

## Co-word Analysis을 통한 신기술 분야 도식화 방법에 관한 연구\*

이우형\*\* · 김윤명\*\*\* · 박각로\*\*\* · 이명호\*\*\*\*

### A Study on the Emerging Technology Mapping Through Co-word Analysis\*

Woo Hyoung Lee\*\* · Yun Myung Kim\*\*\* · Gak Ro Park\*\*\* · Myoung Ho Lee\*\*\*\*

#### ■ Abstract ■

In the highly competitive world, there has been a concomitant increase in the need for the research and planning methodology, which can perform an advanced assessment of technological opportunities and an early perception of threats and possibilities of the emerging technology according to the nation's economic and social status.

This research is aiming to provide indicators and visualization methods to measure the latest research trend and aspect underlying scientific and technological documents to researchers and policy planners using "Co-word Analysis". Organic light emitting diodes(OLED) is an emerging technology in various fields of display and which has a highly prospective market value. In this paper, we presented an analysis on OLED.

Co-word analysis was employed to reveal patterns and trends in the OLED fields by measuring the association strength of terms representatives of relevant publications or other texts produced in the OLED field. Data were collected from SCI and the critical keywords could be extracted from the author keywords. These extracted keywords were further stancardized. In order to trace the dynamic changes in the OLED field, we presented a variety of technology mapping. The results showed that the OLED field has some established research theme and also rapidly transforms to embrace new themes.

Keyword : Co-word analysis, Keyword, Emerging Technology, Visualization Method, SCI

논문접수일 : 2006년 08월 05일 논문게재확정일 : 2006년 10월 30일

\* 이 논문은 2006학년도 한국외국어대학교 학술연구비의 지원에 의하여 이루어진 것임.

\*\* 정보통신연구진흥원

\*\*\* 한국부품소재산업진흥원

\*\*\*\* 한국외국어대학교 경영학부

† 교신저자

## 1. 서론

오늘날 비약적으로 발전한 과학기술 문헌의 범위와 양적 성장을 기반으로 이를 이용한 과학기술 분석들이 등장하였는데 이를 가르켜 “bibliometric” 또는 “scientometrics”라 한다. 이러한 분석의 결과들은 기술분야의 급변성에 부응하고 미래의 발전성을 예측하는 연구자, 연구책임자, 기술정보전문가 그리고 기획자들에게 유용한 가치를 제공한다.

그러나 많은 과학기술 문헌들 속에서 연구자들은 연구 분야를 발견하고 자기 연구분야에서 연구 영역간을 연계하는 것이 쉽지 않으며 정책 입안자들도 연구계획을 수립하기 위해 과학기술분야의 변화양상을 도식화하는데 어려움을 겪고 있다.

이러한 도식화를 위해 몇 가지 방법들이 시도되어졌다. 과학연구와 과학정책에서 사용되어진 전통적인 방법은 소수 전문가들의 의견을 구하는 방법인 Peer Review이다[16]. 그러나 여기에는 다음과 같은 결점이 존재한다. 첫째 전문가의 조사 규모가 매우 작지 않으면 이것은 비용이 많이 드는 방법이며, 둘째 조사규모가 작다면 대표성의 문제가 있게 되며, 셋째 문제의 과학분야가 개발 완료 되었는지 발전중인지와 같은 관점의 범위를 대조하는 문제가 복잡하다[10]. 따라서 bibliometric 방법론은 정량적 측면에서 이러한 작업을 수행하기 위한 또 다른 방법이며 대표적인 정량적인 방법중의 하나가 Co-word 분석방법이다.

저자나 저널의 co-citation analysis과 같은 전통적인 bibliometric 방법들은 과학논문에 내포되어있는 인용자료들에 대한 분석을 근간으로 이루어진다. 이러한 종류의 분석들은 흥미로운 결과를 가져다주는 반면, 논문을 이용한 연구주제의 실제적인 내용에 대한 즉각적인 그림을 공급해 주지는 못한다. 주어진 주제 상에서 출간되어진 논문내에서 키워드들의 동시출현을 측정하거나 분석하는 Co-word 분석은 이러한 종류의 분석적 문제를 간결하게 처리할 수 있는 잠재력을 가지고 있다[2].

Co-word 분석방법은 데이터에 포함되어져 있

는 기본적인 정보를 유지하면서 데이터를 독특한 시각적 방법으로 단순화시켜서 보여준다. 이것은 과학적인 개념이나 아이디어 그리고 지식 등의 중요한 제공자로서 자연어를 근간으로 한다[26].

그동안 많은 연구자들이 Co-word 분석방법을 통해 다양한 분야의 영역분석과 변화양상을 파악하는데 활용하였다. 소프트웨어 공학분야[5], 폴리머 화학[2], 신경망연구[19, 26], 광학공학[18], Bio-electronics[11], 약학[22], 생물학[17, 23], 농축물 질물리학[1] 등의 연구가 있다.

Co-word 분석은 논문의 키워드가 그 논문의 내용을 적절하게 표현하고 있거나, 다양한 문제들간의 연결에 의해 논문이 구성되어진다는 가설 하에 이루어진다. 같은 논문 내에서 두개의 키워드가 동시에 발생한다는 것은 그 논문이 언급한 두개의 주제들과 연관되어져 있다는 것이다[4]. 같은 단어나 같은 쌍의 단어들이 논문들에서 자주 나타난다는 것은, 그 연구의 주제가 논문들내에서 중요하게 연계되어져 있다는 것을 지적해 준다. Co-word 분석은 특정 영역에서 출간되어진 관련 간행물들의 대표적인 용어들 간의 관련된 강도를 측정하여서 특정분야에서의 패턴 및 경향을 보여준다. Co-word분석의 주된 형태는, 하나의 특정한 학문분야의 지적구조를 이 영역의 개념적인 공간의 지도로 표현하거나, 이러한 개념적 공간에서의 변화를 추적하여서 시계열적인 지도에 시각적으로 나타내는 것이다.

따라서 본 연구의 목적은 Co-word 분석방법을 통하여 과학기술연구자와 정책입안자들에게 문헌의 서지적 현상속에 숨겨져 있는 과학기술분야의 새로운 연구동향과 변화양상을 도식화하여 제시하는 것이다. 이를 위하여 본 연구에서 분석하고자 하는 분야는 차세대 디스플레이(OLED) 분야이다. 이 분야를 연구의 대상으로 선정한 이유는 차세대 디스플레이 분야의 경제적 중요성과 다양한 분야들이 결합되어져 있는 넓은 영역이기 때문이다. 아울러 기초연구분야에서 뿐만 아니라 응용 및 기술 분야를 서지적인 방법과 기술로 측정하는 것 또한

중요하기 때문이다.

## 2. 이론적 고찰

Co-word 분석방법은 텍스트에서 제시된 특정 영역 안에서 아이디어간의 관계를 확인하기 위해서 텍스트의 언어 부분에서 아이텀쌍이 동시에 출현하는 패턴을 사용하는 콘텐츠 분석기법이다. 이러한 Co-word 분석은 1980년대 프랑스 Ecole Nationale Superieure des Mines의 Centre de Sociologie de l'Innovation과 CNRS(Centre National de la Recherche Scientifique)가 협력하여 개발한 "LEXIMAPPE"라고 하는 시스템에서 처음 시작되었다. Co-word 분석은 그 후에 Callon 등[3]의 "Mapping of the Dynamics of Science and Technology"라는 책을 통하여 하나의 방법론으로 정리되어 프랑스뿐 아니라 영국, 네델란드, 미국 등 다른 나라로 확산되었고 과정과 측정, 해석적인 측면에서 향상을 가져오게 되었다[10].

Co-word 분석을 사용하여 주제영역의 경향과 시기적 변화 등을 파악한 선행연구들을 살펴보면 다음과 같다.

Tijssen과 van Raan[25]은 생명공학 분야를 대상으로 "LEXIMAPPE"이 생성하는 영역지도와 다차원축적을 이용한 영역지도를 비교하였다. "LEXIMAPPE"이 생성한 영역지도는 원형(circular map)과 수직형 두 가지 유형이 있다. 원형지도는 빈도가 높은 단어를 중심에 위치시키고, 수직형 지도는 맨위에 위치시켜 단어의 빈도가 높을수록 주제 영역에서 영향력이 있는 것으로 표현하였다.

