

Co-word 분석법을 이용한 기업수준의 기술지도: 작성 방법과 활용 전략

김혜리, 박용태

서울대학교 공과대학 산업공학과

초 록

특허 데이터 정보는 경쟁자 모니터링, 기술 평가, R&D 포트폴리오 관리 등 전략적 계획의 여러 목적으로 사용되고 있다. 특정 분야의 기술 구조를 시각화하기 위한 방법론으로 특허를 사용한 기술 지도를 2차원상에 표현하는 co-word 분석법이 이용되고 있으나, 이 방법론을 적용하여 기업이 가진 특허를 지도로 시각화하고 분석을 시도한 연구는 없었으며 특허 포트폴리오는 전체 기술을 시각화하는 데 많은 한계를 가지고 있다.

본 연구에서는 특정 기업의 일부 분야 특허에 대하여 co-word matrix를 작성하고 기술을 클러스터링하여 다차원 축소 기법(Multidimensional Scaling)으로 기술 지도를 작성하였다. 또한, 두 경쟁기업의 기술지도를 비교하여 기업간의 cross-licensing 가능 영역을 보이고 시간 흐름에 따른 기업 기술의 동적 분석 등 다양한 분석방법을 제시하였다.

1. 서 론

기술의 발전과 함께 연구개발 관리 기법도 함께 발전하고 있다. 기술 연구개발 환경을 모니터링하고 타기업과의 경쟁 상황에서 기술의 변화를 예측하는 기술 모니터링 시스템(technology monitoring system)은 기업의 사업 기회를 찾고 전략적 의사결정을 내리는 데 유용하다 (Ernst, 1998).

이러한 시스템에서 특허 데이터 정보는 경쟁자 모니터링, 기술 평가, 연구개발 포트폴리오 관리 등에 사용되어 전략적 계획을 세우는 데 중요한 자료가 된다(Ernst, 2003). 특허 특허 포트폴리오는 기술 매력도(technology attractiveness)와 기업의 특허 강도(relative patent position)을 두 축으로 특허를 기술 영역에 할당하고 각 영역의 중요도를 판단하여 자원을 분배할 수 있는 이점이 있어 관리자에게 중요한 수단이다.

그러나 특허 포트폴리오는 다른 기술 분야에서 경쟁하고 있는 기업의 강점(strength)과 약점(weakness)을 평가하여, 최종 의사결정을 내리는 데 요구되는 제한된 양의 정보만을 제공한다는 한계를 가지고 있다. 따라서 상급 관리자의 토론을 용이하게 하기 위해 복잡한 상황을 정리하는 데 유용한 틀이긴 하나, 중간관리자가 전체적인 기업 보유 기술을 구체적으로 시각화하여 정리할 수 있는 보조적인 방법이 필요하다.

서지상의 데이터 등 공간적으로 표현하기 힘든

추상적 개념을 시각화하는 정보 시각화(information visualization)는 정보를 표현하는 하나의 접근법으로 (Börner et al., 2003), 그 중에서 특정 기술 도메인의 구조를 시각화하는 방법론인 기술 지도학(technology cartography)이 발전해오고 있다 (Engelsman et al., 1994).

많은 연구자들이 다양한 분야의 영역 분석을 위해 키워드, 저자, 인용관계 등의 자료를 분석하여 2차원 상에 시각화하였다. Co-word 분석법은 주제 분야의 문헌에 속한 키워드나 분류코드 등의 동시출현 결과를 이용하여 매핑하는 방법으로 (Ding et al., 2001), 데이터분류 시스템에 의존하지 않고 분석이 가능하므로 보다 다양한 해석이 가능하다는 장점이 있으며 시각화와 검색에 활용할 수 있다.

본 연구는 co-word 분석법을 활용하여 기업 수준의 기술지도 작성에 응용하고, 지도 분석을 통한 전략적 활용방안을 마련하는 데 초점을 둔다. 구체적인 지도 활용 전략으로는 보유 기술의 분류와 기술 간의 연관성 파악, 그리고 기업간 cross-licensing이 가능한 기술 영역을 제안하고자 한다.

