

COMPAS 시스템을 활용한 유사특허탐색과 분쟁특허 유사도 측정

강종석* · 홍성화** · 이혁재***

I. 서론

1. 연구 배경

기업의 경쟁기술 모니터링은 기업 활동(business)에 상당한 영향을 줄 수 있는 과학기술 및 사회경제적인 유발요인을 사전에 감지하는 것으로서 성공적 기업 활동에 잠재된 위협요소를 선제적으로 파악하는 개념으로 정의될 수 있다. 또한 기업의 모니터링 활동은 현재의 경쟁력을 유지하기 위한 활동으로 해당 기술의 특이성과 동향을 지속적으로 추적하고 유의미성을 판별하는 일체의 기술지식활동(technical intelligence activities)으로 해석되기도 한다. 따라서 모니터링 대상을 과학과 기술을 아우르는 영역으로 설정하면 과학기술 모니터링(S&T monitoring)으로 정의 할 수 있고, 반면 기술영역으로 한정한다면 기술 모니터링(Technology monitoring)으로 단편화 할 수 있다(DOE, Monitoring science and technology for competitive advantage, 1993).

최근 들어 해외 다국적 기업이 막대한 자금력을 바탕으로 무차별적인 특허소송을 제기하면서 국내 기업에 대한 특허공세가 날로 심화되고 있고, 이는 특허분쟁 대응 체제의 부족으로 해외 로열티 지급, 지식재산권 침해 소송, 특허괴물(patent troll)을 통한 기술 이탈, 그리고 중국 등 경쟁국에 의한 지식재산권 침해사태가 급증하는 추세에 있다. 외국 기업과의 특허분쟁은 그 소송에서 패한 경우 손해배상의 지불 등은 열외로 하더라도, 기업의 브랜드 이미지 실추에 따른 세계시장의 주도권 상실 우려 고조, 해외 판로개척과 신규 거래처 확보 악영향 등 정상적인 기업 활동 자체를 위축시키게 된다.

- 삼성전자 순이익의 11.8%인 1조 2,813억 원(2004년)과 와 LG전자는 3,603억 원(2003년)을 각각 로열티로 지급(조광현, 지식재산, 2008)
- 삼성전자와 LG전자는 휴대전화 관련 특허침해로 인해 미국의 특허괴물인 인터디지털에 각각 1억 3400만 달러와 2억8500만 달러를 지불(동아일보, 2009)
- 수출금액이 100만 달러 이상 1,000만 달러 미만인 국내 기업이 침해를 가장 많이 당한 국가는 중국(64.7%)(특허청, 2008)

한편, 1990년대 이후부터 세계는 지식재산이 국가발전의 중요 자산이 되는 지식기반 경제로 전환됨에 따라 국내 경제 역시 창의·혁신 역량 강화를 통한 국가경쟁력 제고의 필요성이 강하게 대두되어 왔다. 이를 위해 정부는 「국가경쟁력강화위원회」를 통해 창의자본 활성화 등의 11대 중점추진과제를 포함하는 「지식재산강국실현전략」을 제시하였고(특허청, 2009/7/29), 그 중 하나로 「국가차원의 지재산 분쟁 대응 체제 구축」을 제시하였다. 이를 위해 ‘경쟁사(국) 특허분석 등의 주기적 모니터링을 통한 사전 분쟁 예보시스템의 개발’을 현재 추진 중에 있다.

* 강종석, 한국과학기술정보연구원, 기술기획연구팀, 책임연구원, 02-3299-6048, kangjs@kisti.re.kr
** 홍성화, 한국과학기술정보연구원, 기술기획연구팀, 책임연구원, 02-3299-6048, shong@kisti.re.kr
*** 이혁재, 한국과학기술정보연구원, 기술기획연구팀, 책임연구원, 02-3299-6048, hlee@kisti.re.kr(교신저자)

R&D 관련 특허분석 행위로는 선행기술 조사 등의 출원 사전과정과 기술 모니터링 및 평가 등의 사후 과정으로 크게 양분될 수 있다. 특히 기술 모니터링 과정은 지속성 및 주기성, 기술 구현 전문성 등의 이유로 다수의 고급인력을 필요로 하는 노동집약적 작업이라는 특징이 있어, 해외의 경우 IBM(2500명), Matsushita(800명), Toshiba(340명), NEC(300명), Sony(300명) 등을 비롯한 다수의 기업이 막강한 특허 전문 인력을 확보하고 기술 모니터링 활동을 수행하고 있다. 반면, 국내는 삼성(250명) 및 LG(IP센터 소속 200명) 등의 일부 대기업을 제외하면 크게 부족한 실정으로 파악되고 있다(대외경제정책연구원, 2005). 특히, 코스닥 등록기업(230개사 대상)의 80.6%가 특허관리 전담부서를 두고 있지 않으며, 전담부서가 있는 경우라도 1-4명의 규모를 가지는 경우가 대부분(95.5%)을 차지하는 것으로 보고되고 있다(지식재산권연구센터, 2004).