Peters와 van Raan[20]은 화학공학 분야의 영역분석을 위해 다차원축적을 직접적으로 이용하는 방법(direct-MDS)과 클러스터링과 다차원축적을 결합한 방법(cluster-MDS)을 사용하였다. 그리고 Peters와 van Raan[20]은 3개의 다른 실험집단, 세가지 단어 연관도 측정 지수, 그리고 두 가지 단어 유형을 사용하여 화학공학 분야의 영역지도를 비교하였다.

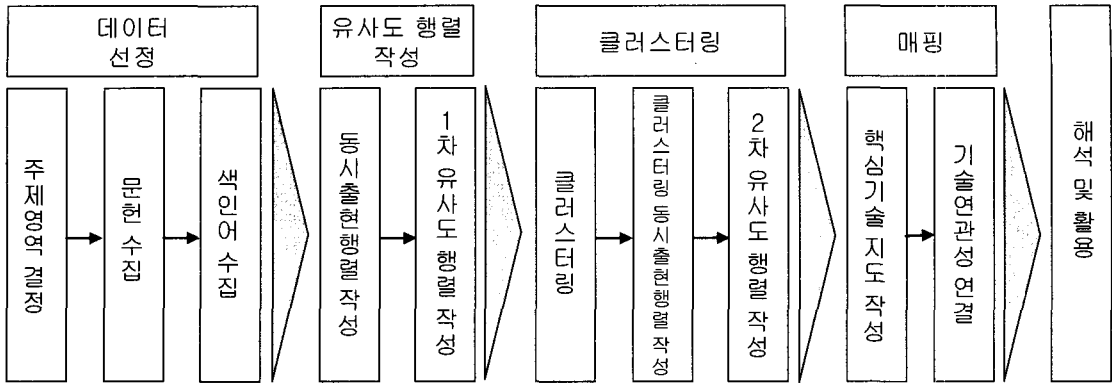
Noyon과 van Raan[19]은 신경망 분야를 대상으로 1989~1990년, 1992~1993년의 두 시기로 나누어 현재의 구조로 과거를 보는 연구, 과거의 구조로 현재 영역을 살펴보는 연구를 하였는데 이것은 현재의 구조로 현재를 보는 연구와 과거의 구조로 과거를 보는 연구를 기초로 한다.

Ding 등[8]은 계량정보학을 검색 분야에 활용하여 온라인 데이터베이스와 웹 정보를 검색하는 시스템인 BIRS(Bibliometric Information Retrieval System)를 제시하였다. BIRS는 계량정보학적 분석방법을 통해 이용자의 질의를 확장시키기 위하여 구현된 시각적 인터페이스를 가진 시스템이다. 이용자는 저자지도(author map), 저널지도(journal map), 키워드지도(keyword map)를 통해 정보를 검색할 수 있다. 이 중 키워드지도는 이용자가 다양한 수준에서 접근하여 결과들을 볼 수 있도록 하였는데, 1단계 수준에서는 넓은 범위의 영역지도를, 2단계 수준에서는 상세한 영역지도를, 그리고 3단계 수준에서는 유사도가 높은 상위 20개의 관련 단어들을 제시한다.

Ding 등[9]은 SCI와 SSCI에서 검색된 정보검색 분야의 문헌을 대상으로 주제 영역과 시기에 따른 변화를 분석하였다. 이들은 2,012개 문헌에서 추출한 3,227개의 키워드를 수집하였는데, LISA 시소러스를 이용한 표준화 작업과 표제와 초록으로부터 수작업으로 키워드를 추가하는 방법을 통하여 총 240개의 키워드를 선택하였다. 이 키워드들간의 동시출현빈도를 이용해 5개의 클러스터를 만들고 다차원축적을 이용해 정보검색분야의 영역지도를 만들었다. 그리고 각 클러스터별로 상세한 영역지도를 만들었다.

## 3. Co-word 분석을 통한 Technology Mapping Method

Co-word 분석방법은 분석분야의 선정, 분석자료 및 유형선정, 단어들의 동시출현빈도 정규화 및 유사도 측정, 그리고 클러스터링과 Mapping을



〈그림 1-1〉 본 연구의 방법론

통하여 하위구조를 만들고 하위구조들간의 상호작용을 표현하는 방법 등의 내용을 포함해야 한다.

### 3.1 데이터 선정

단어는 Co-word 분석에서 가장 중요한 요소이기 때문에 저널 단위로 데이터를 선택하기 보다는 문헌 단위로 데이터를 선정하는 것이 바람직하다. 저널이나, 학회논문, 보고서, 또는 책의 내용 등에서 이러한 단어들을 추출해 내는 것은 두 가지 방법이 있다. 첫 번째는 단어를 키워드 목록, 주제, 초록, 아니면 어떠한 분석코드를 가진 데이터로부터 추출해 내는 것이다. 많은 저널, 논문초록 서비스, 그리고 데이터베이스들은 이미 이러한 키워드들을 제공하고 있다. Cambrosio 등[4]은 자신들이 가지고 있는 연구로 사용되어질 데이터의 취약한 구조로 인한 품질을 보완하기 위해 몇 개의 색인과 주제를 추가시켜 키워드를 추출해낸 바가 있다. 이 결과로 제공되어진 기술어(記述語)들은 다른 형태의 철자와 같은 용어를 사용함에 있어서 다르게 되어진 표현들을 표준화 한 것이었다. Coulter 등[5]은 전문적인 색인작성가에 의해 얻어진 기술어(記述語)를 사용하였다. 그들은 일반적인 명칭들을 요구하는 고정적인 시스템에 대한 연구를 진행함에 있어서 이러한 방법이 상당히 유용할 것이라 믿었다. 전문적인 색인작성가들의 경험들은 분류학적으

로 표준화되어진 도구들을 사용할 것이라는 확신을 준다. Looze와 Lemarie[17]는 전문가들에 의해 제시되어진 키워드들을 기반으로 Co-word 분석을 수행하였다. 몇몇 연구원들은 색인작성가와 저자에 의해 추가되어진 온라인 데이터베이스에서 키워드를 제공받기도 한다[6, 7, 16]. Noyons과 van Raan [19]은 신경정보망 분야의 조악한 전체구조를 분류코드의 동시출현을 이용하여 도표화한 바 있다.

조절되어진 용어로부터의 이러한 데이터 수집에 있어서 가장 중요한 조건은 “색인자 효과(Indexer effect)”의 가능성이다. 즉, 다른 데이터베이스를 가지고 작업하는 전문 색인작성가에 의한 키워드 선택시 그들의 선입관이나 관점이 투영되어질 수 있다는 것이 바로 이러한 문제점이다[13]. 하지만 이러한 영향은 올바르게 진행되어진 인터뷰의 결과로 상쇄되어질 수 있다[4, 6, 16, 24].

두 번째 데이터수집 방법은 NPtools와 같은 소프트웨어를 이용하여 문서전체로부터 단어를 직접 추출해 내는 것이다[27]. 적절한 빈도의 단어나 구절들은 특정분야의 핵심주제를 표현하기 위한 Co-word 분석기법의 주제처럼 선정되어졌다. 이러한 방법들은 색인작성가들의 쓸모없는 노력과, 분류시스템과 컴퓨터에 기억된 색인검색 시스템을 구축하는데 소요되어지는 시간문제나, 새롭게 개발되어질 관련분야의 유지 및 보수에 소요되는 시간상의 문

제 및 어려움 등을 피하기 위해 선택되어진다.

### 3.2 유사도 행렬 작성

Co-word 분석기법에서는, 한번 연구주제가 선택되어지면 단어의 동시발생을 기초로 하는 동시출현 행렬을 작성한다. 두 단어들의 셀에 대한 가치는, 이런 두 단어가 하나의 문서에서 동시에 발생하는 횟수에 의해 결정되어진다. 두 단어간에 나타나는 높은 수준의 동시발생빈도는 그 둘 간의 밀접한 관계가 있다는 것을 보여준다[9]. 이러한 동시출현 행렬을 그대로 1차 유사도 행렬로 사용하기도 한다.