2. 배경이론

기술 지도학 방법론은 연구 분야의 구조를 표현하기 위해 저자나 저널의 동시인용분석 등을 제시하고 있다. 본 연구에서는 직접적인 연구 분야 표현을 위해서 co-word 분석을 이용한다.

Co-word 분석을 이용한 기존의 연구는 화학 분야 (Peters et al., 1993), 인공지능 분야 등의 도메인 분석에 이용되었을 뿐 아니라, Ying ding(2001)은 결과를 활용하여 온라인 데이터 베이스, 웹 정보를 검색하는 시스템을 제시하고, 키워드 클러스터에 접근하여서 각 클러스터의 결과를 확인할 수 있도록 하는 등 검색의 기능을 강화하였다.

Co-word 분석 방법의 순서는 크게 데이터 추출, 데이터 변환, 매핑 세가지의 단계로 구분된다.

첫 번째, 데이터 추출은 일정한 값(threshold) 이상의 출현 빈도를 가지는 키워드를 text-mining으로 추출한다. 이때 무의미한 단어나 중복되는 단어는 정리하여 중요한 키워드의 분산을 막고 결과의 오차를 줄일 필요가 있다.

두 번째, 데이터 변환 단계에서는 키워드 간의 동시출현 횟수를 측정하여 co-word 행렬을 작성한다. 행렬의 각 셀(cell)에 입력되는 $C(i, j)$ 는 키워드

i와 j를 동시에 가지고 있는 문서의 개수이다.

각 키워드 간의 근사도(similarity)를 계산하기 위해 co-word 행렬은 co-efficient 행렬로 변환된다. 피어슨 상관계수를 이용한 행렬은 매핑에 이용되며, 높은 출현빈도를 가진 단어들의 부정적인 영향력을 줄이는 정규화 방법을 적용하기 위해 두 단어의 직접적인 연관관계를 나타내는 지표인 salton index를 고려한다. C_x 는 키워드 x를 포함하는 문서의 전체 개수, C_{xy} 는 키워드 x와 y를 동시에 포함하는 문서의 전체 개수를 의미하며 계산방법은 다음과 같다 (Ding et al., 2001).

$$Sim(x, y) = \frac{C_{xy}}{\sqrt{C_x C_y}}$$

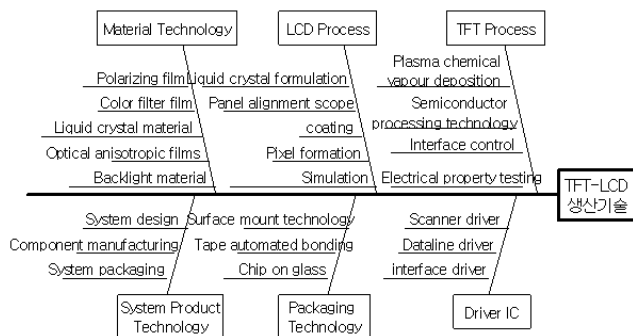
세 번째, 매핑 단계에서는 다차원 축척 기법(MDS)을 사용하여 행렬을 2차원 상에 표현한다. 키워드들만의 관계는 뚜렷한 시각화와 구조 분석에 한계를 가지므로 키워드의 조합인 클러스터의 의미를 파악하여 분야의 세부 부분을 구분할 수 있다 (Peters et al., 1993).

Co-word 분석은 클러스터링을 활용함으로써 여러 수준, 즉 전체적인 지도와 세부적인 지도의 작성이 가능하게 되었으며 이에 따라 접근 수준에 맞추어 기술을 검색할 수 있다.

3. 기술지도 작성과정 및 결과

3.1 데이터

2003년 세계 LCD 시장의 1, 2위로 각각 18%, 17.8%의 점유율을 차지한 우리나라의 두 대기업을 대상으로 실제 LCD 기술의 키워드를 2차원 상에 시각화하였다. LCD 산업은 Figure 1에서와 같은 기술 분류 구조를 가지고 있으며 오른쪽 방향의 기술이 발전하는 양상을 보이고 있다 (Hung, 2002).