상기 이유로 인해 모니터링의 대상이 되는 특허는 전 세계로 확대되는 동시에 매년 기하급수적으로 증가하는 추세이므로 전통적인 방법을 통해 이들 특허 모두를 추적 모니터링 하는 것이 현실적으로 불가능하다. 선진국의 경우는 특허정보의 계량분석적 방법에 의해 특허분쟁 모니터링을 지원할 수 있는 체계가 가동 중이며, 미국의 OceanTomo, iPiQ, Manning & Napier Information Services¹⁾, RightData Corp²⁾, Iowa State University³⁾, Office of Naval Research, 일본의 IPB, 영국의 Horizon Scanning Center 및 캐나다의 CISTI 등이 대표적인 사례이다. 반면, 국내의 경우는 특허정보에 특화된 분석도구를 제공하는 사례로 민간기업인 웹스의 ThinKlear와 특허청의 PIAS가 있으나 분쟁 예견 사례의 모니터링 기능이 부족하거나 사용목적에 최적화되지 못하였고, 민간 컨설팅 기관의 경우는 활동이 미미할 뿐 아니라 일회성인 경우가 대부분을 차지하고 있다.

따라서 한국이 강점을 보이는 IT기술을 적극 활용하여 효율적이고 효과적으로 특허분쟁 가능성이 높은 특허들을 사전 탐지할 수 있는 국가차원의 체제를 갖추는 것이 지식재산강국의 실현을 위해 필수적 사안으로 인식되고 있다. 이에 본 연구는 출원 특허를 대상으로 특허상호간의 권리화된 기술에 대한 유사도 계측의 수학적 모델을 제시하고 전사모사 하여, 특허분쟁예보 체계의 시스템적 구현 가능성을 제시한 것이다.

II. 본 문

1. COMPAS[®] 유사특허탐색 체계

1) 중소기업 니즈 도출

본 연구는 정보 분석 기반의 모니터링 시스템 구축을 위한 초기모델을 검토하고, 사전에 파악된 수요자 니즈(needs)가 반영된 모니터링 시스템 구성을 위한 가상모델 제안하며, 그리고 실제 특허정보를 활용하여 제안된 모니터링 시스템을 구현하는 것이다. 우선, 수요자 니즈를 파악하기 위하여 이미 연구 발표된 학술논문조사, OECD 국가들을 대상으로 유사 정부지원 프로그램을 조사하였다. 미국에서 수행

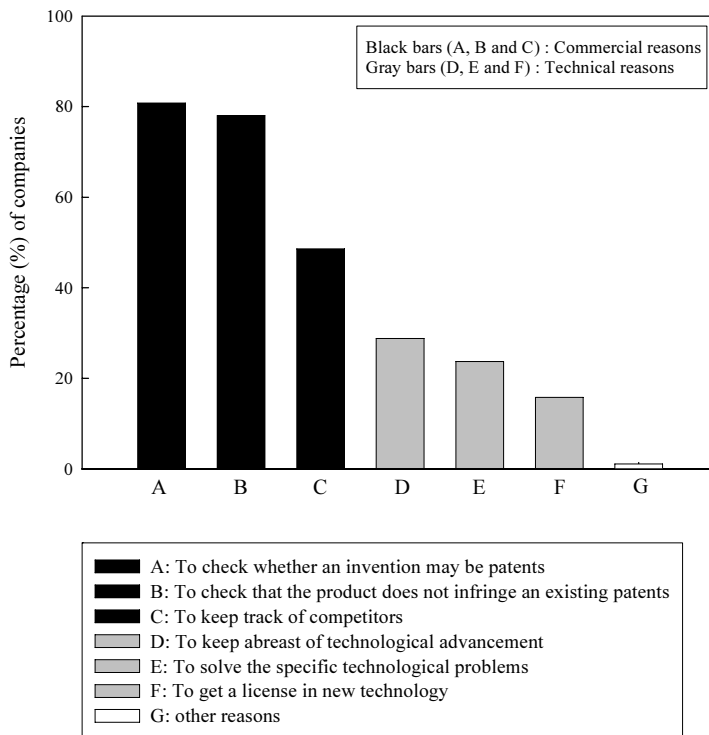
1) 특허분쟁 모니터링 시스템인 "constant watch" 운영 (<http://www.mins.net>)

