

---

# 특허 발명의 명칭에 쓰인 단어를 이용한 기술동향 분석 연구

## Analysis of Technology Trends from Words in Patent Titles

---

김태중, 이명선, 최호남  
한국과학기술정보연구원

Tae-Jung Kim(taejung@kisti.re.kr), Myung-Sun Lee(mslee@kisti.re.kr),  
Ho-Nam Choi(hnchoi@kisti.re.kr)

---

### 요약

특허는 중요한 과학기술 연구 성과물을 담고 있다. 과학기술용어의 사용 빈도를 활용해 기술동향을 분석하는 연구가 있으나 용어는 종종 분야에 따라 다른 의미로 사용된다. 이 논문에서는 2000년부터 2008년까지 매 2년 주기의 미국, 일본, 한국, 유럽(EPO), 국제(PCT) 특허의 제목(발명의명칭)으로부터 단어를 추출하여 WIPO가 제시한 전기공학, 기기, 화학, 기계공학, 기타 등의 5가지 분류로 정리하여 용어의 출현 빈도 변화를 계산하였다. 이 값이 상위에 있는 단어들을 분석하여 기술동향, 특허와 기술개발과의 관련성을 분석하였다.

■ 중심어 : | 특허 발명의 명칭 | 단어 출현빈도 | 기술 동향 분석 |

### Abstract

Patent contains meaningful technical achievement. There are many cases explaining technology trends from the analysis of frequency of term. Term sometimes has different meaning on fields. In this paper, words from patent titles of US, Japan, Korea PCT and EPO are collected by the 5 categories of WIPO. Frequency changes rate of each word were calculated and high ranked words of 5 categories were analyzed to find relationship between patent and technology development as well as technology trends.

■ keyword : | Patent Title | Word Frequency | Technology Trends Analysis |

---

## I. Introduction

Patent is used as a tool analyzing and forecasting the direction of technology development, and patent information is good indicators of R&D output[1]. A patent document contains dozens of items for analysis; some are structured, meaning they are uniform in semantics and in format across patents

such as patent number, filing date, or assignee; some are unstructured, meaning they are free texts of various length and content, such as claims, abstracts, or description of the invention[2]. Patent trend analysis is a management forecasting tool that can be useful in R&D planning and new product development, and patent analysis can indicate the growth pattern of a technology and the technological

---

\* 본 논문은 한국콘텐츠학회 ICC2009 국제학술대회 우수논문입니다.

접수번호 : #100324-002

접수일자 : 2010년 03월 24일

심사완료일 : 2010년 04월 13일

교신저자 : 이명선, e-mail : mslee@kisti.re.kr

shift that are occurring[3]. Analysis of patent data has considered to be an important method of assessing various aspects of technological change. Patenting data provide information on both the technology levels in particular sector[4]. Co-word analysis applied to patents through WPIL normalized title words appears to give a useful picture of a given field, and patent title can be used find the process of development of product and manufacturing method, and select new research theme[5][6]. The author proposes a method to find technology changes by the analysis of words changes in patent documents. Technical terms sometimes are used in different meaning according to fields. So, words from the title are grouped into 5 categories, which is prepared to classify patent documents into industrial aspects by WIPO(World Intellectual Property Organization), electrical engineering(E), instruments(I), chemistry(C), mechanical engineering(M) and other fields(O), such as [table 1][7][8]. Every two year, from year 2000 to 2008, the change rate of frequency and variation in usage of each term was calculated and compared to find relationship between patent and technology changes.

## II. Data collection and preparation

Words from titles of 5 patent documents, 4 applications, Japan, Korea, EPO and PCT, and US registration from year 2000 to 2008, collected on the basis application date from KISTI's NDSL patent service[9]. Every title was classified by the 5 categories according to the IPC 4digit code assigned to the document. Words of a title were separated by space, comma, hyphen and slash and every characters were changed into small letter. [Table 1] and [Table 2] are the number of collected titles by categories,

year and patent offices. Finally, 30,772,685 words were collected except of some meaning-less words like preposition, conjunction and pronoun.

