometric ?, *Journal of Documentation*, **24**, 348-349

Price, D. D. (1961), *Science since Babylon*, New Heaven: Yale University Press

Thagard, P. (1992), *Conceptual Revolutions*, Princeton, NJ: Princeton University Press