그러나 대부분 유사계수를 사용하여 유사도 행렬을 작성하게 된다. 유사계수를 사용하는 목적은 동시출현빈도의 범위를 정규화 함으로써 출현빈도가 높은 단어들과 출현빈도가 낮은 단어들간의 출현빈도 차이를 표준화시킬 수 있기 때문이다. 이러한 유사계수는 벡터 구성요소의 성질에 따라 가중치벡터용, 이진벡터용[3, 5, 20]으로 나눌 수 있고 중요시하는 단어의 빈도수준에 따라 고빈도어 선호, 저빈도어 선호[20]로 나눌 수 있다. 또한 최근에 Ding 등[9]은 피어슨 상관계수를 사용하여 단어간 연관도를 측정하였다. 피어슨 상관계수  $r$ 은 두 단어의 유사성을 측정하기 위해 두 단어와 Co-word 행렬의 다른 단어들과의 동시출현빈도를 이용하여 구해지고 -1에서 +1까지의 값을 갖는다. 하지만 어떤 유사계수를 선택하는 것이 주제분야의 현재 기술상태와 발전되고 있는 상태를 표현하기 위해 가장 좋은지는 명확하지 않다.

### 3.3 클러스터링

유사도 행렬이 작성되면 이를 바탕으로 용어와 문헌들의 분류를 위해 클러스터링을 실시한다. 현재까지 Co-word 분석에서 주로 쓰이는 클러스터링 기법은 주성분분석(Principal Component Analysis; PCA)과 계층적 클러스터링 기법(Hierarchical Clus-

tering)이 있다.

주성분분석이란 해석하고자 하는 다차원의 데이터를 거기에 포함된 정보의 손실을 가능한 한 적게 해서 2 혹은 3차원의 데이터로 축약하는 기법이다. 주성분분석을 활용하면 관측대상이 어떠한 위치에 있는지 시각적으로 파악할 수 있게 된다. 또한 주성분분석은 다수의 지표를 통합한 종합적인 지표를 작성하고, 관측대상을 몇 개의 그룹으로 나누고, 중회귀분석이나 판별분석을 위한 데이터를 다른 관점에서 음미하고자 하는 목적으로 사용된다.

계층적 클러스터링은 더 큰 클러스터를 생성하기 위해 이미 생성된 클러스터들이 일정한 거리값에 따라 단계적으로 결합을 반복하는 것으로 덴드로그램(dendrogram)으로 표현가능하다. 이러한 계층적 클러스터링에는 클러스터내의 구성원들을 적어도 한 개의 링크에 의해 서로 연결시키는 방법인 단일연결(single linkage), 각 클러스터 사이의 유사도를 결정하기 위해 가장 유사도가 작은 쌍을 이용하는 완전연결(complete linkage), 특정 클러스터에 포함된 모든 문서의 평균 유사도로 클러스터 유사도를 결정하는 집단평균(group average), 클러스터 센트로이드간의 유클리드 거리를 최소화하는 방법으로 클러스터를 생성하는 와드기법(Ward's method) 등 4가지 종류가 있다.

이렇게 구해진 클러스터를 기반으로 클러스터 동시출현 행렬을 작성하고 또한 2차 유사도 행렬을 작성한다. 작성방법은 앞서 언급한 유사도 행렬작성법과 동일하다. 한 가지 차이가 있다면 클러스터 내에 속한 단어를 기반으로 한다는 점이다.

### 3.4 맵핑

데이터를 맵핑하는 것은 몇 가지 방법들이 있는데, 그 중 가장 일반적으로 사용되어지는 방법은 다차원 축적법(Multidimensional scaling)이다. 다차원 축적법이란 대상간의 유사성 측도에 의거해서 대상을 다차원 공간속에 배치시키는 방법으로 유사성이 작은(비유사성이 큰) 대상끼리는 멀리, 유사

성이 큰(비유사성이 작은) 대상끼리는 가깝게 위치를 정한다.

다른 방법들은 특정한 프로그램들을 사용하여야 한다. Co-word 맵핑을 위한 LEXIMAPPE 프로그램은 과학정책 도구로 개발되어졌고, 다양한 연구분야에서의 간행물들을 분석하는데 이미 사용되어진바 있다[4, 6, 16, 17]; Content Analysis and Information Retrieval (CAIR)은 Caregie Mellon 대학교 내의 Software Engineering Institute에 의해 개발되어졌고, Co-word 분석을 위해 근무하던 연구원들에 의해 사용되어졌다[5]; Austrian Research Centers에서 개발한 Bibliometric Technology Monitoring(BibTechMon)은 Co-word 분석을 위해 개발되어진 또 다른 소프트웨어 중 하나이다[15, 28].

Kohonen[14]의 Neural network 알고리즘은 데이터를 도표화하기 위한 또 다른 접근법이다. Polanco 등[21]은 Adaptive Resonance Theory(ART), a Multilayer Perceptron (MLP) 그리고 Associative network with unsupervised learning (KOHNRN) 등을 포함한 인공지능망 기술을 과학과 기술의 정보영역 평가 및 도표화에 적용한 바 있다. WEB-SOM Research group이 그러한 예 중 하나인데, 그들은 웹을 기반으로 한 Self-Organizing Map interface를 구축하고 이를 이용하여 많은 양의 논문들로부터 얻어낸 단어들을 도식화한 바 있다[14]. WEBSOM은 Self-Organizing Maps를 이용하여서 완전히 자연발생적이고 통제되지 않은 문서 Set의 전체내용 분석을 수행하였다. 이러한 분석의 결과는, 문서 간격이 정렬된 Map은, 그 문서들의 주제가 되는 내용들의 유사성관계를 직접적으로 나타낸다. 이들은 Document 맵 상에서의 관계들의 거리들을 나타내준다. 문서간격 내 서로 다른 부분의 문서의 밀집도는 문서지도의 표현시 색상 차로 표현할 수 있다[12].

이렇게 다차원 축적 지도가 완성되면 클러스터에 속한 단어를 기반으로 클러스터간의 유사도를 구해야 한다. 여기에는 두 가지 방법이 있다. 먼저

Callon 등[2]은 클러스터간 상이성(dissimilarity)을 측정하기 위해 변화지수(transformation index) t를 사용하였고, Coulter 등[5]은 변화지수에서 파생된 유사성지수 SI(Similarity Index)를 통하여 시기가 다른 네트워크간의 유사도를 구하였다. 다음으로 Peter와 van Rann[20]은 두 클러스터에 속한 단어목록(word-profile)과 가중치를 이용한 코사인계수를 이용하여 클러스터간 유사도를 구하는 방법을 제시하였다.

## 4. 분석 결과

### 4.1 데이터 선정

OLED(Organic Light Electroluminescence Device)

<표 4-1> OLED 관련 문헌 추출을 위한 query key word

1. Organic semiconductors,
2. Electroluminescence,
3. Carrier injection,
4. Optical polymers,
5. Organic thin film transistors
6. Flat panel displays,
7. Optical properties,
8. ITO glass,
9. Organometallic compounds,
10. Organic thin films,
11. Electrical properties with photoluminescence or Electrical properties with electroluminescence,
12. FLEDs,
13. OLED,
14. Organic light emitting diodes,
15. Optoelectronics,
16. Excitons,
17. Electron-hole recombination,
18. Polymer light emitting diodes
19. Thermionic emission,
20. Optical properties with photoluminescence or Optical properties with electroluminescence

〈표 4-2〉 OLED 관련 키워드들의 동시출현 빈도수 행렬

	var00001	var00002	var00003	var00004	var00005	var00006	var00007	var00008	var00009
1	1.00	.00	.00	.00	.00	.00	.00	.00	.00
2	.00	4.00	.00	.00	.00	.00	.00	.00	.00
3	.00	.00	6.00	.00	.00	.00	.00	.00	.00
4	.00	.00	.00	4.00	.00	.00	.00	.00	.00
5	.00	.00	.00	.00	9.00	.00	.00	.00	.00
6	.00	.00	.00	.00	.00	6.00	.00	.00	.00
7	.00	.00	.00	.00	.00	.00	1.00	.00	.00
8	.00	.00	.00	.00	.00	.00	.00	6.00	.00
9	.00	.00	.00	.00	.00	.00	.00	.00	12.00
10	.00	.00	1.00	3.00	1.00	2.00	.00	.00	.00
11	.00	.00	.00	.00	.00	.00	.00	.00	.00
12	.00	.00	1.00	.00	.00	.00	.00	.00	.00
13	.00	.00	.00	.00	.00	.00	.00	.00	.00
14	.00	.00	.00	.00	.00	.00	.00	.00	.00
15	.00	.00	.00	.00	.00	.00	.00	.00	.00
16	.00	.00	.00	.00	.00	.00	.00	.00	.00
17	.00	.00	.00	.00	.00	.00	.00	.00	.00
18	.00	.00	.00	.00	.00	.00	.00	.00	.00
19	.00	.00	.00	.00	.00	.00	.00	.00	.00
20	.00	.00	.00	.00	.00	.00	.00	.00	.00