2) 특허침해 및 분쟁, 경쟁사 감지 시스템 운영 (<http://www.rightdata.com>)

3) 특허분쟁 조정모델연구, Patent Litigation with Endogenous Disputes, The American Economic Review, Vol.96, No.2 (May 2006), pp.77-81

된 초기 연구결과(W.B. Ashton, A. Johnson, G. Stacey, PNL-SA-23128, 1993)를 보면 모니터링 시스템 구성 중, 모니터링 프로세스의 시스템적 구축을 위한 체제를 강조하고, ①서비스대상 정의: who & what, ②시스템 구상 및 설계, ③소스 선정 및 확보, ④결과분석, ⑤모니터링 프로덕트의 평가, ⑥지원 수행 및 리뷰의 프로세스로 진행된다.

중소기업을 대상으로 한 특허정보 및 특허정보 분석 시스템의 활용성에 대한 수요조사는 WIPO (world intellectual property organization)와 영국을 중심으로 한 유럽연합에서 활발한 연구가 진행되었는데, 그 내용을 정리하면 다음 <그림 1>과 같다.



<그림 1> 특허정보의 활용 목적별 분포⁴⁾

<그림 1>에서 알 수 있듯이, 개발된 신기술이 특허를 통한 권리화 확보 가능성 및 경쟁기업 기술에 대한 특허 침해성 확인, 그리고 인지된 경쟁사의 신기술 파악이 가장 높은 우선순위를 보임에 따라 해당기술의 특허 권리화에 대한 방어 및 공격적 목적, 그리고 경쟁자(기술) 탐색으로 요약된다. 따라서 상기 중소기업의 니즈를 반영하여 경쟁사 유사특허탐색 모형을 제시하고 COMPAS[®] 시스템에 반영하였다.

2) 유사특허탐색 모형

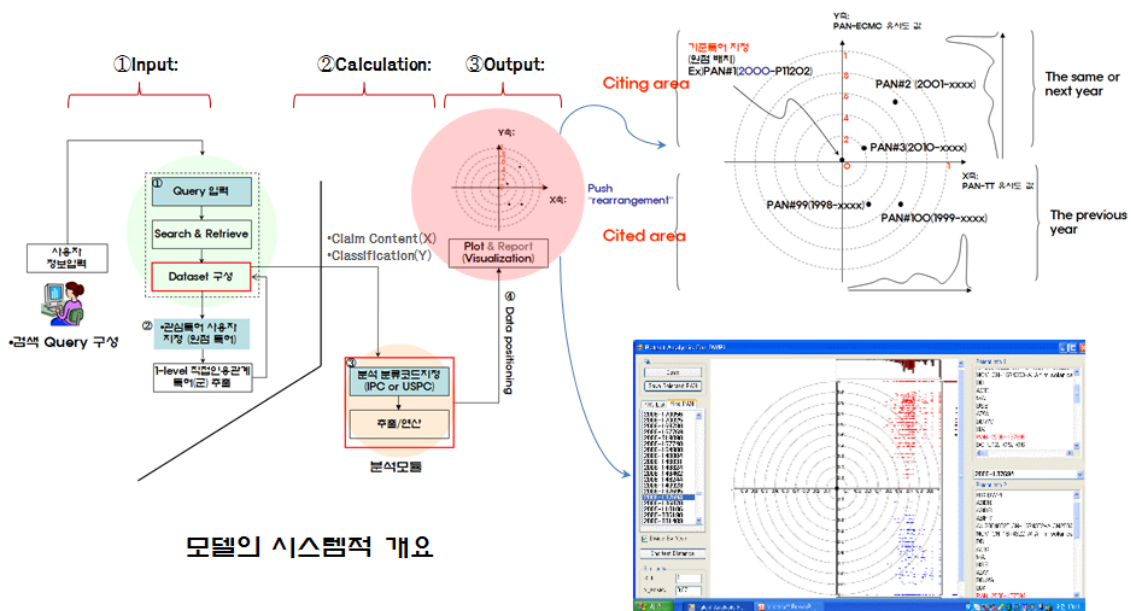
COMPAS 시스템에서 구현되는 경쟁사 유사특허탐색 모형의 프로세스는 6개(A-F)의 단계별 데이터처리가 동반된다. 각 부분모듈별 설명은 다음과 같다.

