

# 국가적 차원의 유망연구영역 탐색: Scopus 데이터베이스를 이용한 과학계량학적 접근

## Identification of Emerging Research at the national level: Scientometric Approach using Scopus

여운동\* · 손은수\*\* · 정의섭\*\*\* · 이창환\*\*\*\*

Woon-Dong Yeo · Eun-Soo Sohn · Eui-Seob Jung · Chang-Hoan Lee

### 차 례

- |            |         |
|------------|---------|
| 1. 서론      | 4. 연구결과 |
| 2. 선행연구 검토 | 5. 결 론  |
| 3. 연구 방법론  | • 참고문헌  |

### 초 록

급속도로 변화하는 과학기술 환경 속에서 기업들은 현존하는 기술의 발전을 모니터링함과 동시에 새롭게 부상하는 유망기술을 찾아야만 경쟁력을 가질 수 있다. 각 국가에서는 경쟁적으로 유망연구영역을 찾고 있는데, 대부분의 국가에서는 전문가 평가를 기반으로 한 델파이 방법으로 유망기술을 탐색하고 있다. 그러나 델파이와 같은 전문가방식은 기본적으로 전문가의 주관적 판단에 의지하기 때문에 편향성과 이에 따른 문제가 발생한다. 본 연구에서는 델파이 방법의 기술예측방법이 가지는 문제점을 개선하기 위해 과학계량학적 방법으로 유망연구영역을 탐색하였다. 탐색과정에서는 다음 3가지의 특별한 노력을 통해 과학계량학적 방법의 성능을 제고하고자 하였다. 첫째, 데이터베이스 선정에 있어서 공학 등 저널수가 적은 기술분야에서도 유망연구영역의 도출이 가능한가를 고려하였다. 둘째, 과학기술 분야를 대상으로 하는 분석에서 과학기술분야별로 가지고 있는 인용수의 차이로 인해 발생하는 문제점을 해결하고자 부분인용계상(fractional citation counting)과 이동평균을 이용한 선형회귀분석을 도입하였다. 셋째, 과학계량학적 분석으로 나온 결과를 정성적으로 검증하여 과학계량학적 방법에 의한 오류를 최소화 하였다. 최종 290개의 유망연구영역을 선정하였으며, 각 영역은 기술의 네트워크상에서 가시화하였다. 본 연구에서는 Scopus 데이터베이스가 사용되었으며, 데이터마이닝과 가시화에는 한국과학기술정보연구원에서 개발한 KnowledgeMatrix가 사용되었다.

### 키 워 드

과학계량학, 서지정보학, 데이터마이닝, 유망연구영역

- \* 한국과학기술정보연구원 계량정보연구팀 선임연구원  
(Senior Researcher, Information Research Team, KISTI, wdyeo@kisti.re.kr)
- \*\* 한국과학기술정보연구원 계량정보연구팀 선임연구원  
(Senior Researcher, Information Research Team, KISTI, essohn@kisti.re.kr)
- \*\*\* 한국과학기술정보연구원 경인지원 책임연구원  
(Principal Researcher, Gyeongin Regional Office, KISTI, esjing@kisti.re.kr)
- \*\*\*\* 한국과학기술정보연구원 계량정보연구팀 책임연구원  
(Principal Researcher, Information Research Team, KISTI, chereel@kisti.re.kr)
- 논문접수일자 : 2008년 6월 5일
- 게재확정일자 : 2008년 9월 22일

## ABSTRACT

In today's environment in which scientific technologies are changing very fast than ever, companies have to monitor and search emerging technologies to gain competitiveness. Actually many nations try to do that. Most of them use Delphi approach based on experts review as a searching method. But experts review has been criticised for probability of inclination and its derivative problems in the sense that it is accomplished only by expert's subjectivity. To overcome such problems, we used Scientometric Method for identifying emerging technology that had been done by Delphi as a rule. We made three particular efforts in order to improve the Quality of the result. Firstly, we selected one alternative database between SCI and Scopus hoping to see evenly-distributing results in wide fields on the front burner. Secondly we used Fractional citation counting in counting citation number in the stage of linear regression analysis. Lastly, we verified Scientometric result with experts opinions to minimize probable errors in a Scientometric research. As a result, we derived 290 emerging technologies from Scientometric analysis with Scopus Database, and visualized them on 2-dimension map with data mining system named KnowledgeMatrix which was developed by KISTI.

## KEYWORDS

KnowledgeMatrix, Scientometric

## 1. 서론

급속도로 변화하는 과학기술 환경 속에서 기업들은 현존하는 기술의 발전을 모니터링함과 동시에 새롭게 부상하는 유망기술을 찾아야만 경쟁력을 가질 수 있다(Wheatley and Wilemon 1999). 이 사실을 직시하고 있는 각 국가에서는 경쟁적으로 미래를 선도할 유망연구 및 기술을 찾기 위해 국가적인 사업을 추진 중이다. 이들 대부분의 국가에서는 전문가 평가를 기반으로 한 델파이 방법으로 유망기술

을 탐색하고 있다(Martin 1995). 그런데 문제는 전문가(동료) 평가는 기본적으로 전문가의 주관적 판단에 의지하기 때문에 편향성과 이에 따른 문제가 존재한다는 것이다(Chubin and Hackett 1990; Kostoff 1994). 예를 들면 기존분야를 유지하려고 하는 경향이나 개인적·조직적 편향, 유명한 연구자나 기관에 높은 평점을 주게 되는 후광효과, 평가자마다 다른 평가기준 등이 있을 수 있다. 전문가 평가는 특정분야의 전문가로 구성되기 때문에 학제연구의 질의판단이나 자신의 분야를 넘어

서는 비교는 어려울 수 있다. 더욱이 한정된 수의 평가대상에 기초해 비교를 해야 하며, 대체로 평가자에게 할당되는 작업량도 크다. 한국에서도 2005년 6월 '미래국가유망기술위원회(과기부)'를 구성하여 '과학기술예측조사(2005-2030)' 결과에서 도출된 유망기술후보군을 바탕으로 『국가미래 국가유망기술 21』을 선정하여 발표한 바 있는데(한국화학공학회 2005), 이 역시 델파이 방법으로 수행되었다.