관련 synthetic bibliometric 연구를 위한 데이터 선정을 위해서 20개의 query key word를 사용하였다. 그 20개의 key word를 <표 4-1>에 나타내었다. 이러한 query key word 선정방법은 OLED를 나타내는 용어인 5개의 key word(OLED, Organic light emitting diode\*, Organic electroluminescence device\*, Polymer light emitting diode)를 이용하여 850 편의 논문을 1차로 추출한 후에 이들 표본 논문의 제목, 저자 키워드, 키워드 플러스, 초록에 포함된 key word 중에서 출현빈도수가 높은 70개의 2차 키워드를 선정하고 이 중 OLED와 관련성이 높은 20개를 전문가의 검토를 거친 후에 query key word로 선정한 것이다. 본 연구의 대상문헌은 1995년부터 2002년까지 8년의 연구기간 동안 발표된 SCI 문헌 중 query key word에 의해 추출된 OLED 관련 문헌 8217건이다).

1) 분석에 사용된 소프트웨어로는 데이터 선정시 MS-SQL, 유사도행렬, 클러스터링, 맵핑에 SPSS 10.0을 사용함.

## 4.2 유사도 행렬 작성

유사도 행렬에 사용된 OLED 관련 8,217 건의 문헌에서 추출한 index key word의 수는 모두 12,558 개이고 이 중 키워드 클리닝(cleansing) 작업<sup>2)</sup>과 시소러스 과정<sup>3)</sup>을 마친 후의 index key word 수는 10,144 개였다. 이들 키워드 중에서 500회 이상의 출현빈도수를 보이는 경우의 키워드들은 유사도 행렬 작성 시에 발생할 수 있는 과반수효과<sup>4)</sup>를 고려하여 제외시켰고, 중간 단계의 출현빈도수를 갖는 500 개의 키워드를 이용하여 이들의 동시출현 빈도수에 관

- 2) 추출된 키워드들의 단·복수, 약어, 각종 입력 오류 등을 교정하는 작업을 통칭함.
- 3) 추출된 키워드들 중 동의어, 반의어, 모호한 표현, 일반적 용어들을 분석에서 제외하거나 교정하는 작업을 의미함.
- 4) 특정키워드(본 연구에서는 OLED)의 출현 빈도수가 나머지 키워드들보다 월등히 많아 전체 빈도수의 50% 이상을 차지할 경우 다른 키워드들간의 관계(본 연구에서는 유사성)를 왜곡시킬 수 있다는 것으로, 이를 해결하기 위해서 통상적으로 과반수 이상의 키워드를 분석에서 제외시킴.

〈표 4-3〉 OLED 관련 키워드들의 동시출현 빈도수에 의한 유사도 행렬

	VAR00001	VAR00002	VAR00003	VAR00004	VAR00005	VAR00006	VAR00007	VAR00008	VAR00009
VAR00001	.000	-.004	-.005	-.003	-.004	-.004	-.003	-.004	-.005
VAR00002	-.004	.000	.031	-.008	-.009	.034	-.007	-.009	.024
VAR00003	-.005	.031	.000	.096	.041	.094	.041	-.011	.028
VAR00004	-.003	-.008	.096	.000	.065	.183	-.006	-.008	.072
VAR00005	-.004	-.009	.041	.064	.000	.084	.062	.023	.006
VAR00006	-.004	.034	.094	.183	.084	.000	.044	-.009	.039
VAR00007	-.003	-.007	.041	-.006	.062	.044	.000	-.007	-.008
VAR00008	-.004	-.009	-.011	-.008	.023	-.009	-.007	.000	-.011
VAR00009	-.005	.024	.028	.072	.006	.039	-.008	-.011	.000
VAR000010	.000	.001	.157	.525	.120	.306	.022	.008	.018
VAR000011	-.006	-.013	.011	.045	.015	.001	.034	-.013	-.008
VAR000012	-.004	-.009	.396	.067	.052	.027	-.007	-.009	.028
VAR000013	-.004	-.010	.023	-.008	.022	.037	.069	-.009	-.005
VAR000014	-.005	.011	.034	.007	.017	.038	.005	.010	.024
VAR000015	-.004	.021	.021	.007	.035	-.008	-.006	.056	.013
VAR000016	-.003	-.006	-.007	-.005	.012	-.006	-.005	-.006	-.007
VAR000017	-.003	-.007	-.008	.021	-.006	-.006	-.005	.033	.003
VAR000018	-.003	-.007	-.008	-.006	.043	-.006	-.005	-.018	-.008
VAR000019	-.004	.016	-.010	.017	.015	.009	-.006	.009	.046
VAR000020	-.004	-.010	-.012	.000	.007	-.001	-.007	.008	.017

한 매트릭스를 만들었다. 그리고 고 빈도 출현키워드들과 저 빈도 출현 키워드들 간의 출현빈도 차이를 표준화하기 위하여 이를 정규화 한 유사도 행렬을 작성하였다.

정규화 유사도 행렬의 작성에는 피어슨(Pearson) 상관계수를 사용하였다. 피어슨 상관계수는 두 키워드들의 유사성을 측정하기 위해서 두 키워드와 동시출현행렬에 출현한 다른 키워드들과의 동시출현빈도를 이용하여 구해지고 +1에서 -1까지의 값을 갖는다. 만약 두 키워드들이 완전한 정(+)의 상관관계에 있다면 상관계수는 +1이며 완전한 음의 상관관계에 있다면 그 값은 -1을 갖게 되는 것이다

〈표 4-2〉는 OLED 관련 키워드들의 동시출현 빈도수를 나타낸 행렬이고 〈표 4-3〉은 피어슨 유

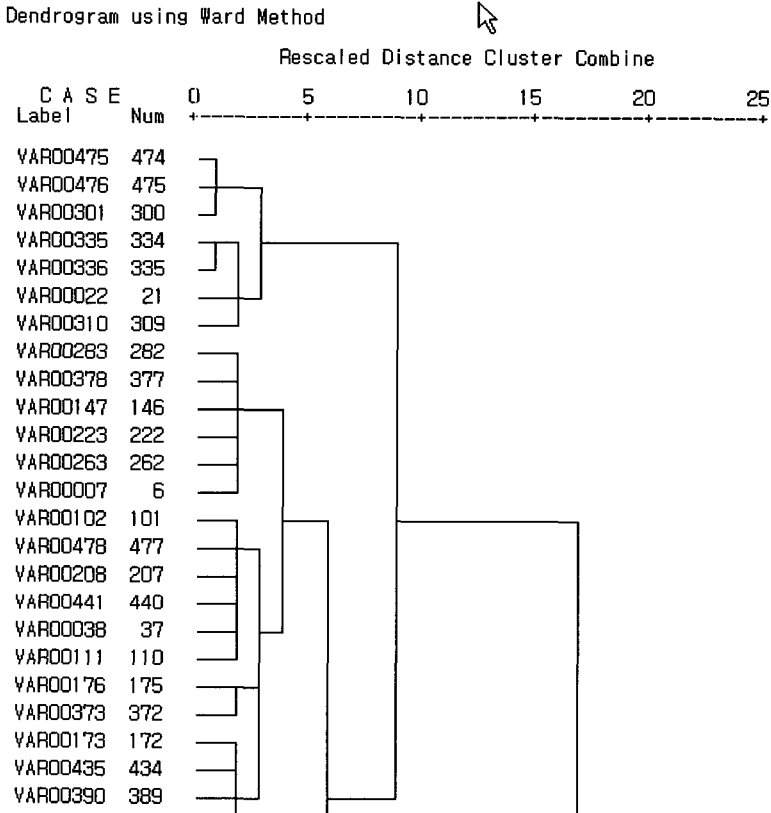
사계수를 사용하여 정규화한 OLED 관련 키워드들의 유사도 행렬의 일부를 나타낸 것이다.