4) 상기 결과는 영국의 177개 중소기업을 대상으로 설문조사를 바탕으로 조사된 것으로, 원문 (M. Hall, C. Oppenheim and M. Sheen, 1999, Barriers to the use of patent information in UK small and medium-sized enterprises. Part 1: Questionnaire survey, Journal of Information Science, Vol. 25, No. 5, pp 335-350)에 발표된 내용을 도표화 한 것임.

- A)특허 데이터 셋 구축 : 기본적으로 가용한 특허 데이터베이스 (각국의 특허청에서 구축된 DB와 상용 특허DB를 포괄함) 에서 검색자의 질문식에 부합되는 특허데이터를 확인·다운로드하여 초기 모집단 데이터 셋의 구축 과정이며, 사용자 관심 특허의 지정에 따라 1차 인용관계 특허(군)은 자동적으로 모집단 데이터 셋에 반영됨.
- B)필드 추출 및 표준화 : 분석하고자 하는 특허명세서 중의 특정 항목 또는 필드 항목의 사용자 지정 및 추출과정.
- C)메타데이터 구성 : 상기 B)단계 수행이후에 가공된 특허 데이터 셋 구축
- D)분석 및 계측 필드 지정 : 특허청구항 내용을 반영한 가공필드에 대한 추출과 기술분류코드(IPC, MC, 미국특허분류코드, 기타 특허분류코드 등)의 추출.
- E)특허 포괄성 및 유사도 계측 : D) 모듈에서 추출된 필드에서 개별 특허에 대한 벡터화를 실시하고, 이것으로부터 특허에서 서술되거나 권리화하고자 하는 기술적 관점에서 특허상호간의 동시발생행렬을 구성 및 유사도 계측
- F)가시화 : 특허상 간의 유사도 계산값에 대한 거리환산 및 가시화 구현

결론적으로, X축은 특허 청구항에서 권리화 하고자하는 기술에 대한 유사도 및 포괄성에 대한 값을 계측하는 것이며, Y축은 특허 분류코드에서 분류된 기술에 대한 유사도 및 포괄성에 대한 값을 계측한 것이다(값의 범위는 $0 \leq X$ 또는 $Y \leq 1$). 이러한 결과값의 해석은 (0, 0)에 근접할수록 유사성 및 포괄성이 낮은 것을 나타내며, 반대로 (1, 1)에 가까울수록 완전히 유사하거나 어느 한 특허의 권리화된 기술내용이 기타의 특허에 포함됨을 의미한다.

M1: 유사특허탐색모듈



<그림 2> COMPAS 「유사특허탐색」 시스템 구성

본 연구에서는 구조화된 특허필드 내의 구성성분 간 유사도 계측과 특허 청구항 부분의 텍스트 기반 유사도 계측을 동시에 실시하고 이를 동일 도메인에 위치하도록 하여 2축 모드로 가시화하는 「2차원 혼성 유사도 계측모델」을 제안하였고 그 수학적 표현은 아래와 같다.

- ◆ x축의 수학적 정의: 특허 청구항의 텍스트 중첩기반 유사도 계측

$$N_{ij} = P_{kw} \cdot P_{kw}^T$$

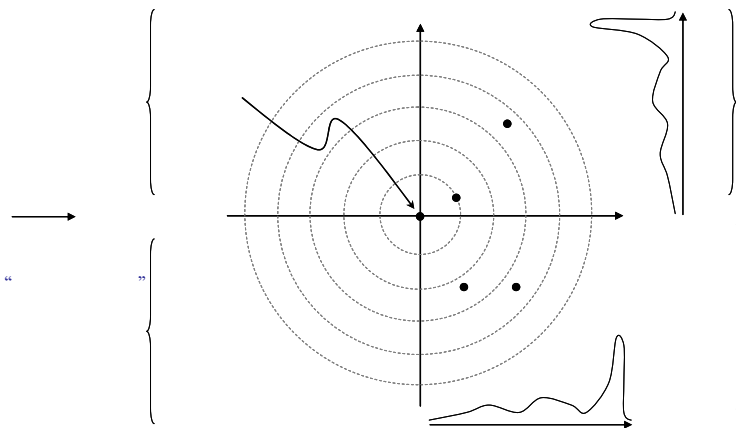
- ◆ y축의 수학적 정의: 특허기술분류 코드(CC, IPC, MC, USPC 등) 중첩기반 유사도 계측

$$M_{ij} = P_{cc} \cdot P_{cc}^T$$

여기에서 특허 i와 j에 대한 유사도 매트릭스를 N, M으로 정의하며, Pkw와 Pcc는 각 특허에서 추출된 키워드 및 분류체계 매트릭스를 의미한다. 최종적으로 계측된 2개의 값(n_{ij} , m_{ij})을 원점기준의 중첩도 분포(레이더 망 구현)로 가시화하였다.