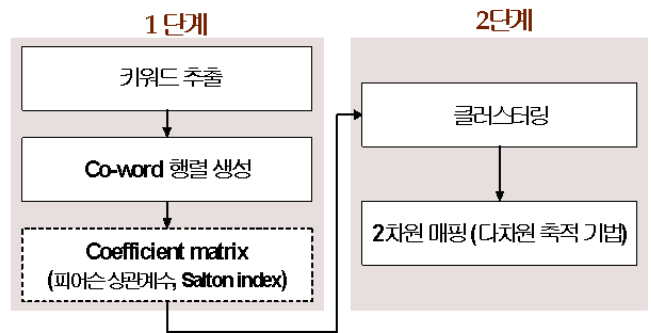


[Figure 1] LCD 산업의 기술 트리

키워드의 선택은, LCD 기술을 대부분 포함하고 있는 US class 349, 345에 속하고 초록에 'LCD' or 'Liquid Crystal Display'의 키워드를 가지고 있으며 2002년 6월에서 2004년 3월 사이에 출원된 것을 기준으로 하였다. 가장 최근의 같은 기간을 기준으로 둔 것은 두 기업의 가장 최근의 경쟁 상황을 살피고 비교하기 위한 것이다. 이같은 기준 아래 선택된 특허 문서는 S기업 100개, L기업 123개였다.

3.2 작성과정

기술 지도의 작성은 크게 키워드 추출과 co-word 행렬 작성, co-efficient 행렬 변환이 이루어지는 전처리 과정인 1 단계와 클러스터링과 매핑을 위한 2단계처리 과정으로 프레임워크를 구성하였다.



[Figure 2] 기술 지도 작성 프레임워크

3.2.1 1단계: 전처리 과정

Text-mining으로 특허 내의 키워드를 추출하는 첫 번째 단계는, 특허의 제목과 초록만을 선택하여 모든 문서에서 5회 이상 출현, 3개의 문장 이상에 소속되는 키워드를 선택하였다. 추출 후 주요 단어인 LCD는 삭제하고 기술과 직접 관련되지 않은 기본 단어 역시 제외하였다. 정리된 키워드는 각각 S기업 84개, L기업 88개였다. 이 중 두 기업에서 일치하는 키워드는 약 30개 정도로 나머지 약 50개가량은 독자적인 키워드이다. 이 단계에서 최근 기업의 역량이 집중된 부분과 소홀한 부분을 1차적인 검색하며, Table 1에서 서로 관심있는 키워드를 확인할 수 있다.

[Table 1] 두 기업의 공통적/독자적 키워드

S	공통	L
ITO angle liquid crystal panel photolithography photoresist lamp PCB mold frame ...	substrate data pad drain electrode gate pad mask passivation -layer ohmic ...	switching mode liquid capacitor domain liquid -capacitance transmissive mode transparentelectrode etching thickness ...

각 기업별로 84*84, 88*88의 co-word 행렬을 작성하고 피어슨 상관계수로 정규화한 co-efficient 행렬로 변환시켜 다음 2단계에서 키워드의 클러스터링을 실행하였다.

3.2.2. 2단계: 클러스터링 및 매핑

-클러스터링 클러스터링은 계층적 기법 중 데이터에 내재하는 클러스터의 구조를 잘 발견하는 것으로 검증된 Ward 기법을 사용하였으며 (Polanco et al., 2001), SPSS 11.5으로 수행하였다. 그 결과 두 기업의 키워드 클러스터는 Table 2와 같이 대표 키워드로 정리되었다.