〈표 4-2〉에 나타난 VAR0001, VAR0002, 등은 분석대상으로 사용된 각 키워드들을 의미한다. 세로축의 1, 2, 3, ... 은 각각 가로축의 VAR0001, VAR0002, VAR0003, ... 과 같은 의미를 갖는다. VAR0001과 VAR0001이 만나는 동시 출현 빈도수 1은 VAR0001을 포함하는 논문의 수를 말하는 것이고, VAR0004와 VAR0010이 만나는 동시 출현빈도수 3은 키워드 VAR0004와 VAR0010이 한 논문에 동시에 발생하는 빈도가 3임을 의미한다.

〈표 4-2〉와 마찬가지로 〈표 4-3〉에 나타나 VAR0001, VAR0002, 등은 분석대상으로 사용된 각 키워드들을 의미하는 것이다.



\*\*\*\*\* H I E R A R C H I C A L C L U S T E R A N A L Y S I S \*\*\*\*\*



<그림 4-1> OLED Ward 클러스터링 결과

### 4.3 클러스터링

4.2절에서 산출된 키워드들 간의 유사도 행렬을 가지고 Ding 등[9]의 연구에서 사용되었던 계층적 클러스터링(Hierarchical clustering) 기법인 Ward 법을 사용하여 클러스터링 하였다. Ward clustering 결과로 산출된 Dendrogram을 <그림 4-1>에 나타내었다. 이러한 클러스터링 결과 20개의 클러스터가 형성 되었으며<sup>5)</sup> 각각의 클러스터에 명명된

클러스터명은 OLED 관련 세부 연구분야를 의미하는 것이다. 일반적으로 클러스터명을 정의하는 방법으로는 높은 출현빈도수를 갖는 키워드가 그 클러스터를 대표하는 방법과 클러스터에 속한 전체 키워드를 대표할 수 있는 하나 혹은 그 이상의 단어를 사용하여 클러스터명을 부여하는 방법이 있는데 본 연구에서는 후자의 방법을 사용하여 클러스터명을 지정하였다[9].

이렇게 선정된 각각의 클러스터명(OLED 관련 세부연구분야)과 클러스터들을 대표하는 높은 출현 빈도수를 갖는 키워드들을 부록의 <표 1>에 나타내었다.

세부연구분야에 대한 분석결과 OLED 분야는 크게 시스템 연구분야와 요소 연구분야로 나누어짐

5) 클러스터를 20개로 설정한 이유는 Ward기법에 의해 제시된 cluster distance를 6으로 setting 했기 때문임. cluster distance를 6으로 setting한 이유는 이 기준으로 클러스터링하는 것이 OLED를 가장 잘 표현하기 때문임.

을 알 수 있는데 시스템 연구분야는 관련 기술을 모두 포함하고 있는 통합적인 연구분야이고 요소 연구분야는 공정기술, 설계기술 및 분석기술 등의 특화된 연구분야를 의미하는 것이다.

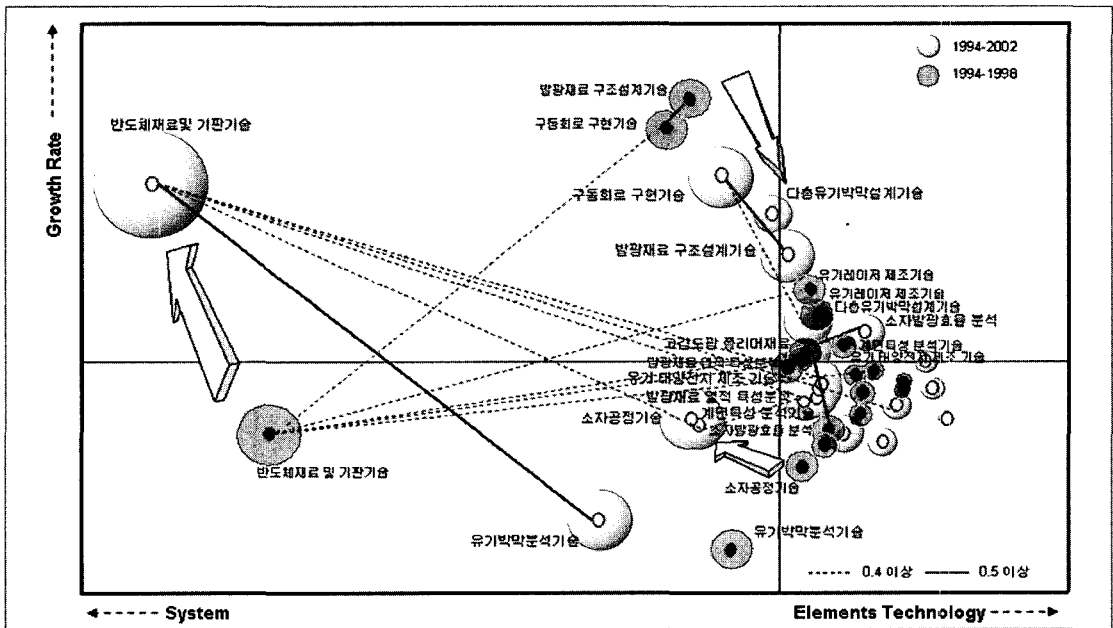
### 4.4 맵핑(Mapping)

4.3절에서 정의된 20개의 클러스터들의 상관관계를 산출해 내고 맵핑을 통해서 세부 연구분야 간 구조와 상관관계를 파악하기 위해서 각 클러스터들이 한 논문에 동시에 발생하는 빈도수를 구하였다.

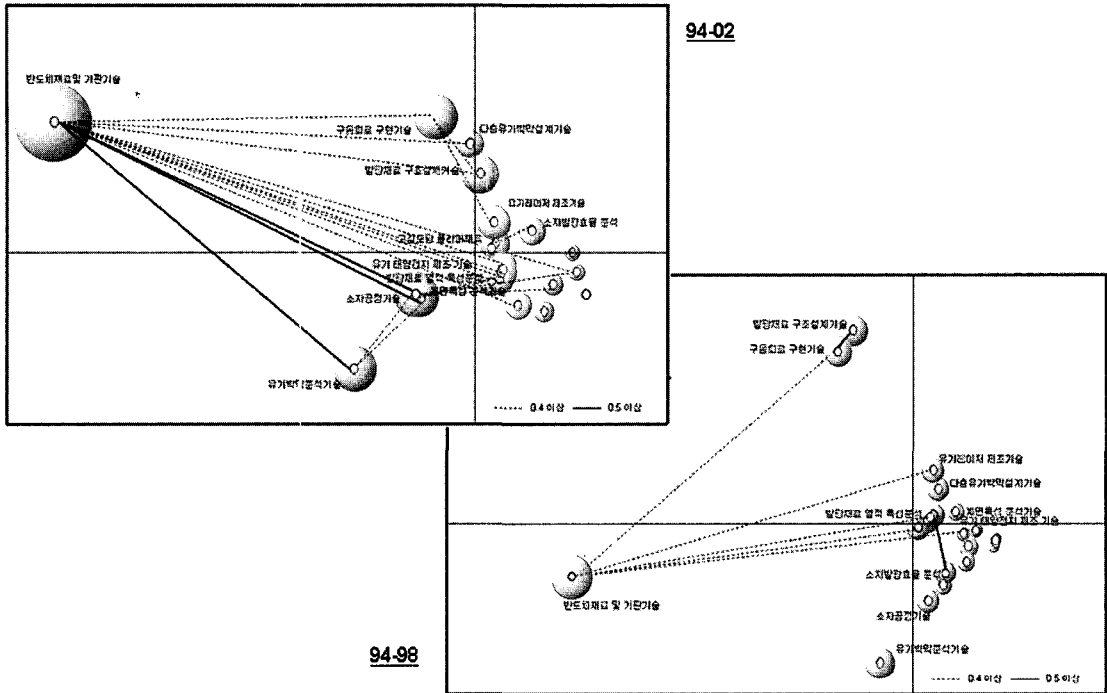
이렇게 산출되어진 클러스터들의 동시발생 빈도수는 이후 클러스터 간 상관도를 측정하기 위한 자료로 쓰이게 된다. 각각의 클러스터들이 2차원 지도 안에서 어느 곳에 위치하고 각 클러스터간의 관계들을 시각화 할 수 있는 맵핑을 하기 위해서는 다양한 유사계수들이 사용되어지는데 본 연구에서는 키워드들 간 유사도 행렬의 정규화에 사용하였던 피어슨 상관계수를 사용하였다. 부록의 <표 2>는 클러스터 별 동시발생 빈도 행렬을 나타낸 것이

고 부록의 <표 3>은 산출된 클러스터간의 피어슨 상관계수 값을 나타낸 것이다.