2. 분쟁특허 유사도 계측

1) 가시화 해석



<그림 3>COMPAS[®] 유사특허탐색 분석결과 가시화

<그림 3>은 해당 COMPAS[®] 모듈의 최종 분석결과를 가시화한 것이다. COMPAS[®] 유사특허탐색 모듈은 사용자가 유사도 연산방법 선택, 분석필드 및 분류레벨 지정, 사용자 관심특허 지정으로 모든 분석수행은 자동으로 실행된다. 결과 화면은 레이더형 유사도 분포로 산출되며 사용자 지정 관심특허는 자동적으로 xy축의 원점에 위치된다. 따라서 사용자는 분석결과를 직관적으로 파악 할 수 있도록 하였다. 다시 말해, 원점(0,0)인 사용자 관심특허를 기준으로 가까울수록 유사성이 높은 특허로 파악할 수 있고 반면, 멀리 위치할수록 유사정도가 낮은 특허(군)로 인지할 수 있도록 구현된다.

해당 모듈의 분석 수행의 과정은 사용자에게 공개되지 않으며, 초기 분석 결과의 초기 내용으로 확인 가능하다. 전체적인 분석내용(각 좌표축을 기준으로 빈도에 따른 분포결과가 사용자 지정 특허에 따라 실시간 반영되도록 설계 됨)을 확인하고, 분절 기능키를 활용하여 선택적으로 4가지 유형(인용관계에 있지 않는 선행특허(군), 인용관계에 있지 않는 비선행특허(군), 인용관계에 있는 선행특허(군), 인용관계에 있는 비선행특허(군))으로 분산 배치된 최종 결과를 얻을 수 있다.(<그림 3> 참조)

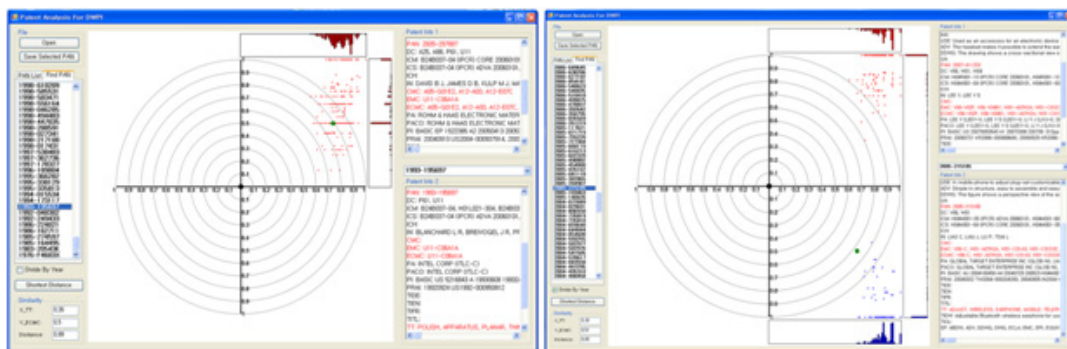
2) 분쟁특허 유사도 측정

COMPAS[®] 유사특허탐색 모듈을 이용하여 실제 분쟁특허쌍의 유사도 계측 및 제안 모형의 유효성을 검토하였다. <표 1>은 본 연구에서 고려된 특허분쟁 사건으로 원고 및 피고측 특허, 원고 및 피고측 인용특허로 구성되며, 분야별 특성을 고려하여 화학분야 3건, 원자력 1건, 기계분야 2건, 전기전자분야 4건으로 국내 대법원 민사본안사건을 대상으로 하였다.

<표 1>분쟁특허 사례활용

No	Case No.	원 고	피 고	유 형	특 성	분야
1	2005허-10459	Abbott Inc.	Hanmi Inc.	균등침해	-	화학
2	2007나-23335	Tonen Chem.	(주)SK	균등침해	피고인용특허	화학
3	2008당-774	SKC Inc.	Toray	균등침해	원고인용특허, 1&12&13&17항	화학
4	EPA	AREVA Inc.	KAERI	균등침해		원자력
5	2007-3356	SKC Inc.	(주)Rom&H	간접침해	원고인용특허	기계
6	2008-3889	Wise&blue Inc.	Dream Tech Inc.	문언침해	원고인용특허	기계
7	2001-32187	Konami Inc.	Amuse Inc.	문언침해	피고인용특허	전기전자
8	2004-2596	FormFact Inc.	FiCcom Inc.	문언침해	피고인용특허	전기전자
9	2006-1831	Canon	Pacano Inc	문언침해	피고인용특허, 25&26항	전기전자
10	2009-46712	EyeCT Inc.	(주)KB Tech	균등침해	피고인용특허	전기전자

