



구조화된 연관맵을 이용한 연구개발 전략 수립

A R&D strategies for development using structured association map

송원호* · 이준석** · 박상성***†

Wonho Song, Junseok Lee, and Sangsung Park †

*고려대학교 공학대학원 지식재산학과, **고려대학교 산업경영공학과

***고려대학교 기술경영전문대학원

* Department of Intellectual Property, Korea University

** Department of Industrial Management Engineering, Korea University

*** Graduate School of Management of Technology, Korea University

요 약

급변하는 글로벌 시장 환경에서 기술은 계속해서 급속히 발전하고 있다. 이러한 급변하고 있는 환경을 반영한 연구개발은 기업이 있어서 필수가 되었다. 즉, 기업의 경쟁력 향상을 위해서는 자사가 보유한 기술에 대한 체계적인 분석이 필요하다. 최근에는 객관적이며 정량화된 기술분류를 위하여 특허문서의 IPC 코드를 이용하여 기술분류를 수행하고 있다. 국제특허분류인 IPC 코드는 국제적으로 규격화된 기술분류 코드이기 때문에, 이를 활용하면 객관적이고 정량화된 기술분석 수행이 가능하다. 본 논문에서는 C사의 특허에 대하여 전수조사를 실시하고, IPC 코드기반 분석 Matrix를 구축한 후 해당특허들을 신뢰도 기반의 연관규칙 마이닝을 실시하며 구조화된 연관맵을 생성한다. 연관맵을 이용하면 해당회사의 특허 현황 파악에 유용하게 활용된다. 또한, 구조화된 연관맵을 이용하면 상호 연관있는 기술에 대하여 군집화를 가능하게 하기 때문에, 본 논문에서 제시한 C사의 기술을 파악할 수 있으며 이를 기반으로 기술 흐름과 향후 기술 전략 수립을 가능하게 한다.

키워드 : 연관규칙마이닝, 연관맵, IPC 코드, 기술군집, 전략

Abstract

A technology is continuously developed in a rapidly changing global market. A company requires an appropriate R&D strategy for adapting to this environment. That is, the technologies owned by the company needs to be thoroughly analyzed to improve its competitiveness. Alternatively, technology classification using IPC codes is carried out recently in an objective and quantitative way. International Patent Classification, IPC is an internationally specified classification system, so it is helpful to conduct an objective and quantitative patent analysis of technology. In this study, all of the patents owned by company C are investigated and a matrix representing IPC codes of each patent is created. Then, a structured association map of the patents is made through association rules mining based on Confidence. The association map can be used to inspect the current situation of a company about patents. It also allows highly associated technologies to be clustered. Using the association map, this study analyzes the technologies of company C and how it changes with time. The strategy for future technologies is established based on the result.

Key Words : ARM(Association Rule Mining), Association map, IPC Code, Associated technologies, Strategy

Received: May, 9, 2016

Revised : Jun, 2, 2016

Accepted: Jun, 13, 2016

† Corresponding authors

hanyul@korea.ac.kr

본 논문은 2015년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(한국연구재단-NRF-2015R1D1A1A01059742). 본 논문은 BK21 플러스 사업(고려대학교, 제조·물류분야에서의 빅 데이터 운용 사업팀)으로 지원된 연구임.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

최근 금융위기와 경기침체로 기업들은 생존을 위하여 핵심 기술 탐색을 이용한 사업 다각화를 시도하고 있다. 핵심기술 탐색을 위하여 과거에는 정성적 방법인 전문가 의견조사, 델파이 기법, 시나리오 분석 등을 사용하였으나 이러한 방법들은 전문가의 주관적인 판단에 의존하며 실험설계를 위한 많은 시간과 노력이 필요하다[1]. 이러한 문제점을 극복하기 위하여 최근에는 특허문서를 이용한 정량적 분석 방법에 관한 연구가 시도되고 있다. 특허문서는 기술문서(Technology Description)이기 때문에 기술 자체의 분석이 매우 용이하다. 특허 기술 문서에는 다양한 내용을 담고 있으나 가장 대표적인 것은 국제특허분류(IPC)이다.