<표 3>의 결과에서 보면, 클러스터 5(유기박막 분석기술)와 클러스터 10(반도체 재료 및 기관기술), 클러스터 8(소자공정기술)과 클러스터 10, 클러스터 10과 클러스터 19(고분자 발광재료 합성기술) 사이의 유사도가 0.5 이상의 유사도를 보여 강한 상관관계에 있는 것으로 나타났으며 나머지 클러스터들 간의 유사도는 0.5 이하의 상관관계를 보였다. 특히 클러스터 10과 다른 클러스터 간의 상관관계가 높게 나타났는데 이는 클러스터 10의 연구분야가 여러 가지 관련 연구분야가 통합된 시스템적인 연구분야임을 의미하는 것이다. 또한 분석 결과 대부분의 클러스터간의 유사도가 낮게 나타났다는 것은 각 연구분야가 높은 연관관계를 가지며 연구가 진행되고 있는 것이 아니고 각각의 연구분야가 특성화 되어 독립적인 특성을 지니며 진행되고 있음을 의미하는 것이다. 이렇게 산출된 클러스터 간 피어슨 상관관계를 이용하여 클러스터(연구분야) 간 구조와 상관관계를 파악할 수 있도록 <그



<그림 4-2> OLED 클러스터(연구분야)의 시기별 도메인 특성비교를 위한 MDS 맵



〈그림 4-3〉 OLED 연구분야의 시기별 상호 연관강도의 변화를 나타낸 MDS 맵

림 4-2>와 같은 2차원 상의 다차원축적지도(Multidimensional scaling; MDS)를 구하였다.

<그림 4-2>와 <그림 4-3>에서 클러스터의 크기는 클러스터에 속한 키워드들의 출현빈도수가 전체 클러스터에서 차지하는 비율, 즉 키워드들이 출현한 문헌의 수를 의미하는 것으로 클러스터 하나의 원의 크기를 기준으로 나머지 클러스터들의 크기를 비례적으로 묘사하였다. 또한 여기서 클러스터 간의 연결은 0.5수준에서 유사도를 표시하는 것이 효율적으로 OLED 분야에서의 단어들간의 유사도 정도를 효과적으로 표현할 수 있다는 전문가들의 의견을 수용하여 피어슨 상관계수가 0.5 이상인 것은 실선으로 0.4 이상인 것은 점선으로 나타내었다.

OLED 관련 연구분야에 대한 클러스터(연구분야)별 MDS 맵을 1994~1993년도와 1994~2002년도 두 개의 시기로 나누어 비교 분석하여 본 결과 <그림 4-3>, 1994~1998년도의 연구기간에는 반도체재료 및 기판기술에 대한 연구분야의 성장률이

매우 낮고 다른 연구분야와의 연관강도도 매우 낮았음을 알 수 있다. 또한 요소 연구분야 부분도 각각의 연구분야가 특성화 되지 않고 혼재된 구조를 보이고 있음을 알 수 있다.

그러나 1990년대 후반부터 2002년까지는 OLED 연구가 기술적 성장기에서 성숙기로 돌입하는 시기로 연구분야의 구조적인 측면에서 1994~1998년도와 비교하여 볼 때 반도체 재료 및 기판기술, 구동회로 구현기술, 소자 공정분야 등 응용 연구분야의 성장이 두드러진 반면에 발광재료 구조설계 연구분야 등 기초 연구분야의 성장률은 감소한 것으로 나타났다.

OLED 연구분야에서 반도체 재료 및 기판기술 연구분야가 가장 두드러진 성장을 보였다는 것은 유기소자의 연구가 기존의 단순한 EL 현상만을 이용한 소자제작에서 기존의 LED 분야의 기술을 접목하여 본격적인 OLED 개념의 소자에 대한 연구가 이루어졌음을 의미하는 것이다. 이러한 반도체

재료 및 기관기술 연구분야가 가장 두드러진 성장은 1990년대 반도체 기관을 이용한 LED 기술의 급성장을 그 원인으로 들 수 있다.

구동회로 구현기술 연구분야의 성장률이 두드러진 것은 EL 기술에 TFT-LCD 기술이 접목된 것으로 이는 OLED 디스플레이 연구가 대면적화를 이루는 방향으로 활성화 되고 있음을 의미하는 것이다.

<그림 4-3>에서 알 수 있듯이 연구분야 간 상호연관관계 측면에서는 1994~2002년 시기에는 반도체 재료 및 기관기술 연구분야를 중심으로 각 연구분야의 연관강도가 강해지고 요소기술 연구분야가 특성화되고 있음을 알 수 있다.

반도체 재료 및 기관기술 연구분야와 유기박막 분석 연구분야의 연결강도가 강해진 이유는 반도체 재료 및 기관기술 분야에서 OLED와 관련된 대표적인 기술은 유기물질의 증착과 화합물반도체에 대한 부분이기 때문에 이러한 박막들의 특성을 측정하기 위한 유기박막분야와 밀접한 관련을 맺고 있는 것으로 판단된다. 또한 반도체 재료 및 기관기술 연구분야는 OLED와 관련된 거의 모든 연구분야와 강한 연관관계를 맺고 있는데 이는 반도체 재료 및 기관기술 연구분야가 어느 특정한 기술만을 나타내는 요소기술분야가 아닌 여러 가지 관련기술을 통합한 시스템의 성격을 지닌 연구분야임을 의미하는 것으로 세계적으로 우수한 반도체 기술을 보유하고 있는 우리나라로서는 OLED 연구분야의 기술경쟁력에 있어서도 매우 유리한 위치에서 있음을 나타내는 것이다.

## 5. 결 론

본 연구에서 개발한 지식맵은 신기술의 사전 타당성 평가에 매우 유용한 정보와 객관적인 근거를 제공할 수 있었으며, 현재의 전문가 회의(Peer review)방식과 지식맵 방식을 통합하는 형태의 새로운 사전 조정 방식은 매우 유용할 것으로 판단된다. 지식맵은 빠르게 변화하는 국내외 기술환경과

관련된 다양한 변수에 관한 객관적 근거를 제공할 수 있다. 또한 지식맵은 신기술의 복잡한 기술구조와 연관도를 용이하게 파악함으로써 전문가의 이해 및 해석 능력을 크게 증대시킬 수 있다. 기술융합이 가속화되고 있는 신기술 분야의 평가에 있어서 이(異)분야 전문가간 의사소통을 원활하게 해주며 광범위한 이해관계자들(정책, 경제, 경영, 일반인 등)이 참여하는 사용자 평가(Merit evaluation)를 가능하게 한다.

지식맵 작성 과정을 전산화하여 신속하게 제공한다면 단기간내에 많은 과제를 평가해야 하는 예산 사전조정 심의에 유효할 것이다. 일단 한번 작성된 지식맵에 대해 추가적인 Update를 통해 해당 기술의 지속적인 변화를 추적할 수 있기 때문에 다년도 과제의 중간평가, 성과평가에도 재차 활용할 수 있을 것이다.

이러한 연구결과를 바탕으로 향후에 연구를 발전시키기 위해서는 다음과 같은 연구가 필요하다. 먼저, 지식맵 작성 S/W Tools 개발을 통해 차세대 성장동력산업관련 등 핵심 신기술에 대한 지식맵을 작성, 보급해야 할 것이다. 다음으로 High quality TRM, 국제협력지도, 기술분류체계 작성에 활용되어야 할 것이다. 마지막으로 세계의 연구개발 프론티어에 대한 지속적인 지식맵 작성 및 축적으로 우리나라 신기술 분야의 취약점 및 기술기회를 조기에 알려주는 과학기술 조직 모니터링시스템(S&T early monitoring system)을 구축해야 할 것이다.