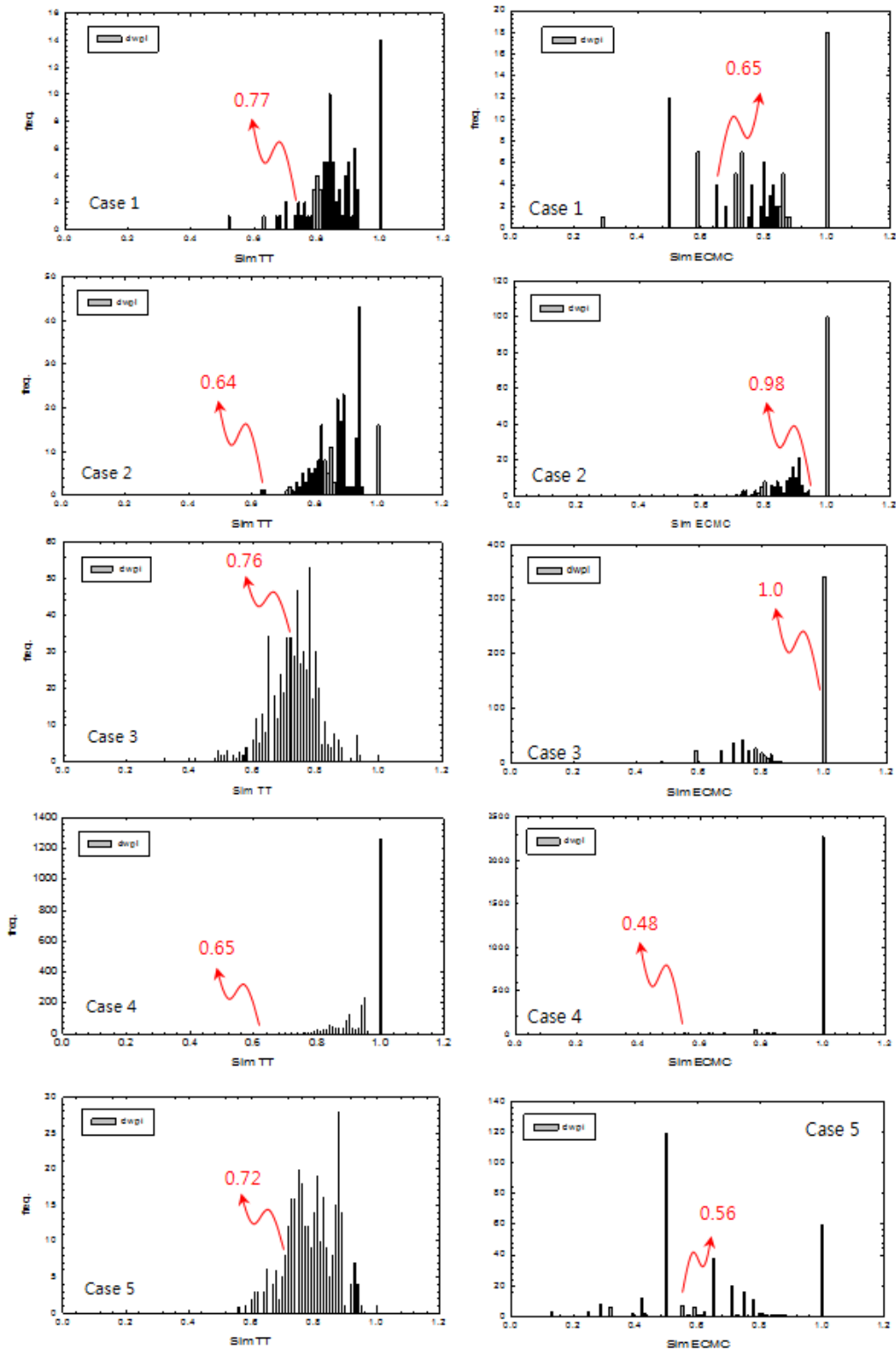
<그림 4>는 실제 COMPAS[®]를 이용한 분석결과를 제시한 것으로, 원점(0,0)은 원고측 특허를 사용자 관심특허로 지정하고 이에 대응되는 피고측 특허의 위치를 추적한 것이다. x축과 y축에 대응되는 분석 모집단 특허와 피고특허에 대한 유사도에 따른 거리환산 값으로 각축의 분포형태는 모집단의 특성에 기인한 것으로 판단되나, 특허분쟁 대상인 피고특허의 위치는 고유한 것으로 해석될 수 있다. <그림 5>와 <그림 6>은 x축(청구항 기술키워드 유사도)과 y축(특허분류코드 유사도)의 분쟁대상 특허군의 분포형태와 피고특허의 위치를 계산한 것이다.