Table 1. No. of titles by 5 categories

code	No. of titles	remarks
C	1,594,542	Organic fine chemistry, biotechnology, pharmaceuticals, polymers, food chemistry, materials, chemical engineering, environmental technology
E	2,241,418	Electrical machinery, A-V technology, energy, telecommunications, computer, semiconductor
I	1,082,802	Optics, measurement, control, medical technology
M	1,579,617	Handling, machine tools, engine, turbines, machines, transport
O	540,771	Furniture, games, other consumer goods, civil engineering

Table 2. No. of titles by year and patent office(country)

Year	USA	EPO	Japan	Korea	PCT	total
2000	191,165	149,895	654,233	85,592	130,824	1,213,709
2002	205,493	156,863	752,663	88,881	137,777	1,343,679
2004	200,877	223,932	701,902	87,300	215,599	1,431,614
2006	217,525	222,285	668,665	166,571	277,707	1,554,759
2008	204,981	240,758	583,420	112,155	361,087	1,504,409
total	1,020,041	993,733	3,360,883	541,499	1,122,994	7,039,150

## III. Analysis and results

### 1. By occurrence

To find any special feature in changes of word occurrence, words are grouped by year, field and patent office, and sorted[Table 3-5]. The value of yearly fields as a sample machinery and electrical engineering are calculated and compared[Table 6]. Words commonly used in patent document such as 'method', 'device', 'manufacturing', 'program', 'use', 'system', apparatus', 'image' and that sort of words are excluded.

Table 3. Top 10 occurrences classified by year

	2000	2002	2004	2006	2008
1	semiconductor	information	display	display	display
2	circuit	Semiconductor	semiconductor	semiconductor	semiconductor
3	display	display	information	vehicle	vehicle
4	data	data	data	information	information
5	vehicle	circuit	vehicle	data	data
6	film	vehicle	circuit	communication	communication
7	information	film	film	film	film
8	resin	communication	communication	circuit	circuit
9	power	resin	power	power	power
10	sheet	power	resin	cell	cell

Table 4. Top 10 occurrences classified by category

	C	E	I	M	O	total
1	film	semiconductor	display	vehicle	game	display
2	resin	data	sensor	engine	door	semiconductor
3	water	information	semiconductor	air	water	information
4	acid	circuit	film	valve	wall	data
5	metal	display	laser	heat	building	vehicle
6	sheet	network	circuit	ink	panel	circuit
7	gas	power	lens	gas	golf	film
8	polymer	memory	vehicle	fuel	vehicle	communication
9	cell	terminal	fiber	sheet	body	power
10	protein	computer	data	power	storage	resin

Table 5. Top 10 occurrences classified by patent office(country)

	USA	EPO	Japan	Korea	PCT
1	data	vehicle	display	display	data
2	semiconductor	data	information	semiconductor	communication
3	circuit	communication	semiconductor	water	display
4	display	information	film	circuit	vehicle
5	memory	display	vehicle	air	network
6	power	circuit	resin	panel	information
7	vehicle	power	circuit	mobile	cell
8	information	gas	sheet	vehicle	power
9	communication	motor	data	data	film
10	network	engine	communication	power	circuit

Table 6. Top 10 occurrences classified by yearly category(machinery and electrical engineering)

	2000	2002	2004	2006	2008	
M	1	vehicle	vehicle	vehicle	vehicle	vehicle
	2	engine	engine	engine	engine	engine
	3	air	air	air	air	air
	4	valve	ink	valve	valve	power
	5	ink	valve	ink	heat	heat
	6	heat	heat	fuel	fuel	valve
	7	gas	sheet	sheet	gas	gas
	8	fuel	gas	gas	power	fuel
	9	resin	fuel	heat	tool	bearing
	10	motor	resin	motor	sheet	sheet
E	1	semiconductor	information	semiconductor	semiconductor	semiconductor
	2	circuit	semiconductor	data	data	data
	3	data	data	information	display	communication
	4	display	circuit	circuit	information	display
	5	communication	communication	communication	communication	information
	6	power	display	display	circuit	circuit
	7	memory	power	network	network	network
	8	signal	network	memory	terminal	power
	9	network	computer	terminal	mobile	memory
	10	storage	terminal	computer	power	cell

2. By the comparisons of change ratio

To find year to year changes in word occurrence explicitly, the change ratio is proposed and calculated as follows:

$$CR(\text{change ratio}) = (A/Ta) - (B/Tb)$$

\*Here, 'A' and 'B' is occurrences of words 'a' and 'b', and 'Ta' and 'Tb' is total occurrences of group belonging word 'a' and 'b'. [Table 7] and [Table 8] are samples of the comparison of CR by year.