본 연구에서는 델파이 방법의 기술예측이 가지는 문제점을 개선하기 위해 과학계량학을 이용하여 한국의 미래 과학기술 유망연구영역을 탐색하였다. 그러나 여기에서 주지해야 할 사실은 과학계량학만을 사용하는 평가는 분명히 문제점을 안고 있다는 것이다. 특히 최근 연구성과의 질적인 측면의 중요성을 강조하면서 '인용'이 널리 사용되고 있는데(Cole 2000; Moed 2005), 인용 목적의 상이성(Small 1982; Liu 1993; Case and Higgins 2000)과 분야별 인용수 차이(Egghe and Rousseau 1990)의 문제가 제기되는 등 '인용'이 평가지표로서 적합한가에 대한 논란이 계속되고 있다. 본 연구에서는 과학계량학이 가지는 한계를 극복하기 위하여 다음 3가지의 특별한 노력을 하였다. 첫째, 데이터베이스 선정에 있어서 공학 등 저널수가 적은 기술분야에서도 유망연구영역의 도출이 가능한가를 고려하였다. 둘째, 과학기술 전 분야를 대상으로 하는 분석에서 과학기술분야별로 가지고 있는 인용수의 차이로 인해 발생하는 문제점을 해결하고자 부분인용

계상(fractional citation counting)과 이동평균을 이용한 선형회귀분석을 도입하였다. 셋째, 과학계량학적 분석으로 나온 결과를 정성적으로 검증하여 과학계량학적 방법에 의한 오류를 최소화 하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 과학계량학(scientometrics)을 적용한 유망연구 및 기술의 선정 사례를 국내외적으로 살펴본다. 3장에서는 유망연구탐색 모델(프로세스)에 대해 설명한다. 우선 전체 프로세스에 대해 기술하고, 다음으로 각 모듈별로 상세하게 기술한다. 4장에서는 유망연구 탐색결과를 기술하는데, 지면 제약의 이유로 인해 연구결과의 일부만 나타내었다. 마지막으로 5장에서는 결과의 검증에서 연구가 가지는 한계와 의미에 대해 논한다.

## 2. 선행연구 검토

과학(Science)의 계량(Metrics)을 뜻하는 과학계량학은 1983년 발표된 Martin and Irvane의 논문에서 처음으로 연구평가에 응용된 후로, 최근에 컴퓨팅 기술의 발달과 함께 더욱 다양한 방향으로 연구가 시도되고 있다. 유망연구를 탐색하는 분야에서도 과학계량학이 응용된 사례가 몇몇 보고되고 있다. 그 중 가장 대표적인 것은 일본 NISTEP(National Institute of Science and Technology Policy)에서 수행된 연구다. 일본은 전통적으로 델파이

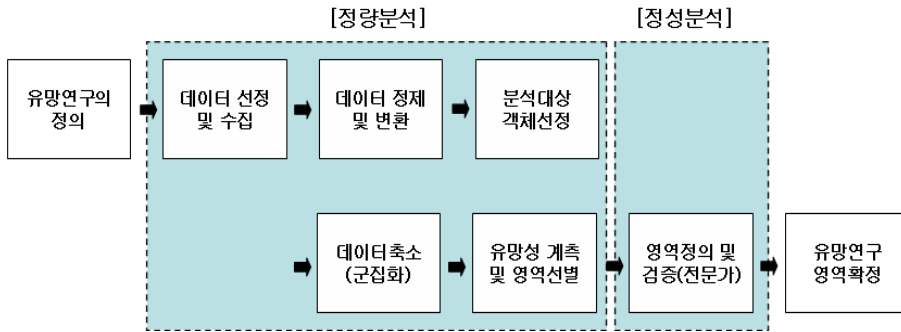
이 조사 방법으로 미래기술을 예측하고 있는데, 2005년에는 과학계량학을 활용한 결과(Study on Rapidly Developing Research Areas)를 델파이 기법을 위한 기초데이터로 사용하였다(NISTEP 2005; Okuwada 2008). 이 연구에서는 Small(1997)이 사용한 것과 같은 “humpty-dumpty”라는 블록모델(Block-model) 기법을 사용하여 동시인용분석을 수행하였다. Small의 방식과 다른 점은 1차 군집화 후 회귀모델로 분석하여 우수한 특성을 보이는 군집들만을 가지고 2차 군집화를 한다는 것이다. Bengisu and Nekhili(2006)는 터키에서 델파이 기법으로 나온 Turkey's Vision 2023(TUBITAK 2004) 기술예측 결과 중 “Machine and Materials”에 속하는 20개의 기술에 대해 과학계량학적으로 유망성을 검증하였다. 연도별 논문 및 특허의 발생 빈도와 이들의 연도별 유사성, Gompertz와 logistic 성장곡선 예측법이 사용된 검증에서 20개의 기술 중 적어도 5개의 기술에 대해서는 델파이 기법과 다른 결과(유망성 부적합)를 보여, 과학계량학의 필요성을 증명하였다. 이 밖에도 Kuusi and Meyer(2007)는 Nantero Inc.가 보유하고 있는 특허에 대해 서지결합 분석으로 NRAM 기술의 예측연구를 수행하였으며, Miranda Santo 등(2006)은 브라질의 Center for Management and Strategic Studies(CGEE)에서 수행된 예측연구에 대한 보고로써 과학계량학과 전문가 방법 병행에 대한 중요성을 강조하고 있다.

한국에서는 정보통신연구진흥원(IITA)에서 구축한 논문과 특허 데이터를 분석하는 레이더스(RADERS: Research Area DEtection through R&d information Scanning)가 과학계량학적 기술예측의 대표적인 사례이다(정보통신진흥연구원 2007). 2007년에 구축된 이 시스템은 미래유망기술 발굴, 연구자간 공동연구 추진현황, 전문가 발굴, 인용지수(impact factor)에 의한 질적 평가, 인용도 분석 등 추적평가 등에 적용하고 있다. 이 시스템에서 핵심을 차지하는 미래유망기술 발굴 프로세스는 동시단어분석을 기반으로 한다.

### 3. 연구 방법론

#### 3.1 전체 프로세스

과학계량학을 이용한 유망연구영역 탐색 프로세스는 <그림 1>과 같다. 연구에서 가장 선행되는 것은 유망성에 대한 개념을 정립하는 것이다. 다음으로 그러한 ‘유망한 연구영역’을 찾기 위해 사용할 데이터 셋을 선정하고 수집하여 분석 가능하도록 정제, 변환을 한다. 기본적인 데이터 변환이 완료되면 분석대상이 되는 객체를 선정한다. 여기에서 객체라 함은 키워드나 문헌과 같이 분석결과로 도출되는 요소를 일컫는다. 동시단어분석에서는 키워드가 객체가 되고, 저자동시인용분석에서는 저자가 객체가 되는 것이다.



〈그림 1〉 유망연구영역 탐색 프로세스

객체가 선정되면 그 객체를 대상으로 데이터 축소를 하게 된다. 과학계량학의 실현을 위한 데이터마이닝은 대규모의 데이터에서 중요한 정보를 추출하는 작업으로 전체 과정을 데이터를 축소해가는 과정이라 할 수 있다. 따라서 데이터축소의 작업이 정량분석에서 가장 중요한 작업이라 할 수 있다. 데이터축소 알고리즘은 여러 가지가 있는데 본 연구에서는 군집화 알고리즘을 사용한다. 군집화 결과는 개별 논문들이 군집의 형태로 나오게 된다. 이 군집을 본 연구에서는 연구영역이라고 칭하고 있으며, Morris(2005)는 “Research Fronts”라고 정의하여 그 중요성을 부각하였다.