국제특허분류 코드는 기술문서에서 발생할 수 있는 언어장벽, 기술의 이해도 차이, 부정확한 용어 사용 등의 문제를 해결할 수 있으며 분석기술에 기술영역을 명확히 한정할 수 있다[2]. IPC를 이용한 최근 연구로는 IPC 코드 기반의 시계열, 군집분석한 결과와 키워드 및 전문가 설문조사를 병행한

IP R&D 방법론이 있다 [3]. 이 방법은 기존의 정성적 분석 방법을 보완하기 위하여 정량적 분석 방법을 병행함으로써 전문가의 경험과 객관성을 지닌 정량적 분석 데이터를 통합하여 새로운 기술예측을 실시하여 그 효과를 검증한 연구이다.

본 연구에서는 연구개발의 다각화를 위한 자사의 기술 분석이 필요한 C社의 특허를 전수조사하고, 해당특허들을 이용하여 IPC 기반의 분석 Matrix를 구축한다. 구축된 Matrix는 연관규칙마이닝 (Association Rule Mining, ARM)을 통해 신뢰성 Matrix를 생성한다. 본 논문에서는 기술개발 방향을 제시하기 위한 방법으로 구조화된 연관맵을 이용한다. 연관맵은 연관규칙마이닝의 신뢰성 Matrix를 도식화 한 것으로서, 조사대상 기업의 기술현황 파악을 용이하게 한다. 신뢰성 기반 도식화한 연관맵을 활용하여 본 연구의 최종목표인 C社의 향후 기술개발 방향을 제시하고자 한다.

본 논문의 2장에서는 특허 데이터, 연관규칙마이닝 및 기술군집화에 관한 이론적 배경 설명, 3장에서는 연구방법에 대해 기술하고, 4장에서는 특허데이터 수집, 정량분석 및 제안된 방법에 대한 실험 및 결과에 대해 설명한다. 마지막으로 5장에서는 결론 및 향후 연구방향에 대해 논의하고자 한다.

2. 이론적 배경

2.1 특허 데이터

특허문서는 기술문서이기 때문에 기술 자체의 분석이 매우 용이하다. 특허분석에서 활용하는 방법은 크게 통계적 방법 (statistical method), 경제성 분석 방법(economical analysis method) 그리고 시각화 방법(visualization method)이 있다 [4]. 특히, 시각화 방법은 최근에 부각되고 있는 방법으로, 다양한 특허지도나 특허 네트워크를 활용하여 특허간의 연관관계를 파악하는 데 용이하게

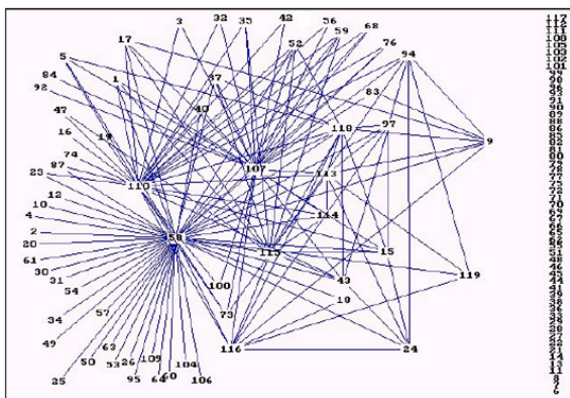


그림 1. 특허 연관관계 시각화 (예시)
Fig. 1. Association correlation of patents (example)

활용된다. 다음의 그림 1은 특허들과 이들의 상호 연관관계를 표시하는 특허 네트워크를 도시 한 것이다.

본 논문에서는 특허 기술의 상호 연관관계를 파악하기 위하여 기술영역을 명확히 정의 할 수 있는 장점을 가진 IPC코드를 기반으로 하고, 이를 시각화 방법에 적용함으로써 나타내고자 한다.

