

KDD 방법을 이용한 Biochip 관련 분야들의 초기와 현재의 구조적 비교와 분석

The structural comparison and analysis of the early and current status in
biochip research using KDD

오해영 · 박각로 · 박상진 · 윤문섭
과학기술정책연구원 기술경제팀

Hae Young Oh, Gak Ro Park, Sang Jin Park, Moon Seob Youn
Techno-Economics Team

Abstract

The aim of this study is to map the structure of biochip research field and analyse ex-ante feasibility in R&D planning stage using Co-word analysis. We used SCI Database as a source data to analyze the intellectual structure of biochip research in two different periods (1994~1999, 1994~2002).

1. 서론

1.1 연구배경

1995년 국민소득 1만 불 시대에 들어선 우리나라는 WTO 출범 및 인터넷을 통한 지식 정보 혁명에 의해 지역을 초월하는 글로벌 경제체제를 맞이하게 되었다. 1997년 IMF 경제 환란이후 우리나라는 연구개발투자 위축 등으로 인해 선진국과의 기술적 격차는 줄어들지 않으면서 후발국들의 강한 도전을 받게 되었고 국가경제 미래에 대한 불안감은 점차 확산되어가는 상황 하에 놓이게 되었다.

따라서 정부는 “혁신” 중심의 산업구조로의 전환, 신 성장 원천 확충 등을 통해 돌파구를 찾아야 한다는 공감대를 형성하였고, 선진경제로의 도약을 목표로 향후 5~10년 동안 우리나라에 강점이 있고 부가가치가 큰 차세대 성장동력을 발굴하여 여기에 국가 역량을 집중하고자 2002년부터 2006년까지 신기술

분야에 약 13조원을 집중 투자하는 계획을 수립한 바 있다.

이러한 신기술 개발에 있어서 불확실성으로 인한 실패위험은 항상 따르기 마련이다. 따라서 본격적인 투자가 시작되기 이전인 연구기획단계에서부터 기술적·경제적 타당성을 지속적으로 분석할 수 있는 방안이 강구될 필요가 있고, 미국과 같은 선진국에서는 1990년대부터 사전타당성 분석을 의무화 하고 있다.

사전타당성 분석을 위해 과학연구와 과학정책에서 사용되어진 전통적인 방법은 상대적으로 소수 전문가들의 견해를 구하는 방법이 었다(Law & Whittaker, 1992). 하지만, 본 연구에서 사용되어질 Co-word analysis는 정량적 측면에서 이러한 작업을 수행하기 위한 또 다른 방법으로 이는 사회적이고 정책적인 질문들에서 과학과 기술을 개발하고 지식의 발전을 목표로 하고 있다(Van Raan, 1997). 해외의 많은 연구자들이 다양한 분야에서 학문적 구조 및 개념간의 연결 관계를 찾아내기 위한 중요한 방법으로 Co-word analysis를 사용하고 있고, 예를 들면 Polymer chemistry (Callon et al., 1991), 신경망 연구(Noyons & Van Raan, 1998a; Tijssen, 1993), Software engineering(Coulter et al., 1998) 등이 있다.

1.2 연구목적

본 연구의 목적은 불확실성이 가장 높은 연구기획단계에서 사전 타당성 분석에 적합한 것으로 제시되어진(윤문섭 외, 2003) 다량의 데이터를 누적하고 있는 데이터베이스로부터 지식을 발견해내는 Knowledge Discover in Database를 구현하는 방법론 중 하나인 Co-word analysis 기법을 Biochip 분야에 적용하여 학문적 구조 및 연계성 분석을 시도하였다. 또한 이를 통해 Biochip 분야 내 세부 분야의 연구동향변화를 도식화(Knowledge Map)하여 추적하였고, 이를 연구기획 및 활동에 보충적 참고자료로 활용하고자 하는데 그 부차적인 목적을 두고 있다.

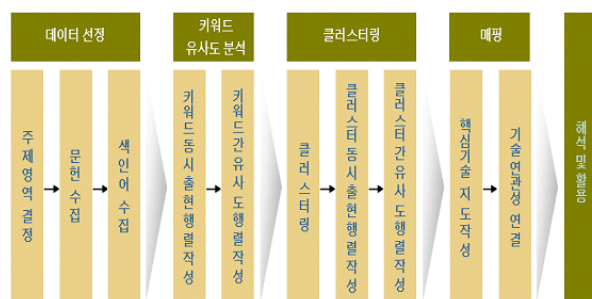
2. 연구방법

본 연구에서는 Biochip 분야의 세부분야 구조를 초기(1994~1999)와 전 기간(1994~2002)을 대상으로 Ding(2000), Rann & Velde(1991)의 연구방법론을 사용하여 비교 분석하였다. 분석대상으로 사용되어진 데이터는 1994~2002년까지 SCI DB에 수록된 문헌의 서지정보이고 데이터 분석을 위해 MS-SQL 2000 및 SPSS 11.5 등을 사용하였다.

2.1 지식맵 작성법

2.1.1 데이터 선정

Biochip분야와 관련된 핵심검색어를 선정하기 위해 학계 및 기업의 전문가들에게 의뢰하여 26개의 검색어를 선정하였다.