군집화를 통해 도출된 연구영역은 영역들의 유사성과 유망성이 검증되지 않은 영역이다. 그래서 이 영역들을 본 논문에서는 유망연구 영역후보군이라고 정의하였다. 유망연구영역 후보들은 몇 가지 통계 기법으로 그 유망성을 판단할 수 있다. 마지막 단계에서는 계량분석을 통해 나온 정량적인 결과를 전문가를 활용하여 정성적으로 검증함으로써 계량분석이 가

지는 근본적인 약점을 보완하고자 하였다. 최종 확정된 유망연구영역은 영역별로 기술구성을 찾고 세부기술 간의 관계를 쉽게 파악할 수 있도록 2차원 매핑 알고리즘을 사용하여 가시화 하였다.

### 3.2 유망연구의 개념

유망연구의 개념은 널리 사용되고 있음에도 불구하고 정의가 제대로 내려져 있지 않다. 최근 일부 연구자들에 의해 이러한 개념을 정리하고자 하는 시도가 이루어지고 있다. 이러한 연구자들의 시도를 토대로 부상기술이라는 용어가 가지는 개념을 본 연구에서는 다음과 같이 정의하였다(Park et al. 2007).

유망연구의 개념에서 가장 먼저 언급될 것은 “최근 들어 급격한 성장을 보이는 영역”이라는 개념이다(Poter et al. 2002; Corrocher et al. 2003). 이는 신·구 연구를 가능하는 잣대로 사용될 수 있을 뿐만 아니라 최근 들어 부상하는 연구를 파악하는 데 유용한 개념이

다. 또한 이 개념은 어떤 정량적인 방법을 통해 유망연구를 도출할 지에 대한 방법을 결정하는데 유용한 정보를 제공한다.

유망연구는 유망기술을 견인할 수 있는 연구이다. 유망연구는 기존과 비교하여 다른(또는 새로운) 분야로의 전이나 변화를 의미한다(Garfield 1993; Hung and Chu 2006). 이러한 전이나 변화는 동태적(dynamic)인 것으로서 변화가 멈추었다면 더 이상 유망기술로 보기 어렵다. 따라서 특정 시점에서의 유망연구는 그 시점에서 전이 또는 변화하는 과정에 있어야 한다. 이러한 전이나 변화는 점진적(incremental)일수도 있으며 급진적(radical)일 수도 있다.

변화나 전이는 흔히 잠재력이라는 개념과 연관되곤 한다. 이는 이러한 변화나 전이가 기존의 것과 비교하여 경쟁력을 가지는 새로운 영역을 창조하는 것을 의미하기 때문이다. 따라서 유망연구는 현 시점에서는 파악하기 어려울 수도 있으나 시장성 또는 경제적 잠재력을 지니고 있게 된다(Day and Schoemaker 2000; Porter et al. 2002; hung and Chu 2006).

유망연구의 마지막 특징은 과학에 대한 의존도가 점차 높아진다는 점이다(Day and Schoemaker 2000). 이미 선행연구에서도 지적되었듯이 사회·경제적으로 큰 영향력을 발휘할 수 있는 기술개발 여부는 보유하고 있는 과학적 지식의 선도성에 영향을 받기 때문이다. 또한 나노 또는 바이오와 같은 영향력이 큰 분야

에 있어서 기초과학에 대한 의존도가 높다는 점은 익히 잘 알려진 사실이다.

### 3.3 데이터 선정 및 수집

과학계량학의 대상이 될 수 있는 과학기술 문헌 데이터베이스는 다수가 있다. 그 중에서 과학계량학이나 계량문헌학에서 자주 사용되는 것이 SCI(Science Citation Index) 데이터베이스이다. 명칭에서 알 수 있듯이 인용(citation)에 관한 정보를 얻을 수 있는 과학기술문헌 데이터베이스라는 점과 과학기술 전반을 대상으로 하고 있는 점, 수집 대상이 되는 저널의 선택이 타당하다는 점 등에 의해 과학계량학에 있어 기본적인 도구(bibliographic instrument)로서 뿐만 아니라 서지적 레퍼런스를 이용하여 과학적 그림을 만들어 내는 도구로서도 인정받고 있다(Wouters 1999).

인용정보를 제공하는 또 다른 데이터베이스로 2004년 개설된 Scopus가 있다. Scopus는 SCI에 비해 더 많은 저널을 포함하고 있으며 인용정보와 함께 저자의 소속기관 파악이 가능하게 되어 있다. 분석의 편의성에 있어서도 분석모집단을 설정하고 이들을 인용하는 논문들을 연도별로 리스팅할 수 있으며, 인용논문들만 다운받아 분석에 이용할 수도 있다. 또한 외적인 측면에서 뿐만 아니라 질적인 측면에서도 평가용 데이터베이스로서 SCI를 대체할 수 있는 Scopus의 가능성이 비교연구를 통해

발표되고 있다(Norris 2007).

본 연구에서 SCI와 Scopus 데이터베이스를 선택함에 있어서 가장 중요하게 고려한 사항은 데이터베이스의 “분류” 체계였다. 연구분야별 저널수 편중은 바이오분야와 같은 일부 영역에 다수의 유망연구영역이 집중되어 나타나는 문제점을 발생시키므로 기술을 세분화하여 표본집단을 설정할 필요성이 있다. SCI는 171개로 논문들을 소속 저널분류에 따라 분류하고 있는 데에 반하여 Scopus는 294개로 분류하고 있다. 이 점이 Scopus를 더 선호하게 만든 이유이다.

본 논문에서는 피인용율 상위 1%에 해당하는 논문집단인 고인용논문(highly cited papers)을 분석을 위한 기초데이터로 활용하였다. 고인용논문은 탁월한 성과를 보이는 과학적 연구를 모니터링하고 찾아내기 위한 잠재적인 대표집단으로 여겨지고 있다. 최근에 European Commission에서 이뤄진 벤치마킹 연구에서 유럽나라들의 연구성과를 측정하기 위한 지표로서 고인용논문이 사용되기도 하였다(European Commission 2001). 고인용논문은 연구그룹의 사례연구에서 지표로 사용되기도 했으며(Martin and Irvine 1983), 세계적으로 우수한 연구를 찾는 유용한 지표로서 사용될 수 있다고 보고되고 있다(Tijssen et al. 2002).

본 논문에서는 Scopus를 이용하여 최근 5년간(2002~2006년)에 걸쳐 각 연도별, Scopus 데이터베이스의 294개 분야별 피인용율이 상위 1%에 속하는 고피인용논문 7만2,299편과 이를 인용하는 147만250편의 논문을 취하여 정보분석을 실시하였다(〈표 1〉 참조).