2.2 연관규칙마이닝

시장바구니분석(Market Basket Analysis)라고도 알려진 연관규칙마이닝(Association Rule Mining)은 트랜잭션 사이의 선행사건 X와 후행사건 Y의 관계를 분석하는 대표적인 데이터마이닝 기법이다 [5]. X와 Y의 관계는 지지도(Support), 신뢰도(Confidence)를 이용하여 측정할 수 있다.

규칙 (Y→X)의 지지도는 X와 Y를 동시에 포함한 트랜잭션 수를 전체 트랜잭션 수로 나눈 값으로, 일반적으로 지지도는 자주 발생하지 않는 규칙을 우선적으로 제거하는데 주로 사용되며, 신뢰도는 다음과 같은 식(1)의 형태로 나타낼 수 있다. 하지만 지지도는 P(X→Y)와 P(Y→X)를 구분하지 못하는 단점이 존재한다.

$$\text{Support}(X \rightarrow Y) = P(X \cap Y), (0,1) \quad (1)$$

이와 같은 문제점을 보완하기 위해서 연관규칙 마이닝에서는 신뢰도를 이용한다.

본 연구에서는 기술의 선후관계 뿐만 아니라 서로 차지하는 비중을 파악하기 위해 신뢰도를 이용하여 분석을 수행한다. 신뢰도는 다음과 같은 식(2)의 형태로 나타낼 수 있다.

$$\text{Confidence}(X \rightarrow Y) = p(Y|X) = \frac{P(X \cap Y)}{P(X)}, (0,1) \quad (2)$$

연구개발 방향을 분석하기 위해 신뢰도를 기준으로 연관규칙을 시각화한다. 시각화는 n개의 사건이 존재할 때, 더욱 명확하게 나타낼 수 있어 연관규칙 파악에 유리하다. 이렇게 발생한 연관규칙에서 의미 있는 연관규칙을 찾아내기 위해서는 평가기준 (evaluated criteria)이 필요하다. 연관규칙 마이닝에서는 확률측도에 대하여 최소 확률 값을 정하여 이 값보다 큰 규칙들에 대하여 의미를 부여한다. 이 값을 연관규칙 마이닝에서는 최소 임계값(minimum threshold)이라 한다.

2.3 기술군집화

군집화(Clustering)는 주어진 데이터의 각 개체에 대하여 주어진 유사성측도(Similarity measure)를 이용하여 서로 유사한 것들끼리 그룹화 하는 과정이다 [6]. 기술군집은 기술을 나타낼 수 있는 키워드,

특허분류코드 등을 이용하여 서로 다른 기술이 연관이 있거나 또는 거리로 표현하였을 때 그 위치가 서로 가까워 그 여러 기술을 한 집단으로 나타낼 수 있을 때 그 요소의 집합을 기술군집이라고 말한다. 기술군집에 포함되어 있는 키워드 혹은 특허분류코드를 통해 기술군집을 정의할 수 있다. 기술군집은 목적에 따라 다양한 방면으로 사용될 수 있다. 본 논문에서는 기술군집한 특허분류코드를 통하여 자사의 특허현황과 주요 핵심기술을 파악할 수 있으며, 이를 기반으로 향후 개발될 특허 기술에 대해서도 예측을 할 수 있는 자료로 활용될 수 있다.

3. 연구방법

본 연구는 C社의 특허를 전수조사하여 보유특허를 파악하고 이를 연관규칙 마이닝을 통하여 구조화된 연관맵을 작성한 후, 이를 바탕으로 향후 C社의 기술이 어떠한 기술을 중심으로 개발되는지에 대한 기술개발 방향을 제시하고자 한다. 제안된 방법은 다음 그림2의 Flow Chart와 같은 순서로 진행된다.