[Table 2] 두 기업의 클러스터별 대표 키워드

클러스터	S	L
1	substrate data pad drain electrode	substrate pixel electrode switching
2	display panel voltage monitor	passivation Transistor transflective
3	mold-frame chassis wall	IP capacitance switching device
4	panel TFT array plurality	alignment domain liquid dielectric
5	display unit polarizing angle	transmissive region polarizer
6	font graphic interface	transverse direction connector longitudinal direction

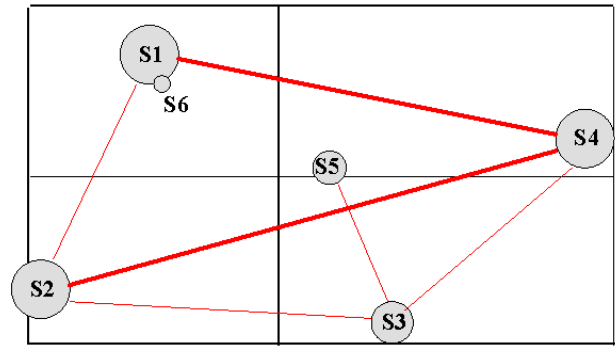
두 기업 모두 6개로 구분한 것이 가장 적당하였으며 각각의 클러스터는 출현 빈도가 높은 대표 키워드 3개로 표현하였다. 클러스터가 1:1로 매치되는 것은 아니며 클러스터 1의 경우에만 약 50%의 키워드가 일치하고 나머지 클러스터의 키워드는 30% 이상의 일치도를 보이지 않는다. 따라서 두 기업의 최근 기술 발전은 서로 다른 양상을 보이고 있다.

S, L 기업의 클러스터 1은 TFT 제조 후반 공정 기술로 가장 역량이 집중되어 있다. 각 기업에서 특이할 만한 것은, S기업에서는 클러스터 3이 차체 공정기술, 클러스터 6이 interface 기술에 관련된 키워드 집합으로 B에 비해 윤곽이 뚜렷하며, L기업에서는 전체적으로 display 기술과 회로 기술 등에 집중되어 있었으며 매칭되는 기술을 확인하기 어려운 클러스터(L6)도 있었다.

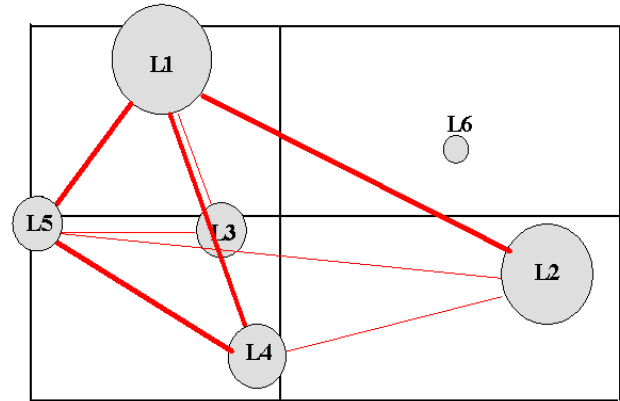
-매핑 각 클러스터들을 2차원 상에 매핑하기 위한 단계에서는 새로운 전처리 과정이 필요하다. 6*6의 새로운 클러스터 행렬에서 각 cell의 value는 양쪽 클러스터에 속한 키워드들을 양쪽에 적어도 하나씩 가진 특허문서의 개수가 된다. 두 기업의 클러스터를 매핑하는 것은 SPSS 11.5의 다차원 축적 기법을 이용하였다. 각 클러스터의 크기는 특허 문서의 갯수로 나타내며 salton index로 개개 클러스터 사이의 연관 정도는 선으로 표현하였다. 클러스터가 나타내는 기술 특허의 크기는 특허 포트폴리오의 기업 특허 강도를 나타낸다.

Figure 3에서 두 기업의 전체적 지도를 비교한 결과 비슷한 개수의 키워드를 가지고 있음에도 불구하고 L기업의 클러스터 크기가 더 크며 서로 연관성이 크게 나타났다. 이것은 각 분야의 관련 정도가 더 크다는 것을 의미하며, 상대적으로 S기업은 연구 분야가 다양하다고 말할 수 있다.