### 3.4 데이터 정제 및 변환

데이터의 불완전성, 잡음, 일치성의 부재는 대규모 데이터베이스에서 흔히 발생하는 문제이다. 논문의 서지정보를 데이터베이스에 입력하는 과정에서 발생하는 정보누락이나 기입 오류는 논문데이터베이스의 주된 정제 이유가 된다. 논문데이터베이스를 활용하는 계량정보분석에서는 동음이의어나 이음동의어 등의 처리도 필요하다. 논문 저자를 분석할 경우는 동음이의어나 동의어와 마찬가지로 고유식별자를 사용하여 저자별로 구별이 가능하도록 해주어야 한다. 국내외적으로 정량적인 성과평가의 중요성이 대두됨에 따라 저자구분을 자동 혹은 반자동으로 수행하려는 연구들이 지속적으로 이루어지고 있다. 저자소속기관의 정제에서도 여러 지역에 분포하는 연구소나 계열기관을 어느 정도까지 같은 기관으로 봐야 할지에 대한 명확한 기준이 있어야 한다. 국가별 연구동향

〈표 1〉 데이터셋 현황 (연도별 논문수)

연도	2002	2003	2004	2005	2006
논문수	13,075	13,273	14,144	15,638	16,135

의 파악을 위해서는 해당논문에 대한 국가별 기여도를 할당해야 하는데, 이것은 주저자의 소속국가를 기준으로 하거나 교신저자를 기준으로 하는 것이 일반적이지만, 선행연구들을 바탕으로 한 새로운 지표를 적용할 수도 있다.

계량정보분석에는 정형화된 데이터와 비정형화 데이터를 사용할 수 있다. 비정형화 데이터의 경우 정형화 과정이 수반되어야 한다. 정형화 과정에서는 경우에 따라서는 자동형태소 분석과 같은 자연어처리 기법을 활용하기도 한다. 정형화된 데이터도 데이터분석에 활용되는 소프트웨어가 인식할 수 있는 형태로 변환을 해 주어야 한다. 본 연구에서는 한국과학기술정보연구원에서 개발한 KnowledgeMatrix를 사용하여 정량분석을 수행하였다. KnowledgeMatrix는 최종분석에서 연구영역별 세부 기술의 네트워크 가시화에도 사용된다. KnowledgeMatrix는 국내외 유수의 특허논문데이터베이스에 대해 데이터 입력이 가능하도록 설계되어 있으며, 필요시에는 사용자가 입력 형태를 간단히 설정할 수 있도록 되어있다. 본 연구에서는 Scopus 데이터베이스를 XML 형태로 수집하였기 때문에 서지정보를 KnowledgeMatrix가 인식할 수 있도록 텍스트형태로 변환하는 작업을 수행하였다.

### 3.5 분석대상 객체선정

유망연구영역선정을 위한 계량정보분석의 객체로는 키워드와 논문, 이 두 가지를 고려할

수 있다. 최종결과물이 키워드의 집합으로 표현되는 경우는 필드의 지식 구조로 되어 있어 논문의 집합으로 표현되는 경우에 비해 직접적인 해석이 가능하다는 장점이 있다(Bauin 1991; Callon et al. 1991). 즉, 논문집합은 전문가를 활용하여 그 논문들이 의미하는 연구가 무엇인지 해석을 해야 하나 키워드의 집합은 가장 빈도가 높은 키워드를 선택하는 방법 등을 이용하여 직접적으로 영역의 정의가 가능하다. 그러나 대규모 데이터를 가지고 과학기술 다분야에 대한 계량정보분석에서는 키워드를 활용한 분석의 신뢰성이 떨어지게 된다. 그 이유는 키워드 분석에서는 색인자효과(indexer effect)가 발생할 뿐만 아니라(Whittaker 1989) 키워드 정제 작업이 가장 중요한데, 대규모 데이터의 키워드에 대해 기준을 설정하고 키워드를 정제하는 것은 거의 불가능하기 때문이다. 특히 동일한 철자를 가진 단어가 분야별로 다른 의미로 사용되는 경우가 빈번하다. 따라서 본 연구에서는 논문을 분석대상 객체로 선정하였다. 동시발생을 이용하여 논문을 분석하는 경우에도 키워드와 인용을 활용하여 분석하는 것이 가능한데 키워드의 경우 위에서 설명한 것과 같은 이유로 대규모 데이터에서는 적절치 않는 분석기법이 된다.

### 3.6 군집화

인용패턴의 유사성으로 논문의 군집관계를 살펴보는 방법에는 크게 공인용분석(Small



1973)과 서지결합분석(Kessler 1963)의 2가지 기법이 있다. 공인용분석과 서지결합분석 간의 우수성에 대한 논쟁은 Small이 공인용 분석을 제안함과 동시에 시작되었다. Glanzel 등(1996)은 공인용에서는 출판직후의 분석이 불가능한 것에 비해 서지결합에서는 가능하므로 최첨단 분야의 분석에서는 서지결합법이 이용될 가능성은 크다고 주장하였으며, 이것은 이미 Hicks(1987)가 공인용을 위해서는 많은 인용이 필요하다는 이유로 공인용 분석의 문제점을 제기한 것과도 일치한다. Kuusi and Meyer(2007)도 이와 같은 이유에서 공인용분석보다 서지결합법을 선호한다고 밝히고 있다. Small(2003)은 2002년에 공인용분석에 의해 군집화된 4,270개의 연구영역의 2달간 변화를 조사하였는데, 새롭게 나타난 영역이 521개, 크기가 증가한 영역이 515개, 크기가 감소한 영역이 923개 등 영역에 변화가 있는 것이 총 1,969개로 45.9%에 달하는 것으로 나타났다. 중장기 유망연구영역의 선정이라는 관점에서는 이렇듯 변화무쌍한 공인용 분석은 오히려 단점으로 작용한다. 이에 본 논문에서는 서지결합법을 이용하며, 그 연구기간을 일정기간으로 절단하고 통계적 분석을 사용하여 영역들을 선별하고자 하였다.

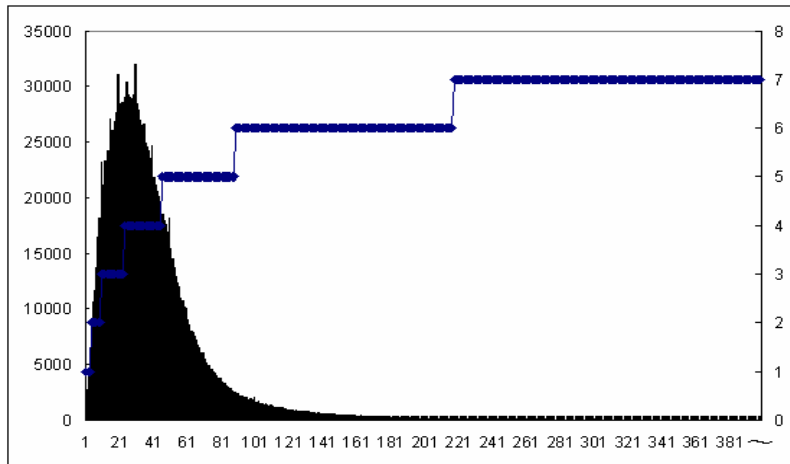
논문 간의 유사도를 측정하기 위해서는 유사도 계수를 사용하는데, 사용용도에 따라서 다양한 유사계수 공식이 제시되어 있다. 모든 유사계수가 공통적인 특성을 나타내므로, 군집화에 있어 어떠한 유사계수를 선택하는가는

문제가 되지 않는다고 인식되고 있다(Salton and McGill 1983). 그러나 실제로 둘 이상의 유사계수를 이용한 실험적 연구에서는 이용된 유사계수에 따라 결과가 달라지는 현상이 나타나기도 했다. 인용분석에서는 Cosine 계수(Salton 1989)가 주로 사용된다.