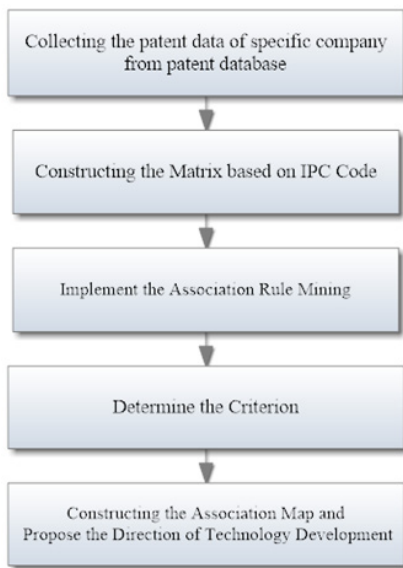


그림 2. 제안된 방법
Fig. 2. Methodology

우선, C社에서 출원한 국내특허에 대하여 특허 검색 데이터베이스를 이용하여 전수조사를 실시하여 C社의 유효특허를 수집한다. 수집한 C社의 유효특허에 대하여 IPC코드의 서브클래스 수준으로 각각의 특허에 대한 분석 Matrix를 생성한다. 다음으로는 생성된 Matrix에 대하여 연관규칙마이닝(Association Rule Mining)을 실시한다. 이때, 특허 기술의 선행개발과 후행개발을 파악하기 위하여 신뢰성(Coherence) 기반으로 마이닝을 수행한다.

선행개발과 후행개발의 유의한 규칙성을 발견하기 위해서, 임계값의 조정을 통하여 C社에 의미 있는 기술의 그룹핑이 가능하도록 임계값을 선정하여 규칙의 수를 결정한다. 선정된 규칙에 따라 확보된 Data를 기반으로 연관맵을 생성하고 이를 기반으로 C社의 현재 기술현황을 파악하고 향후 기술개발 방향을 제시한다.

4. 실험 및 결과

4.1 특허데이터 수집단계

특허검색 데이터베이스인 “FOCUST” 를 활용해 1996년에서 2016년까지 C社에서 출원한 특허에 대하여 전수조사를 실시하였다. 국가는 한국으로 설정하여 검색한 결과 총 217건의 특허데이터를 확보하였다.

표 1. 특허수집
Table 1. Patent Collection

County	Database	Search Query	Result
Korea	FOCUST	(COOOOOOO), AP	217

4.2 C社 특허데이터 기반 정량분석단계

먼저, C社의 1996년에서 2016년까지의 국내특허를 전수조사하여 총 217건의 특허자료를 수집하였으며, 이를 시계열로 도시하면 아래의 그림3과 같이 나타난다.

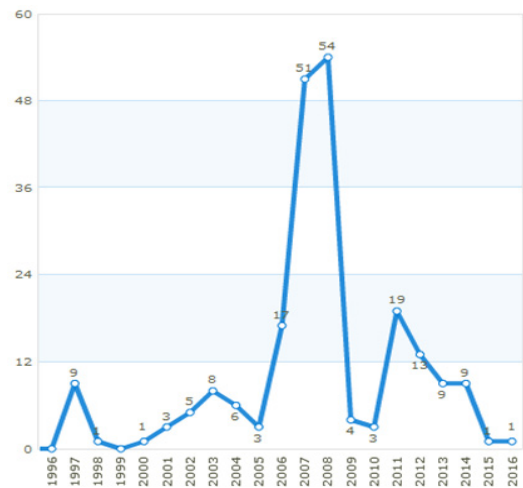


그림 3. C社 특허출원 현황
Fig. 3. Patents application status of C社

C社의 1996년 최초 특허출원 이후 2016년까지의 특허출원 현황을 살펴보면, 주목할 만한 점이 2006년부터 2008년까지 특허출원 건수가 급격하게 증가하는 것을 확인할 수가 있다. C社는

반도체/디스플레이 장비전문기업으로써, 2006년부터 2008년까지 반도체/디스플레이 장비시장이 세계적으로 호황이었던 시기에 이 분야 연구개발을 활발하게 진행하였으며, 또한 이시기에 개발한 기술에 대한 권리확보를 위하여 특허 출원을 활발하게 진행하였던 것으로 보인다.