Table 7. Top 10 CR based on year 2000 and 2006 in total

	2000-2002	2000-2004	2000-2006	2000-2008	2006-2008
1	information	information	display	display	power
2	game	fuel	mobile	wireless	vehicle
3	server	game	information	cell	sensor

4	terminal	server	cell	communication	wireless
5	communication	display	wireless	information	content
6	computer	terminal	terminal	mobile	communication
7	network	wireless	communication	content	crystal
8	wireless	communication	network	network	solar
9	cell	mobile	content	vehicle	energy
10	mobile	vehicle	game	game	media

Table 8. Top 10 CR based on year 2000 and 2006 in field E(electrical engineering)

	2000-2002	2000-2004	2000-2006	2000-2008	2006-2008
1	information	information	mobile	wireless	communication
2	service	fuel	wireless	fuel	wireless
3	server	wireless	fuel	communication	content
4	terminal	server	display	content	power
5	wireless	content	terminal	mobile	media
6	fuel	mobile	content	display	sensor
7	internet	terminal	information	network	solar
8	content	service	communication	information	vehicle
9	network	communication	network	service	video
10	communication	network	service	terminal	rfid

#### IV. Conclusion

From the [table 3-5] author can find that the Japan, USA and Korea have relatively strength in electrical engineering and instruments, and these areas lead technology development over last 10 years. By the comparison of CR, changes in interesting technology can be identified explicitly like 'wireless', 'LCD(display)', 'fuel cell(cell or fuel)', 'content' and so on.

There are two limits of this study or directions for future research. One of them is that CR needs more evidence for the usefulness by much more experimental study. The other one is the limitation of word. This study is based on the word which is extracted from title of patent documents. So, in case

of counting the occurrence of words like 'liquid crystal display(LCD)' or 'internal combustion engine', it was needed careful attention. For the better and easy processing, it might be improved to the basis of terms or product.

#### 참고 문헌

- [1] "IP and business: Patent Information: Buried Treasure," WIPO Magazine, 2005(1).
- [2] Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, "Text mining techniques for patent analysis," Information Processing and Management, Vol.43, pp.1216-1247, 2007.
- [3] R. S. Campbell, "Patent trends as a technological forecasting tool," World Patent Information, Vol.5, No.3, pp.137-143, 1983.
- [4] B. P. Abraham and S. D. Moitra, "Innovation assessment through patent analysis," Technovation, Vol.21, pp.245-252, 2001.
- [5] J. P. Courtial, M. Callon, and A. Sigogneau, "The use of patent titles for identifying the topics of invention and forecasting trends," Scientometrics, Vol.26, No.2, pp.231-242, 1993.
- [6] S. Jung, "Importance of using patent information," WIPO-MOST Intermediate training course on practical intellectual property issues in business, organized by the World Intellectual Property Organization (WIPO), Geneva, 11.
- [7] Korea Institute of Patent Information, "Patent trends of Korea 2005(I)-Nation/Technology (한국의 특허동향 2005(I)-국가별/기술별)", 특허청, p.184, 2005.
- [8] U. Schmoch, "Concept of a Technology Classification for Country Comparisons," Final report to the World Intellectual Property

Organization(WIPO), Fraunhofer Institute for Systems and Innovation Research, German, 2008.

[9] <http://patent.ndsl.kr/index.do>

저 자 소 개

김 태 중(Tae-Jung Kim)

종신회원



- 1978년 2월 : 성균관대학교 금속공학과 (공학사)
- 1990년 2월 : 성균관대학교 경영대학원 정보처리학과 (경영학석사)
- 2008년 2월 : 경희대학교 정보통신대학원 멀티미디어시스템공학과 (공학박사)

- 1981~1995 : 한국과학기술정보연구원 선임연구원
  - 1996~1999 : 한국데이터베이스진흥센터 부장
  - 1999~현재 : 한국과학기술정보연구원 책임연구원
- <관심분야> : 과학데이터, 저작권보호 및 관리 기술, 정보검색

이 명 선(Myung-Sun Lee)

정회원



- 1983년 2월 : 아주대학교 전자공학과 졸업(공학사)
- 1996년 8월 : 한남대학교 컴퓨터공학과 졸업(공학석사)
- 2005년 2월 : 한남대학교 컴퓨터공학과 졸업(공학박사)

- 1983년 3월 ~ 현재 : 한국과학기술정보연구원 책임연구원
- <관심분야> : 정보보안, 정보처리, 정보통신, 정보시스템

최 호 남(Ho-Nam Choi)

정회원



- 1977년 2월 : 연세대학교 도서관학과 (문학사)
- 1993년 8월 : 충남대학교 대학원 문헌정보학과 (석사)
- 2009년 2월 : 충남대학교 대학원 문헌정보학과 (박사)

- 1981~2005 : 한국과학기술원 정보개발팀장
  - 2006~2009 : 한국과학기술정보연구원 지식기반실장
  - 2006~현재 : 한국과학기술정보연구원 책임연구원
  - 2005~현재 : 과학기술정보관리협회 회장
  - 2009~현재 : 한국전문도서관협의회(KSLA) 초대 회장
  - 2009~현재 : 한국도서관협회(KLA) 부회장
- <관심분야> : 계량정보학, 디지털 라이브러리