군집화는 데이터간의 “유사성”을 바탕으로 상대적인 “거리”를 계산하여 유사한 것끼리 묶어주거나 떼어내는 방법이다. 군집화에는 “거리”를 정의하는 방법, 즉 상대적인 거리를 측정하는 방법에 따라서 여러 가지 알고리즘이 존재한다. 공인용분석과 서지결합분석에 있어서 흔히 사용하는 군집화 알고리즘은 Hierarchical 방식인 Single-linkage와 Ward's method이다. Single-linkage는 유사도 매트릭스(dissimilarity, distance)를 활용하여 항목 간 최단거리를 기준으로 군집화하는 방식으로, 계산알고리즘이 단순하여 대규모 데이터 처리에 자주 이용된다. 본 연구에서는 두 논문의 유사성을 정의하기 위해 코사인 계수가 0.2미만인 경우는 유사도를 0으로 설정하였으며, Single-linkage에 발생하는 연쇄(chaining) 문제를 없애기 위하여 영역의 최대 크기를 논문 50개로 제한하였다.

### 3.7 유망성 계측 및 영역선별

군집화를 통해 도출된 영역들은 인용수가 많은 고인용논문의 모임이다. 인용수는 분야마다 다르고 세부기술별로 연구하는 연구자



〈그림 2〉 인용논문의 레퍼런스와 가중치

수에 따라 차이가 난다. 예를 들어 화학이나 물리 등의 분야에는 수학 분야에 비하여 동일한 기간 동안 훨씬 많은 수의 논문이 발표된다. 또한 화학이나 물리 분야의 논문에는 논문당 20개 내외의 참고문헌이 기재되어 있는 반면 수학 분야의 논문은 대부분 10개 내외이다 (Egghe and Rousseau 1990). Small(1985)은 다학문 분야에서 고르게 분포하는 연구영역을 도출하기 위하여 부분인용계상방식을 사용하여 고인용논문을 선정하였다. 부분인용계상은 논문이 가지고 있는 피인용수를 그대로 인정하는 것이 아니라, 그 논문을 인용하고 있는 논문들이 가지는 레퍼런스의 역수만큼만 인정하는 것이다. 예를 들어 A라는 논문을 B (레퍼런스 수 N)와 C(레퍼런스 수 M)라는 논문이 인용하고 있으면, 종래의 방식으로는 A의 피인용수는 2라는 단순한 계산으로 도출이 가능하다. 하지만 부분인용계상에서는 B와 C

의 레퍼런스 수를 고려하여 A의 피인용수를  $(1/N+1/M)$ 로 계산하게 된다.

본 연구에서는 서지결합법을 통해 도출된 유망연구영역 후보군 중에서 최종 유망연구영역을 선별할 때, 회귀분석 과정에서 이용된 피인용수 계산에 부분인용계상방식을 적용하였다. 본 연구에 사용된 고인용논문 7만2,299편을 인용하고 있는 147만349편의 인용논문에 대한 레퍼런스 수의 분포는 〈그림 2〉와 같다. 3,730개의 레퍼런스가 가장 많은 수치이며, 최소 1편의 논문을 인용하고 있다. 논문들 중에서 30개의 레퍼런스를 가지는 논문이 3만 1,955개로 가장 많은 분포를 차지하고 있다.

그런데 이와 같은 부분인용계상방법에서는 레퍼런스 수가 1개인 것의 가중치가 1인 반면에 레퍼런스 수가 10개 것의 가중치는 0.1로 계산되어 가중에 편차가 과도하게 커지게 된다. 이것은 마태효과(matthew effect)에 의

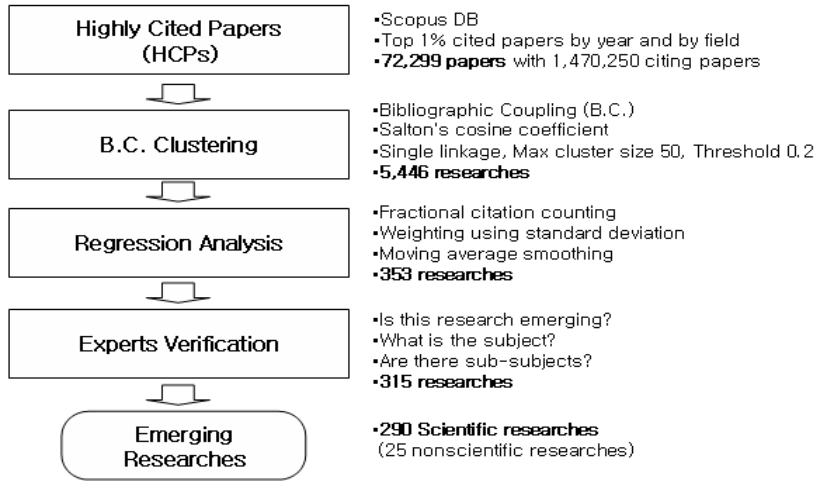
해 인용이 가지는 분포의 편차(skew)가 실제 논문의 질(quality) 차이에 비해 크게 나타나는 것이다(Akesnes 2003). 본 논문에서는 레퍼런스의 분포를 계산하여(〈그림 2〉 참조) 표준편차 구간으로 계급화하여 가중치를 결정함으로써 부분인용계상방법을 보완하고자 하였다. 분포의  $-2.5\sigma$  이하는 1,  $-2.5\sigma \sim -1.5\sigma = 2$ ,  $-1.5\sigma \sim -0.5\sigma = 3$ ,  $-0.5\sigma \sim 0.5\sigma = 4$ ,  $0.5\sigma \sim 1.5\sigma = 5$ ,  $1.5\sigma \sim 2.5\sigma = 6$ ,  $2.5\sigma$  이상은 7로 계급화 하였다. 이로써 레퍼런스가 80개인 논문이 인용한 경우는 가중치를 1/5로, 레퍼런스가 300개인 논문의 인용가중치는 1/7과 같이 적용된다.