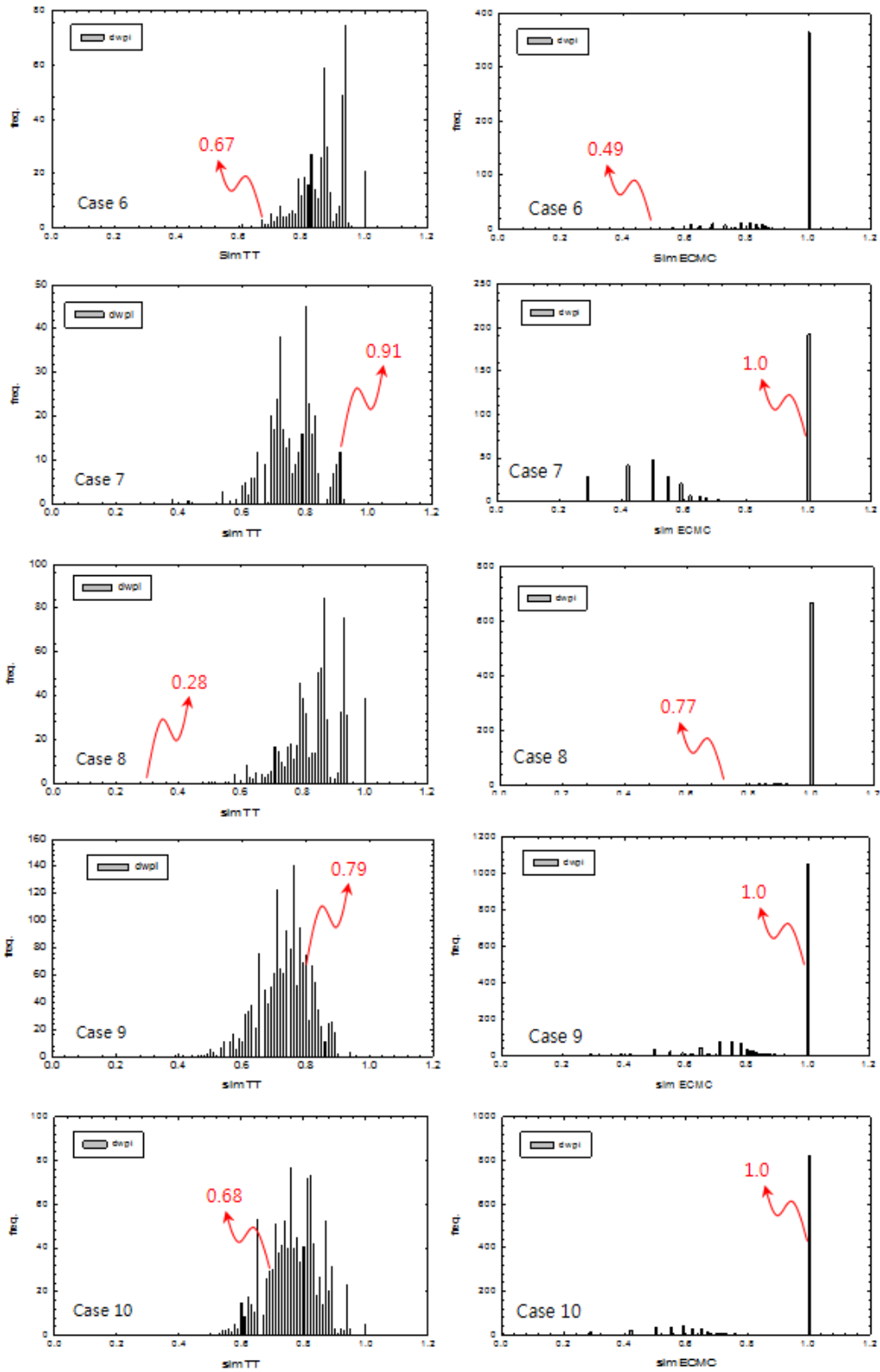
(가)사건번호 2007후 3356 (문헌침해)