<그림 1> Knowledge map분석 방법론

이를 통해 5,949건의 Biochip관련 문헌을

검색하였고 이들 문헌이 가진 키워드를 대상으로 전문가의 thesaurus 작업을 통해 총 15,540개의 키워드를 추출하였다.

2.1.2 키워드 유사도 분석

키워드 간의 유사도 측정을 하기 위해 한 문헌에서 두 개의 키워드들이 동시에 발생하는 빈도를 산출하였다. 이 중 동시출현빈도수가 11개 이상 500개 이하인 총 2304개의 키워드를 가지고 2304X2304 동시출현행렬을 작성하였고, Pearson상관계수를 이용하여 키워드 간 유사도 행렬을 산출하였다.

2.1.3 클러스터링

키워드 간 유사도를 이용하여 Hierarchical Clustering을 실시하였고, 전문가에 의한 Dendrogram 분석을 통해 10개의 클러스터를 선정하였다. 이렇게 구해진 클러스터를 가지고 각 클러스터에 속한 문헌들의 동시출현 빈도를 구한 후, 다시 Pearson 상관계수를 이용하여 유사도 행렬을 산출해 내었다.

2.1.4 매핑

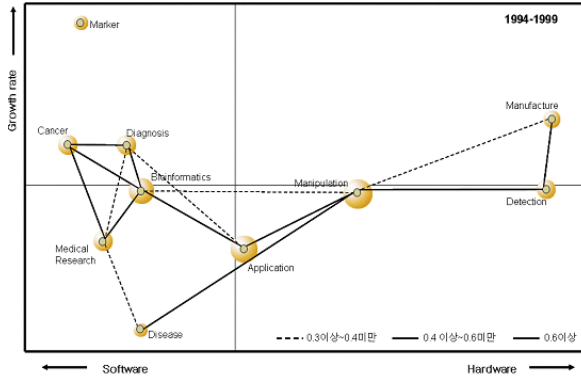
학문적 구조 및 개념간의 관계를 도식화하기 위해 MDS(multidimensional Scale) 분석을 실시하였다. 여기서 클러스터간의 연관성을 나타내기 위해 Proximity Index값이 0.3이상 0.4미만인 것들은 점선, 0.4이상 0.6미만인 것들은 가는 실선, 그리고 0.6 이상인 것은 굵은 실선으로 나타내었다.

3. 분석결과

Biochip 연구의 분야별 특성과 현황을 MDS를 이용한 Knowledge map을 통해 두 기간으로 나누어 시기별 특성을 비교, 분석하여 보았다. 10개로 분류되어진 클러스터명은 각 클러스터들에 포함된 키워드들의 특성을 고려하여 명명 되어졌다.

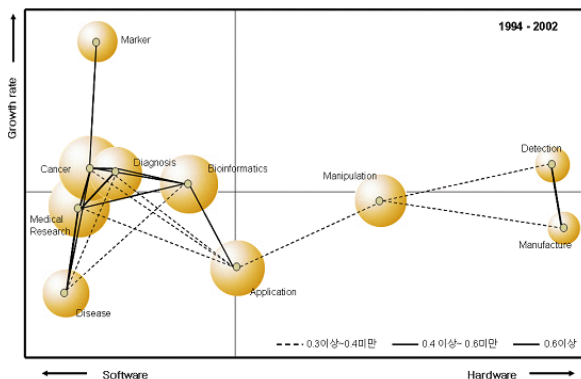
Knowledge map상에 나타나는 클러스터간의 연결선은 각 클러스터 간의 연관도를 나타낸 것이고 클러스터의 크기는 각 클러스터에 속한 문헌의 수를 나타내는 것이다.

클러스터링 결과를 가지고 전문가들과 인터뷰를 한 결과 10개의 클러스터들은 2개의 큰 분류인 Hardware (Detection, Manipulation, Manufacture)와 Software (Marker, Bioinformatics, Cancer, Diagnosis, Medical Research, Disease, Application)로 나누어짐을 알 수 있었다.



<그림 2> Biochip분야의 Knowledge map(1994~1999)

<그림 2>의 Knowledge map을 살펴보면 Biochip분야 연구의 동태기 이므로 상대적으로 하드웨어 개발 및 응용에 초점을 맞춘 연구가 이루어졌음을 알 수 있다. 또한, 개발되어진 하드웨어를 Cancer, Diagnosis, Medical Research, 그리고 Disease와 같은 분야에 응용하여 운용할 것인가에 대한 연구가 높은 관계를 가지고 이루어졌음을 이 분야에 전문적인 지식을 가지고 있지 않은 사람들이더라도 Knowledge map상에서 쉽게 살펴볼 수 있다.