## Reference

- [1] Bhattacharya, S. and Basu, R. K., "Mapping a research area as the micro level using co-word analysis," *Scientometrics*, Vol.43, No.3(1998), pp.359-372.
- [2] Callon, M., Courtial, J.P. and 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(1991), pp.153-203.
- [3] Callon, M., Law, J., and Rip, A.(eds), "*Mapping the Dynamics of Science and Technology : Sociology of Science in the Real World*," London: Macmillan., 1986.
- [4] Cambrosio, A., Limoges, C. Courtial, J.P. and Laville, F., "Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis," *Scientometrics*, Vol.27, No.2(1993), pp.119-143.
- [5] Coulter, N., Monarch, I. and 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(1998), pp.1206-1223.
- [6] Courtial, J.P., "A cword analysis of scientometrics," *Scientometrics*, Vol.31, No.3 (1994), pp.251-260.
- [7] Courtial, J.P., Cahlik, T. and Callon, M., "A model for social interaction between cognition and action through a key-word simulation of knowledge growth," *Scientometrics*, Vol.31, No.2(1994), pp.173-192.
- [8] Ding, Y., Chowdhury, G.G. and Foo, S., "Journal as markers of intellectual space : Journal co-citation analysis of Information Retrieval area," *Scientometrics*, Vol.47, No.1 (2000).
- [9] Ding, Y., Chowdhury, G.G. and Foo, S., "Bibliography of information retrieval research by using co-word analysis," *Information Processing & Management*, Vol.37 (2001), pp.817-842.
- [10] He, Q., "Knowledge Discovery Through Co-word Analysis," *LIBRARY TRENDS*, Vol.48, No.1(1999).
- [11] Hinz, S., "Bibliographical cartography of an emerging interdisciplinary discipline : the case of bioelectronics," *Scientometrics*, Vol.29, No.3(1994), pp.353-376.
- [12] Honkela, T., Kaski, S. Lagus, K. and Kohonen, T., "Newsgroup exploration with WEBSOM method and browsing interface," Report A32, *Laboratory of Computer and Information Science*, Helsinki University of Technology, Finland., 1996.
- [13] King, J., "A review of bibliometric and other science indicators and their role in research evaluation," *Journal of Information Science*, Vol.13(1987), pp.261-276.
- [14] Kohonen, T., "*Self-Organizing Maps*", Springer, Berlin, Heidelberg, 1995.
- [15] Kopcsa, A. and Schiebel, E., "Science and technology mapping : A new iteration model for representing multidimensional relationships," *Journal of the American Society for Information Science*, Vol.49, No.1(1998), pp.7-17.
- [16] Law, J. and Whittaker, J., "Mapping acidification research : A test of the co-word method," *Scientometrics*, Vol.23, No.3(1992), pp.7-17.
- [17] Looze, M.D. and Lemarie, J., "Corpus relevance through co-word analysis : An application to plant proteins," *Scientometrics*, Vol.39, No.3(1997), pp.267-280.
- [18] Noyons, E.C.M. and Van Raan, A.F.J., "Bibliometric cartography of scientific and technological development of an R&D field," *Scientometrics*, Vol.30, No.1(1994), pp.157-173.
- [19] Noyons, E.C.M. and Van Raan, A.F.J., "Advanced mapping of science and technology," *Scientometrics*, Vol.41, No.1-2(1998),

pp.61-76.

- [20] Peters, H.P.F. and van Raan, A.F.J., "Co-word based science maps of chemical engineering, Part I : Representations by direct multidimensional scaling," *Research Policy*, Vol.22(1993), pp.23-45.
- [21] Polanco, X., Francois, C. and Keim, J.P., "Artificial neural network technology for the classification and cartography of science and technology information," *Scientometrics*, Vol.41, No.1-2(1998), pp.69-82.
- [22] Rikken, P., Kiers, H.A.L. and Vos, R., "Mapping the dynamics of adverse drug reactions in subsequent time periods using INDSCAL," *Scientometrics*, Vol.33, No.3(1995), pp.367-380.
- [23] Rip, A. and Courtial, J.P., "Co-word maps of biotechnology : An example of cognitive scientometrics," *Scientometrics*, Vol.6 (1984), pp.381-400.
- [24] Tijssen, R.J.W., "A scientometric cognitive study of neural network research : expert mental maps versus bibliometric maps," *Scientometrics*, Vol.28, No.1(1993), pp.111-136.
- [25] Tijssen, R.J.W. and A.F.J. van Raan, "Mapping co-word structures : a comparison of multidimensional scaling and leximappe," *Scientometrics*, Vol.15(1989), pp.283-295.
- [26] Van Rann, A.F.J. and Tijssen, R.J.W., "The neural net of neural network research," *Scientometrics*, Vol.26, No.1(1993), pp.169-192
- [27] Voutilainen, A., "NPtool. A detector of English noun phrases. In : Proceedings of the workshop on very large corpora," Columbus, Ohio : Ohio State University, June 22, 1993.
- [28] Widhalm, C., "Personal Communication," Nov., 1999.

## 부 록

〈표 1〉 클러스터명 및 클러스터별 주요 키워드

클러스터	기술 분류	KEY WORD
1	양극기판(ITO) 처리기술	Oxide, Plasma, Treatment, Anode, Performance
2	발광효율 기능 개선 기술	TPD, Layer, Dopant, Hole, Transport material, High Efficiency, Spin coating
3	다층유기박막설계기술	Transport Measurement, Copper Phthalocyanine, Hole Injection Layer, Insulator, Hole Transport Layer, Transport layer, Injection, Transport, Electron, Efficiency
4	유기박막형성기술	PAH, Self-assembly process, By-layer manipulation, Adsorption
5	유기박막분석기술	X-ray Photoelectron spectroscopy, Molecular structure, Materials, Crystal structure, X-ray, X-ray diffraction, structure, Crystal, Organic material
6	인광발광재료	Display, EU, TB, Phosphor, Phosphorescence
7	저분자 진공 증착 기술	Deposition process, Structural properties, Thermal evaporator, Vacuum, Metal shadow mask, Small molecular
8	소자공정기술	Experimental, Microscope, Indium Tin Oxide (ITO), Processing, Thin film devices, Lithography, G갯소, Junction, Hetero structures, Hydride, Glass, Surface
9	계면특성 분석 기술	ATOMIC FORCE MICROSCOPY(AFM), SCANNING ELECTRON MICROSCOPY (SEM), INTERFACE FORMATION, SCANNING TUNNELING MICROSCOPY (STM)
10	반도체 재료 및 기판 기술	Liquid crystalline (LC), Electroluminescence material, Carbon oxide, Semiconductor materials, Epitaxy process, Conductivity materials, Microcavity structure, Nitride, Quantum well, Silicon oxide, LANGMUIR-BLODGETT, LB films, Buffer layer, Photooxidation, Hetero junction, Fabrication, Multilayer
11	발광재료 구조설계 기술	Computer simulation, Model, HUBBARD model, MANY-BODY AND QUASI-PARTICLE Theories, Polymer chains, Structure property relationship, Design
12	유기 태양전지 제조 기술	Photoinduced charge transfer, Organic photovoltaics, Plastic solar cells, Photovoltaic cell, Solar cell, TiO <sub>2</sub> , Conversion, Device performance, Control
13	구동회로 구현 기술	Alignment, Spin cast, Matrix, Circuit, TFT, OTFT, AM matrix, PM matrix, Modulation, Optoelectric device
14	재료구조분석	Magnetic resonance, Optically detected magnetic resonance, ODMR, Electron paramagnetic resonance, Spin resonance, Magnetic properties, Resonance
15	고감도 광 폴리머 재료	Sensor, Photoisomerization, Size, Photolysis, Migration, Annihilation, Naphthalene, Localization
16	유기 레이저 응용 기술	Femtosecond, Dynamics, Singlet exciton, Laser flash photolysis, Laser, Stimulated emission, Stimulated luminescence, Laser-dye, Wave-guide, Electrooptic, Proton transfer
17	소자발광효율 분석	LIFETIME, PHOSPHORESCENCE LIFETIME, LUMINESCENCE PROPERTIES, DECAY, PROBE, QUENCHING
18	광 기능성 유기/금속 중합기술	Reactivity, Ligand, Coordination, Complexes, Metal complexes, Terbium complex, Emitting layer, Chelate complexes, Substituent
19	고분자 발광재료 합성기술	Molecular-orbital Method, Derivatives, Length, Light, Poly, Side groups, Polymer synthesis, Blue, Electrochemical property, Model compounds
20	발광재료 열적 특성분석	Metathesis, Catalysts, Transition metal catalysts, DSC, TGA, Coupling reactions, Transition-metal-catalyzed reactions, Chemistry, Cross-coupling reactions