본 논문에서는 논문들의 성장속도를 예측하기 위하여 선형회귀 방정식을 이용하였다. 논문발표수는 저널의 발행기간이 정해져 있고, 기술적 이슈 발견과 연동하여 자기상관을 가질 수 있다. 그러나 이러한 계절 혹은 계열적인 요인을 배제하면 간단한 선형회귀만으로도 중기적 관점의 예측이 가능하다. 선형회귀에서는 데이터가 직선으로 모델링된다. 선형회귀는 가장 단순한 회귀분석이다. 이변량 선형회귀모형은 랜덤변수  $Y$ 를 다른 랜덤변수  $X$ 에 대한  $Y = \alpha + \beta X$  같은 선형함수로 모델링한다. 여기에서  $Y$ 의 분산은 상수로 가정하며,  $\alpha$ ,  $\beta$ 는 각각  $Y$ 절편과 직선의 기울기를 나타내는 회귀계수이다. 본 연구에서는  $x$  항목은 연도를 의미하고,  $y$ 는 피인용수를 의미한다. 그리고 연구영역간의 상대적인 발전가능성의 비교만을 목적으로 하였기 때문에 회귀기울기

인  $\beta$ 에만 관심을 두었다. 여기에서 회귀기울기는 집단의 평균피인용수로 나눈 표준화된 값을 사용하였다. 피인용수는 부분인용계상과 함께 3기 이동평균하여 계절변동이나 불규칙변동을 최소화한 값이다.

### 3.8 영역정의 및 전문가 검증

앞서 우리는 유망연구영역을 시장성 혹은 경제적 잠재력을 지니고 있는 연구로 정의하였다. 유망성의 검증에서는 여기에 추가적으로 한국의 상황이 고려되어야 한다. 글로벌 연구에서 유망한 것이 결코 한국에서도 유망한 연구일 수는 없는 것이다. 그러나 정량분석만으로는 충분히 이를 반영하여 유망연구영역을 선정할 수가 없다. 뿐만 아니라 정량적 분석에 의해 도출된 연구영역은 논문의 집단이다. 이것은 자동주제발견(auto topic discovery)이나 다문서요약(multi-document summarization) 기법을 사용하여 어느 정도 묘사가 가능하나 영역의 정확한 정의를 위해서는 전문가들이 직접 명명하는 것이 가장 최선의 방법일 것이다. 따라서 본 논문에서는 유망연구영역 선정의 최종단계에서 한국과학기술정보연구원 소속 연구원, 대학교수, 퇴직과학자(www.reseat.re.kr) 등의 관련 분야 전문가들로 하여금 영역을 정의하게 하고, 그 영역이 전문가적 관점에서 한국의 미래 유망연구영역으로서 타당한지와 영역을 이루는 논문집단은 공통된 주제를 내포하고 있는지에 대해 검증하도록 하였



〈그림 3〉 연구결과 요약

다. 이 과정에서는 전문가의 주관적인 의견이 검증에 다소간 반영될 것이다. 이를테면 최종 결과에 나타나는 화성탐사와 관련된 연구가 한국의 미래 유망연구영역인가에 대한 검증은 오로지 전문가의 판단에 의존할 뿐이다. 그러나 본 논문에서는 이러한 부정적인 측면보다는 정량분석으로 인해 발생하는 문제점을 보완할 수 있는 긍정적인 측면이 더욱 클 것이라고 기대한다.

#### 4. 연구결과

본 논문에서는 Scopus 데이터베이스에서 최근 5년간(2002~2006), 연도별 294개 분야별 피인용율이 상위 1%에 속하는 고인용논문 7만2,299를 서지결합법으로 군집화하고, 최소 연결계수 0.2와 최대 클러스터 사이즈 50

을 적용하여 5,446개의 연구영역을 확보하였다. 다음 단계로 영역별로 인용하고 있는 논문에 대해 부분인용계상과 3기 이동평균을 적용하여 선형회귀분석을 수행하고 회귀기울기가 0.5 이상인 영역을 추출, 353개의 유망연구영역 후보군을 도출하였다. 353개의 후보군 중에서 전문가 검증에서 유망하지 않다고 판단된 38개 영역과 최다분류가 과학기술이 아닌 25개 영역을 제외하고 최종 290개의 과학기술 유망연구영역을 선정하였다(〈그림 3〉 참조).

본 논문에서는 영역별로 포함된 논문들을 분류(Scopus에서 제공하는 27개 분류)별로 합하여 가장 많이 나온 분류를 영역의 중심분류로 하여 〈표 2〉에 나타내었다. 의학이 49개 영역으로 16.4%를 차지하고 있으며, 다음으로 화학이 41개 영역 13.8%, 생화학, 유전학 및 분자 생물학가 36개 영역 12.1%, 물리학 및 천문학가 30 영역 10.1%로 상위 4개 분야

영역이 전체의 53.8%를 차지하는 것으로 나타났다. 기술분야별 분포의 편중은 과학기술 활동의 편중과도 상관이 있겠지만, Scopus 데이터베이스에 포함된 분류별 논문의 개수와도 밀접한 관계가 있을 것이다. 본 논문에서는 이러한 영향을 최소화하기 위하여 분류를 세분화하여 데이터를 수집하였고, 유망성 계량과 정에는 부분인용계상과 선형회귀분석을 사용하였다. 영역에 속하는 분류가 다 분류로 나타나기 때문에, 실제 영역의 분류별 분포는 <표 2>와 차이가 있을 가능성도 있다. 또한 두 세계 기술이 고르게 분포하는 경우 최근에 이슈가 되고 있는 융합기술(convergence technology)로 생각해 볼 수도 있으나(Okuwada 2008), 융

합기술에 대한 판단 기준의 확립과 선행연구 등이 필요하므로, 차후 연구로 미뤄 두고자 한다.

<표 3>은 290개 유망연구영역 중에서 예시로 논문 출판연도가 최근인 연구영역 10개를 나타낸 것이다. 본 논문에서는 지면상의 이유로 전체 도출된 연구영역들 중 일부만을 나타내었으며, 전체 리스트는 KISTI 미래 포털(miso.yeskisti.net)에서 제공하고 있다. <표 3>에서 핵심논문연구분야는 영역의 포함된 논문들의 각 분류를 합하여 상위 2개의 비율을 차지하는 분류를 나타낸 것이다. 핵심논문수는 영역에 포함된 논문의 수이며, 총피인용수는 핵심 논문을 인용하고 있는 전체 인용논문의 수인데, 이것은 인용논문간의 중복을 제거한 수치이다. 논문당 피인용수는 핵심논문이 가지고 있는 평균 피인용수를 말하며, 가중 피인용수는 부분인용계상을 적용하여 분포에 따라 가중치를 둔 논문당 피인용수이다. 핵심논문평균연도는 핵심논문들의 출판연도를 평균한 수치이다.