S1, S2, S4 및 L1, L2의 클러스터는 현재 기업의 역량이 집중된 발전 기술을 나타낸다. 또한 S6과 L6 등의 클러스터는 문서의 크기가 매우 작으므로 신생 기술 혹은 퇴보 기술로 분류할 수 있다. 기술 트리에서 기술의 발전 방향을 보면 interface 기술을 나타내는 S6은 비교적 최근의 기술임을 확인할 수 있었다.



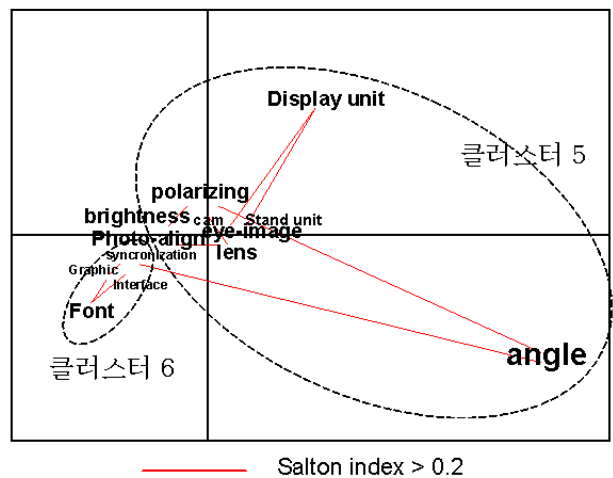
(a) S 기업



(b) L 기업

[Figure 3] 두 기업의 전체적 지도

세부적 지도에서는 각 클러스터에 속한 키워드와 그들간의 연관성을 확인할 수 있다. Figure 4는 두 개의 클러스터를 동시에 표현한 것으로 interface 기술과 display 기술의 세부적 키워드 지도이다.



[Figure 4] S기업의 세부적 지도(클러스터 5,6)

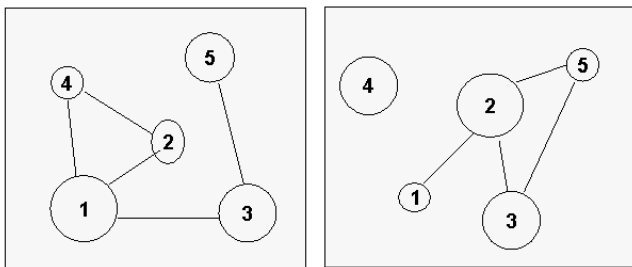
두 클러스터는 인접하여 있으나 클러스터 간 키워드 사이에는 큰 상관성을 보이지는 않았다. 키워드의 크기는 출현 빈도를 나타내며 salton index의 크기에 따른 선의 굵기는 두 키워드 간의 연관 정도를 보여줌으로써 구체적인 기술 구조와 대략적인

핵심기술을 한눈에 파악할 수 있다.

4. 전략적 활용 방안

전체적/세부적 지도의 표현을 통해 자/타사의 기술 현황 확인 및 비교가 가능했다. 즉 보유한 기술의 발전 혹은 퇴보 정도와 기술간의 관련성을 확인할 수 있었다. 클러스터의 크기가 작은 기술은 Figure 1에서 제시한 기술 트리의 발전 양상에 비추어 신생 기술 혹은 퇴보 기술로 확인할 수 있을 것이다. 또한 포함 범위가 큰 클러스터는 내부적으로 그룹핑을 실시하여 기술의 분류가 가능하며, 시간에 따른 동적 분석은 특허의 출원연도를 달리하여 키워드를 추출하고 매핑하여 비교가 가능하다.

이와 같은 방안들은 세부적 기술 내용에 대한 확인이 불가능한 특허 포트폴리오가 가진 한계를 보충하거나 대안으로 쓰여서 중간 관리자의 포트폴리오 작성에 도움이 될 수 있을 것이다.