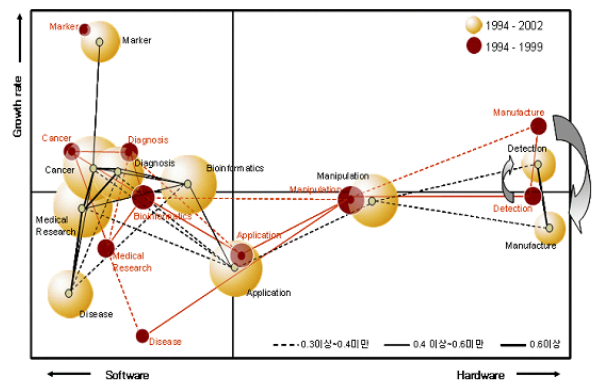
<그림 3> Biochip 분야의 Knowledge map(1994~2002)

<그림 3>의 Knowledge map은 연구 초기 단계의 데이터를 포함하고 있는 2002년까지의 Knowledge map이다. 전반적으로 보았을 때 동태기에서 보여주었던 Hardware 분류의 강세는 조금씩 수그러들고 Software 분류와 관련된 연구가 집중적으로 이루어져 왔음을 알 수 있다.

몇 가지 차이점을 살펴보면 Cancer, Diagnosis, Medical Research의 긴밀한 연관관계 집단에 Disease 클러스터가 높은 연관관계를 가지며 연구가 진행되어짐을 알 수 있다. 또한 초기 단계에 상관관계를 가지지 않고 진행되던 Marker분야가 Cancer와 밀접하게 관련되어져 활발한 연구를 수행하고 있음은 인상적인 부분이라 할 수 있다.

또한, 생물정보를 수록한 DB를 구축하여 생명현상과 관련된 새로운 법칙을 찾아내는 Bioinformatics 분야는 기간의 영향을 받지 않고 핵심적인 Biochip분야임을 살펴볼 수 있다.

<그림 4>에서는 이상에서 분석하여 예시한 Knowledge map을 하나의 Map안에 함께 보여줌으로 연구동향을 보다 쉽게 이해하고 비교할 수 있도록 하였다.



<그림 4> 구간별 연구동향비교

4. 결론

생물 세포내 기존 유전자를 재조합 또는 융합하는 기술에 한정되었던 바이오 기술은 21세기 Post-genome 시대를 맞이하면서 기술 시장 내에서 생물의 유전암호 해독 및 작용기작 해석을 통해 유전정보를 사회 및 경제적 수요에 맞도록 재 디자인하여 입력하는 기술

로 보다 광범하게 적용되어가고 있다.

본 연구는 Post-genome시대의 핵심 연구 분야인 Biochip 분야를 Knowledge map을 이용하여 분석함으로써 세계적 연구동향 변화의 흐름을 보다 손쉽게 파악할 수 있게 해주었다는데 그 의의를 가지고 있다.

이러한 연구 결과는 전문가 회의(Peer review)방식의 한계점인 전문가들 간의 편견 및 후광효과 등(Kostoff, 1998)을 조율해 줄 수 있는 객관적성을 가진 좋은 도구로 작용할 수 있다.

하지만 과거의 데이터만을 통한 분석에 의존함으로써 미래의 변화 예측에는 다소 무리가 있는 것으로 보인다. 또한, 학술관련 데이터만을 다루고 시장성을 기반으로 한 특허 분석을 실시하지 않음으로 기술변화의 흐름을 완전히 반영하지 못하는 것이 연구의 한계로 지적될 수 있다.

따라서 향후 연구에서는 특허구조분석을 통한 과학과 기술 간의 구조적 유사성 및 차이점에 대한 비교분석이 반드시 이루어져야 할 것이다.

Reference

윤문섭, 이우형, 이종석, 장진규, "신기술분야 대형연구개발과제의 사전심의를 위한 타당성 분석방법", 정책연구, 2002-27, pp.133-134, 2003

Callon, M., Courtial, J. P. & Laville, F., "Co-word analysis as tool for describing the network of interactions between basic and technological research : The case of polymer chemistry", *Scientometrics*, Vol.22, No.1, pp.153-203, 1991

Coulter, N., Monarch, I. & Konda, S., "Software engineering as seen through its research literature : A study in co-word analysis", *Journal of the American Society for Information Science*, Vol.49, No.13, pp.1206-1223, 1998

Ding, Y., G. Gobinada G. Chowdhury, S. Foo., *Bibliography of information*

retrieval research by using co-word analysis, *Information processing and management*, 2000

Law, J. & Whittaker, J., "Mapping acidification research : A test of the co-word method", *Scientometrics*, Vol.23, No.3, pp.7-17, 1992

Noyons, E. C. M. & Van Raan, A. F. J., "Advanced mapping of science and technology", *Scientometrics*, Vol.41, No.1-2, pp.61-76, 1998

Tijssen, R.J.W., "A scientometric cognitive study of neural network research: expert mental maps versus bibliometric maps:", *Scientometrics*, 28(1), pp. 111-136, 1993

Kostoff, R.N., "Database Tomography for Technical Intelligence: A Roadmap of the Near-Earth Space Science and Technology Literatur.," *Information Processing and Management*, 34(1), 1998

Van Raan A. F. J and Van der Velde J. G. M., *Bibliometric mapping and early warning for emerging topics : Exploring developments in material science. Research Evaluation*, 1991

Van Raan, A. F. J., *Scientometrics: State-of-the-art. Scientometrics*, 38(1), 1997