(나)사건번호 2008허 3889 (문헌침해)

<그림 4>COMPAS[®] 분석결과 화면



<그림 5>COMPAS[®] 유사특허탐색 모듈의 x-y축 분포형태 및 피고특허 위치 (case 1-5)



<그림 6>COMPAS[®] 유사특허탐색 모듈의 x-y축 분포형태 및 피고특허 위치 (case 6-10)

III. 결론

실제 분쟁특허 쌍에 대한 유사도 측정을 COMPAS[®] 시스템을 활용하여 수행하였다. 각 특허의 청구항에서 추출된 기술키워드를 대상으로 한 거리(x축 위치)와 특허기술분류코드를 이용한 거리(y축)의 분포를 확인해 보면, x축은 0.28-0.91의 범위를 가지며, 약 0.68의 평균값을 보인다. 반면, y축은 0.56-1.0의 범위로 평균값 0.79로 산출되었다. 이것은 기술분류코드의 분류과정에서의 오류를 인정하더라도 코드체계가 가지는 불연속성이 반영된 결과로 해석될 수 있다. 또한 산출된 거리의 평균값에 대한 분쟁특허 간의 특허유사성을 해석해 보면, 균등 및 문헌침해의 경우, 분쟁대상 특허쌍의 청구항 구성 기술키워드와 특허기술분류가 동시에 약 25%정도의 일치성을 보여야 함을 의미하는 것이며, 또한 청구항 기술키워드가 약 32-35% 유사성을 보일 경우 기술분류코드의 유사성은 배재되어도 무방하다. 이러한 해석은 일반화된 개념으로 확대시키기는 어려우며, 본 연구의 사례에 해당함을 밝혀 둔다.

<표 2>분쟁특허 유사도 계측

No	Case No.	원 고	피 고	x축	y축
1	2005허-10459	Abbott Inc.	Hanmi Inc.	0.77	0.65
2	2007나-23335	Tonen Chem.	(주)SK	0.64	0.98
3	2008당-774	SKC Inc.	Toray	0.76	1.0
4	EPA	AREVA Inc.	KAERI	0.65	0.48
5	2007-3356	SKC Inc.	(주)Rom&H	0.72	0.56
6	2008-3889	Wise&blue Inc.	Dream Tech Inc.	0.67	0.49
7	2001-32187	Konami Inc.	Amuse Inc.	0.91	1.0
8	2004-2596	FormFact Inc.	FiCcom Inc.	0.28	0.77
9	2006-1831	Canon	Pacano Inc	0.79	1.0
10	2009-46712	EyeCT Inc.	(주)KB Tech	0.68	1.0

상기 연구결과의 활용은 1)연구개발과제 관리업무에서 연구과제의 중복성 검토, 2)특허 및 논문을 대상으로 하는 선행조사, 3)표절(plagiarism) 가능성 탐지, 4)분쟁가능 특허 탐색 및 분쟁예보에 적용될 수 있을 것으로 기대한다.

참고문헌

- DOE (Department of Energy, USA) (1993), "Monitoring science and technology for competitive advantage".
- 조광현 (2008), "기업의 지식재산관리시스템 설계 및 진단모형 개발(1)" 「지식재산」, 21(103)
- 특허청 (2008), 「해외지재산권 피침해 실태 조사 보고서」
- 지식재산권연구센터 (2004), 「기업특허관리 전담부서의 발전전략」
- Hall M., Oppenheim C., and Sheen M., (1999), "Barriers to the use of patent information in UK small and medium-sized enterprises. Part 1: Questionnaire survey", Journal of Information Science, 25(5), 335-350.
- James B. and Michael J. Meurer, (2006), "Patent Litigation with Endogenous Disputes", The American Economic Review, 96(2), 77-81.

- Kang J. and Lee H. (2012), "Systematic Monitoring of Competitors' Patents Using 2-dimensional Hybrid Similarity Method", *Information*, 15(7), 2813-2822.
- Kang J., Lee H. and Moon Y., (2010), "Systematic Approach for Monitoring Competitor's Technological Challenges based upon Patent Analysis", *Information*, 13(2), 339-352.
- Christopher A. Cotropia. (2005), "Patent Claim Interpretation Methodologies and Their Claim Scope Paradigms", *William & Mary Law Review*, 47, 05-05.
- Michael J. Meurer and Craig A. Nard. (2005), "Invention, Refinement and Patent Claim Scope: A New Perspective on the Doctrine of Equivalents", *Georgetown Law Journal*, 93, 04-05.