[Figure 5] Cross-licensing을 위한 기술지도 비교

기본적인 기술 확인 이외에 적극적인 전략적 활용방안은 특허기술의 licensing 혹은 cross-licensing에 응용하는 것이다 (Brockhoff et al., 1999). Licensing의 경우 자사가 보유하지 않은 타사의 키워드를 검색하여 그 키워드가 속한 기술군을 찾아내어 역으로 키워드 보유 문서를 추적할 수 있다. Cross-licensing의 경우에는 여러 가지의 방안을 모색할 수 있다. Figure 5에서 기술 3은 양사 모두 역량을 모으고 있는 부분이므로 반도체와 같은 첨단 축적기술의 경우에 서로 교환이 가능하다. 그리고 기술 1, 2은 쌍방을 보충해줄 수 있는 경우이며, 기술 5, 6은 주요기술 3과 밀접하면서 서로 가지고 있지 못한 경우로 기술의 쌍방 도입이 가능하다. 특허 교섭을 위해서는 관심 키워드를 가진 특허를 역추적한다. 그러나 cross-licensing은 두 기업의 클러스터 비교시에 클러스터의 일치를 확인하기 위한 더욱 세심한 접근이 필요하다는 한계를 가지고 있다.

5. 결론

본 연구에서는 특허 포트폴리오의 한계를 극복하기 위하여 co-word 분석법을 활용한 기업의 기술지도를 제안하였다. 전체적/세부적 지도의 작성과 지도 분석을 통한 기업의 활용 방안을 제시하였으며, 경쟁 기업간의 기술 비교를 통해 cross-licensing이 가능한 기술 영역을 제안하였다. 그러나 클러스터에 기술을 매칭할 때의 자의성, 두 기업의 기술 비교시 어느 정도 유사해야 같은 기술로 취급할 수 있는지에 대해서는 논란의 여지가 남아있다. 또한 데이터 전처리에서 키워드 삭제 및 정리 과정의 임의성을 줄일 필요가 있다. 따라서

후후 연구 과제는 키워드의 범위를 축소하는 것으로, 특정 연구 분야의 기술 용어 사전(technology vocabulary dictionary)을 작성하여 이와 같은 한계를 극복할 수 있을 것이다.

참고 문헌

- Börner, K., Chen, C., and Boyack, K. W. (2003), Visualizing Knowledge Domains, *Annual Review of Information Science & Technology*, 37.
- Brockhoff, K.K., Ernst H., and Hundhausen E. (1999), Gains and Pains from Licensing-patent-portfolios as Strategic Weapons in the Cardiac Rhythm Management Industry, *Technovation*, 19 (10), 605-614.
- Ding, Y., Chowdhury G. G., and Foo S. (2001), Bibliometric Cartography of Information Retrieval Research by Using Co-word Analysis, *Information Processing and Management*, 37 (6), 817-842.
- Engelsman, E. C. and van Raan, A. F. J. (1994), A Patent-based Cartography of Technology, *Research Policy*, 23 (1), 1-26.
- Ernst H. (1998), Patent Portfolios for Strategic R&D Planning, *Journal of Engineering and technology management*, 15 (4), 279-308.
- Ernst, H. (2003), Patent Information for Strategic technology management, *World Patent Information*, 25 (3), 233-242.
- Peters, H. P. F. and van Raan, A. F. J. (1993), Co-word-based Science Maps Engineering. Part 1 : Representations by Direct Multidimensional Scaling, *Research Policy*, 22 (1), 23-45.
- Peters, H. P. F. and van Raan, A. F. J. (1993), Co-word-based Science Maps Engineering. Part II : Representations by Combined Clustering and Multidimensional Scaling, *Research Policy*, 22 (1), 47-71.
- Polanco, X., Boudourides, M. A., and Roche, I. (2001), Clustering and Mapping Web Sites: For Displaying Implicit Associations and Visualizing Networks, *Institute for systems, informatics and safety of the European Commission*, 1 (2). 1-25.
- Shih-Chang Hung (2002), The Co-evolutions of Technologies and Institutions: a Comparison of Taiwanese Hard Disk Drive and Liquid Crystal Display Industries, *R&D Management*, 32 (2), 179-190.