<표 3>에서 나타난 수치들은 영역들이 가지고 있는 특성을 파악하는 지표로 활용 할 수 있다. 예를 들어 논문당 피인용수가 높다는 것은 연구의 활용도가 높다는 것이며, 핵심논문평균연도가 최근일수록 연구가 최근에 형성되고 있는 것으로 해석하는 것도 가능하다. 본 논문에서는 여기에 추가적으로 영역별로 세부 연구가 어떻게 구성되어 있는지 파악하기 위하여 KISTI에서 개발한 KnowledgeMatrix를 사용하여 핵심논문들을 서지결합의 네트워크

<표 2> 유망연구영역의 분류별 분포

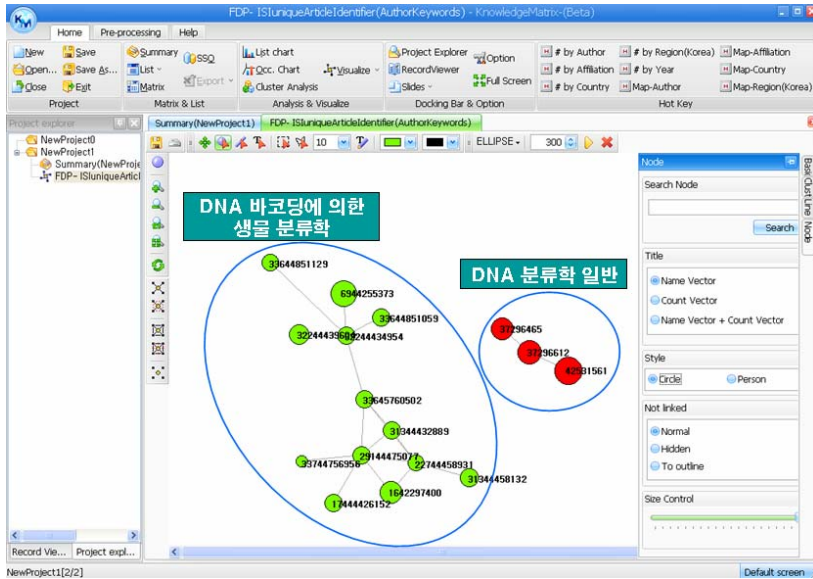
Scopus 중분류	영역수	비율
의학	49	16.4%
화학	41	13.8%
생화학, 유전학 및 분자 생물학	36	12.1%
물리학 및 천문학	30	10.1%
공학	24	8.1%
재료과학	21	7.0%
농학 및 생물학	17	5.7%
면역학 및 미생물학	14	4.7%
지구과학 및 행성과학	12	4.0%
화학공학	10	3.4%
환경과학	9	3.0%
에너지	7	2.3%
신경과학	7	2.3%
수학	5	1.7%
약물학, 독성학 및 제약학	5	1.7%
컴퓨터과학	3	1.0%

〈표 3〉 유망연구영역 예 - 논문 출판연도가 최근인 연구영역 상위 10

순위	클러스터 번호	핵심논문연구분야	연구영역명	핵심 논문수	총 피인용수	논문당 피인용수	가중 피인용수	핵심논문 평균연도
1	ERA C-120	생화학, 유전학 및 분자생물학(50%); 물리학 및 천문학(50%)	CuI 할로겐 화합물과 방향성 N-공여 리간드를 이용한 무기물질-유기물질의 중합체 합성	15	171	11.4	63.5	2005.7
2	ERA C-281	지구과학 및 행성과학(94.23%); 환경과학(5.77%)	전지구 모델을 이용한 기후 및 기후변화 연구	13	379	29.2	90.9	2005.6
3	ERA C-190	의학(100%)	내시경을 이용한 수술	9	217	24.1	68.4	2005.6
4	ERA C-079	물리학 및 천문학(60.17%); 수학(32.83%)	비선형 동역학(Nonlinear dynamics)	10	144	14.4	33.0	2005.6
5	ERA C-276	지구과학 및 행성과학(87.91%); 물리학 및 천문학(12.09%)	화성탐사선(Spirit와 Opportunity)에 의한 화성탐사 및 자료해석	13	184	14.2	42.4	2005.6
6	ERA C-134	생화학, 유전학 및 분자생물학(77.02%); 의학(15.99%)	MicroRNA 조절에 따른 생체 기능	49	2689	54.9	586.9	2005.5
7	ERA C-307	화학(45.83%); 생화학, 유전학 및 분자생물학(37.5%)	유기촉매를 이용한 알돌반응	8	178	22.3	38.1	2005.5
8	ERA C-072	물리학 및 천문학(81.63%); 공학(11.02%)	E-infinity 이론	35	677	19.3	203.4	2005.5
9	ERA C-184	의학(100%)	약물독성에 대한 지방유제주입 소생술	6	97	16.2	39.7	2005.5
10	ERA C-078	물리학 및 천문학(96.43%); 수학(3.57%)	블랙홀, 초끈이론, 게이지-중력 양면대칭성	7	264	37.7	61.2	2005.4

관계로 가시화하였다(〈그림 4〉 참조). 그런데 유망연구영역은 매우 밀집도가 높은 관계를 가지는 논문집단이므로 있는 그대로 가시화하게 되면 모든 논문들이 한 곳에 집중되어 네트워크의 가독성이 떨어지게 된다. 그래서 가시화에 앞서 PathFinder network( $q = n-1$ ,  $r = \infty$ ) 알고리즘을 적용하여 위상학적으로 중요하지 않는 관계를 제거하였다(Schvaneveldt

1989; Chen and Morris 2003). 〈그림 4〉에서 색깔은 클러스터를 의미하며, Knowledge-Matrix에서 연결의 강도에 대한 문턱값(threshold)에 따라 자동적으로 할당된다. 그리고 원의 크기는 핵심논문의 피인용수를 나타낸다. 그림에서는 하나의 연구영역이 두 개의 세부영역으로 구성되어 있는 것을 알 수 있다.



〈그림 4〉 연구영역 가시화 예  
 - DNA 인자를 바코드 시스템을 활용, 분석하여 생물종을 생태학적으로 분류

## 5. 결론

본 연구에서는 과학계량학을 이용하여 논문을 객관적으로 정량화하여 탐구적으로 유망연구영역을 발굴하고자 하였다. 그러나 이쉽게도 유망성이라는 것은 현재시점에서 결론 내리기 쉬운 문제가 아니다. 따라서 본 연구결과의 우수성을 비교연구를 통해 검증하는 것은 또 다른 어려운 과제가 될 것이다. 현 시점에서 가능한 연구는 텔파이 연구와의 정합성을 논하는 정도가 될 것이다. 그럼에도 불구하고 본 연구를 통해서 다음과 같은 몇 가지 확고한 결론을 얻을 수 있었다.

첫째, 본 연구에서와 같은 과학계량적 분석 프로세스가 체계적으로 정립될 경우 텔파이 방

식과 비교해 볼 때 유망연구영역 선정에 있어서 시간과 비용 절감의 효과를 기대할 수 있다.

둘째, 산업적 효과의 불확실성으로 인해 기존에 전문가 집단에 의해서는 부각되지 못했던 소외된 연구에 대해서도 유망성을 검증할 기회를 가질 수 있다.

셋째, 과학계량학적 결과물은 전문가 평가를 위한 보조지표로 뿐만 아니라 평가를 위한 주지표로 사용할 수 있다.