〈표 2〉 클러스터 별 동시발생 빈도 행렬

	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
cluster1	205	40	71	5	78	25	30	94	52	93
cluster2	40	936	296	20	227	91	40	225	120	520
cluster3	71	296	1688	26	364	82	60	372	248	798
cluster4	5	20	26	151	55	6	21	79	41	98
cluster5	78	227	364	55	2033	183	136	516	275	1110
cluster6	25	91	82	6	183	584	59	143	48	298
cluster7	30	40	60	21	136	59	365	155	60	206
cluster8	94	225	372	79	516	143	155	1704	306	896
cluster9	52	120	248	41	275	48	60	306	860	462
cluster10	93	520	798	98	1110	298	206	896	462	3779
cluster11	13	95	346	16	393	35	38	251	165	685
cluster12	19	95	181	17	275	22	26	218	104	422
cluster13	28	178	369	22	448	41	61	399	184	840
cluster14	0	2	5	0	9	0	0	3	0	13
cluster15	26	87	240	27	311	63	38	208	86	522
cluster16	18	119	246	14	227	44	35	253	101	533
cluster17	28	64	179	11	189	71	20	130	51	406
cluster18	14	102	136	10	276	115	12	90	48	393
cluster19	47	221	342	25	488	72	51	333	208	917
cluster20	23	99	194	22	325	59	39	184	121	575

	cluster11	cluster12	cluster13	cluster14	cluster15	cluster16	cluster17	cluster18	cluster19	cluster20
cluster1	13	19	28	0	26	18	28	14	47	23
cluster2	95	95	178	2	87	119	64	102	221	99
cluster3	346	181	369	5	240	246	179	136	342	194
cluster4	16	17	22	0	27	14	11	10	25	22
cluster5	393	275	448	9	311	227	189	276	488	325
cluster6	35	22	41	0	63	44	71	115	72	59
cluster7	38	26	61	0	38	35	20	12	51	39
cluster8	251	218	399	3	208	253	130	90	333	184
cluster9	165	104	184	0	86	101	51	48	208	121
cluster10	685	422	840	13	522	533	406	393	917	575
cluster11	1537	138	493	12	231	253	93	101	339	229
cluster12	138	814	272	5	115	142	103	40	221	78
cluster13	493	272	1812	19	301	402	202	109	382	214
cluster14	12	5	19	40	19	6	11	7	4	4
cluster15	231	115	301	19	1109	203	315	137	183	136
cluster16	253	142	402	6	203	1196	133	56	278	151
cluster17	93	103	202	11	315	133	891	143	128	89
cluster18	101	40	109	7	137	56	143	716	205	175
cluster19	339	221	382	4	183	278	128	205	1672	360
cluster20	229	78	214	4	136	151	89	175	360	1038



〈표 3〉 피어슨 상관계수를 이용한 클러스터 간 유사도 행렬

	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8	cluster9	cluster10
cluster1	1	0.155636	0.224689	0.064977	0.248973	0.076293	0.189075	0.353615	0.237243	0.254209
cluster2	0.155636	1	0.380924	0.123376	0.271202	0.180305	0.112695	0.292464	0.226676	0.469469
cluster3	0.224690	0.380924	1	0.109969	0.288038	0.078344	0.106607	0.324982	0.331609	0.473021
cluster4	0.064977	0.123376	0.109969	1	0.291131	0.037328	0.233946	0.436578	0.288547	0.404458
cluster5	0.248973	0.271202	0.288038	0.291131	1	0.279523	0.321129	0.429198	0.353695	0.599348
cluster6	0.076293	0.180305	0.078344	0.037328	0.279523	1	0.211838	0.216236	0.066037	0.350751
cluster7	0.189075	0.112695	0.106607	0.233946	0.321129	0.211838	1	0.399961	0.204163	0.379118
cluster8	0.353615	0.292464	0.324982	0.436578	0.429198	0.216236	0.399961	1	0.429889	0.535167
cluster9	0.237243	0.226676	0.331609	0.288547	0.353695	0.066479	0.204163	0.429889	1	0.451253
cluster10	0.254209	0.469469	0.473021	0.404458	0.599348	0.350751	0.379118	0.535167	0.451253	1
cluster11	-0.05365	0.108372	0.323016	0.044783	0.340648	-0.02139	0.047779	0.228190	0.224596	0.437328
cluster12	0.025981	0.173247	0.246234	0.104017	0.367661	-0.01482	0.065130	0.315212	0.209332	0.434644
cluster13	0.004567	0.196504	0.312175	0.078552	0.348792	-0.01707	0.099492	0.335217	0.228610	0.478211
cluster14	-0.29739	-0.14015	-0.03661	-0.21429	0.037699	-0.19877	-0.22092	-0.10197	-0.18996	0.084161
cluster15	-0.01811	0.076765	0.233610	0.080948	0.297741	0.049018	0.024256	0.189252	0.075735	0.390391
cluster16	-0.04612	0.146984	0.250500	0.020233	0.208260	-0.01587	0.028305	0.253482	0.128060	0.391006
cluster17	-0.01595	0.043339	0.173478	-0.02074	0.181238	0.083779	-0.04133	0.099264	-0.00171	0.329696
cluster18	-0.03294	0.179894	0.160775	0.014493	0.387803	0.275753	-0.00991	0.096258	0.053114	0.422682
cluster19	0.118722	0.294957	0.310189	0.129010	0.424681	0.088574	0.107690	0.307501	0.293434	0.565247
cluster20	0.014351	0.158946	0.217328	0.109442	0.378168	0.087762	0.080024	0.210917	0.201804	0.490577

	cluster11	cluster12	cluster13	cluster14	cluster15	cluster16	cluster17	cluster18	cluster19	cluster20
cluster1	-0.05365	0.025981	0.004566	-0.29739	-0.01810	-0.04611	-0.01595	-0.03293	0.118721	0.014350
cluster2	0.108372	0.173246	0.196504	-0.14015	0.076765	0.146983	0.043339	0.179893	0.294957	0.158946
cluster3	0.323016	0.246234	0.312175	-0.03661	0.233609	0.250500	0.173478	0.160775	0.310189	0.217327
cluster4	0.044783	0.104016	0.078551	-0.21429	0.080948	0.020232	-0.02074	0.014493	0.129009	0.109441
cluster5	0.340647	0.367661	0.348792	0.037698	0.297741	0.208259	0.181237	0.387803	0.424680	0.378168
cluster6	-0.02139	-0.01482	-0.01706	-0.19877	0.049017	-0.01587	0.083779	0.275753	0.088574	0.087762
cluster7	0.047779	0.065129	0.099492	-0.22091	0.024256	0.028305	-0.04133	-0.00991	0.107690	0.080023
cluster8	0.228190	0.315212	0.335217	-0.10197	0.189252	0.253481	0.099264	0.096258	0.307501	0.210917
cluster9	0.224595	0.209331	0.228610	-0.18996	0.075734	0.128059	-0.00171	0.053114	0.293434	0.201804
cluster10	0.437327	0.434644	0.478211	0.084161	0.390391	0.391005	0.329695	0.422682	0.565247	0.490577
cluster11	1	0.216859	0.471309	0.167197	0.257286	0.303046	0.076511	0.126184	0.340632	0.297789
cluster12	0.216859	1	0.385185	0.004667	0.169298	0.236380	0.140658	0.049574	0.328703	0.140289
cluster13	0.471309	0.385185	1	0.278444	0.317405	0.439453	0.209668	0.111007	0.342396	0.244159
cluster14	0.167197	0.004667	0.278444	1	0.329987	0.048593	0.189487	0.020500	-0.04801	-0.04462
cluster15	0.257286	0.169298	0.317405	0.329987	1	0.248185	0.491238	0.212875	0.180582	0.163638
cluster16	0.303046	0.236380	0.439453	0.048593	0.248185	1	0.152312	0.032776	0.303160	0.196204
cluster17	0.076511	0.140658	0.209668	0.189487	0.491238	0.152312	1	0.246672	0.113958	0.088976
cluster18	0.126184	0.049574	0.111007	0.020500	0.212875	0.032776	0.246672	1	0.316654	0.345425
cluster19	0.340632	0.328703	0.342396	-0.04801	0.180582	0.303160	0.113958	0.316654	1	0.467806
cluster20	0.297789	0.140289	0.244159	-0.04462	0.163638	0.196204	0.088976	0.345425	0.467806	1