## 참고문헌

- 정보통신연구진흥원, 2007. 『Raius를 통한 IT 미래유망기술 발굴』. 대전: 동연구소.  
 한국화학공학회, 2005. 과학기술정책동향-과

- 학기술부, 『NEWS & INFORMATION FOR CHEMICAL ENGINEERS』, 23(5): 512-513.
- Aksnes D. W. 2003. "Characteristics of highly cited papers." *Research Evaluation*, 12(3): 159-170.
- Bengisu, M. and R. Nekhili. 2006. "Forecasting emerging technologies with the aid of science and technology databases." *Technological Forecasting and Social Change*, 73(7): 835-844.
- Bauin, S., B. Michelet, M. G. Schweighoffer and P. Vermeulin. 1991. "Using bibliometrics in strategic analysis: Understanding Chemical Reactions at CNRS." *Scientometrics*, 22: 113-137.
- Callon, M., J. P. Courtial, and F. Laville. 1991. "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry" *Scientometrics*, 22(1): 155-205.
- Case, D. O., and G. M. Higgins. 2000. "How can we investigate citation behavior? A study of reasons for citing literature in communication." *Journal of the American Society for Information Science*, 51(7): 635-645.
- Chen, C. and S. Morris. 2003. "Visualizing evolving networks: Minimum spanning trees versus Pathfinder networks." *Proceedings of IEEE Symposium on Information Visualization*, IEEE Computer Society Press, 67-74.
- Chubin D. E., and E. J. Hackett. 1990. "Peer review and the printed word." In: Chubin DE, Hackett E.J. *Peerless Science: Peer Review and U.S. Science Policy*. Albany, NY: SUNY Press.
- COLE, J. R. 2000. "A short history of the use of citations as a measure of the impact of scientific and scholarly work." In: B. CRONIN, H. B. ATKINS (Eds), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield*. Medford, NJ, USA, Information Today, 281-300.
- Corrocher, N., F. Malerba, and F. Montobbio. 2003. "The emergence of new technologies in the ICT field: main actors, geographical distribution, and knowledge sources." *Working Papers of Faculty of Economics Universit' degli Studi dell'Insubria*, 37 Varese, Italy.
- Day, G. S. and P. J. H. Schoemaker. 2000. "A different game. In G. S. Day &



- P. J. H. Schoemaker(Eds.), "Wharton on Managing Emerging Technologies, New York: John Wiley & Sons.
- Egghe, L., and R. Rousseau, 1990. "Introduction to informetrics, Quantitative Methods in Library." *Documentation and Information Science*, Amsterdam: Elsevier, 203-290.
- European Commission, 2001, Key Figures 2001, Special edition, Indicators for benchmarking of national research policies(Brussels).
- Garfield E, 1993, "Essays of an Information Scientist." ISI Press, 15: 75-83.
- Glanzel, W., and H. Czerwon, 1996. "A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level." *Scientometrics*, 37: 195-221.
- Hicks, D, 1987. "Limitations of co-citation analysis as a tools for science policy." *Social Studies of Science*, 17: 295-316.
- Hung, S. C., and Y. Y. Chu, 2006. "Stimulating new industries from emerging technologies: challenges for the public sector." *Technovation*, 26(1): 104-110.
- Kuusi, O, and M. Meyer, 2007, "Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm." *Scientometrics*, 70(3): 759-777.
- Kessler, M, M, 1963, "Bibliographic coupling between scientific papers." *American Documentation*, 14: 10-25.
- Kostoff, R, N, 1994, "Assessing research impact: US. government retrospective and quantitative approaches." *Science and Public Policy*, 2(1).
- Liu, M, 1993, "Progress in documentation: the complexities of citation practice: a review of citation studies." *Journal of Documentation*, 49(4): 370-408.
- Martin, B, R, 1995, "Foresight in Science and Technology." *Technology Analysis & Strategic Management*, 7(2): 139-168.
- Martin, B, R., and J. Irvine, 1983, "Assessing basic research. Some partial indicators of scientific progress in radio astronomy." *Research Policy*, 12: 61-90.
- Miranda Santo, M., G. Massari Coelho, D. M. dos Santos, and L. Fellows Filho, 2006, "Text mining as a valuable tool in foresight exercises: a study on nanotechnology." *Technological Forecasting & Social Change*, 1013-1027.

- Moed, H. F. 2005. Citation analysis in research evaluation, New York: Springer.
- Morris, S. A. 2005. Unified mathematical treatment of complex cascaded bipartite networks: The case of collections of journal papers, Ph.D. diss., University of Oklahoma.
- NISTEP, Japanese website. <<http://www.nistep.go.jp/index-j.html>>.
- Norris, M., and C. Oppenheim, 2007. "Comparing alternatives to the Web of Science for coverage of the social sciences' literature." *Journal of Informetrics*, 1-9.
- Okuwada, K. 2008. "Elements of Converting Technologies." *CACCI Journal*, 1: 1-6.
- Park, H. W., C. W. Lee and W. D. Yeo. 2007. "Development of Information Model to Identify Emerging Technologies." *Journal of Information Management*, 38(4): 1-21.
- Porter, A. L., D. Roessner, X. Y. Jin & N. C. Newman. 2002. "Measuring national emerging technology capabilities." *Science and Technology Policy*, 29(3): 189-200.
- Salton, G. 1989. Automatic text processing: the transformation, analysis, and retrieval of information by computer. Reading: Addison-Wesley.
- Salton G., and M. J. McGill 1983, Introduction to Modern Information Retrieval, McGraw-Hill, New York.
- Schvaneveldt, R. W., F. T. Durso and D. W. Dearholt. 1989. "Network structures in proximity data." *The psychology of learning and motivation: Advances in research and theory*, Academic Press, San Diego, Calif., 24: 249-284.
- Small, H. 1973. "Co-citation in the scientific literature: a new measure of the relationship between two documents." *Journal of the American Society for Information Science*, 24(4): 265-269.
- Small, H. 1982. "Citation context analysis." *Progress in Communication*, 3: 287-310.
- Small, H. 1997. "Update on Science Mapping: Creating Large Document Spaces." *Scientometrics*, 38(2): 275-293.
- Small H. 2003. "Paradigms, citations and maps of science: a personal history." *Journal of the American Society for Information Science and Technology*, 54(5): 394-399.
- Small, H., and E. Sweeney. 1985. "Clustering the Science Citation Index Using

- Co-Citations I. A Comparison of Methods.” *Scientometrics*, 7: 391-409.
- Tijssen R. J. W., M. S. Visser and T. N. van Leeuwen. 2002. “Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference?” *Scientometrics*, 54(3): 381-397.
- TUBITAK. 2004. Vision 2023 Technology Foresight Delphi Survey Final Report.
- Wheatley, K. K. and D. Wilemon. 1999. “From Emerging Technology to Competitive Advantage.” in Portland International Conference on the Management of Engineering and Technology (PICMET).
- Whittaker, J. 1989. “Creativity and conformity in science: Titles, keywords and co-word Analysis.” *Social Studies of Science*, 19: 473-496.
- Wouters, P. 1999; The citation culture. Unpublished doctoral thesis, University of Amsterdam, [cited 2006.4.25]. <<http://www.garfield.library.upenn.edu/wouters/wouters.pdf>>.