세계적으로 반도체/디스플레이 장비의 호황기였던 2006년부터 현재의 2016년까지 C社의 특허출원 기술을 파악하기 위하여, 이 기간 동안의 출원된 특허에 대하여 IPC코드 기반으로 특허 정량분석을 시계열로 실시하였다. 아래의 그림4와 같이 IPC 코드의 H01L(전기적/반도체 특성 이용 장치), G02F(전자부품, 기타 광학기기 제조업) 그리고 H05H(특수기계 제조업) 특허가 상당수 출원되었음을 확인할 수 있다.

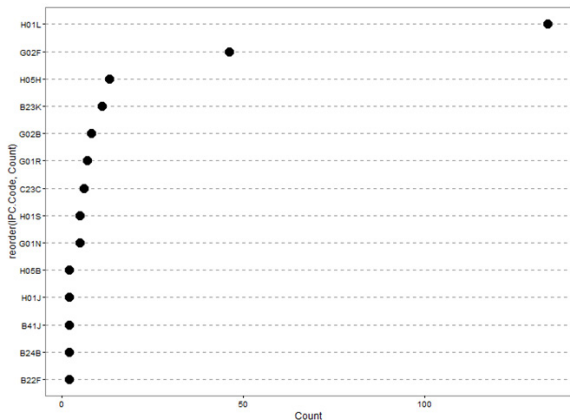


그림 4. C社 특허 분류 (기준: IPC 코드)
Fig. 4. Patent classification of C社. (Criteria : IPC Code)

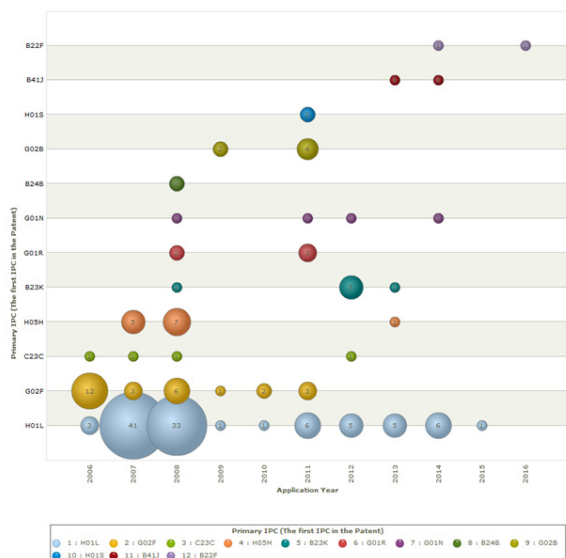


그림 5. C社의 2006년 이후 출원현황 (기준: IPC, 시계열)
Fig. 5. Application status of C社. Since 2006
(Criteria : IPC, Time series)

그림4와 그림5를 참조하면, 출원된 기술의 종류는 IPC 코드를 확인함으로써 정량적으로 파악할 수 있으나 어떠한 선행 기술이 개발되면 이와 관련된 후행 기술이 개발되는지는 파악할 수가 없다는 한계점이 존재한다.

따라서, 본 연구에서는 이를 보완하기 위하여 구조화된 연관맵을 이용한다. 연관맵은 연관규칙마ining(ARM)의 신뢰성을 도식화한 것으로써, 조사대상 기업의 기술현황 파악을 용이하게 한다. 본 논문에서는 신뢰성을 도식화한 연관맵을 활용하여 본 연구의 최종목표인 C社의 향후 기술개발 방향을 제시하고자 한다.

4.3 연관맵을 이용한 특허분석 단계

우선, C社에서 출원한 국내특허에 대하여 특허 검색 데이터베이스를 이용하여 전수조사를 실시하여 C社의 유효특허를 수집한다. 수집한 C社의 유효특허에 대하여 IPC코드의 서브클래스 수준으로 각각의 특허에 대한 분석 Matrix를 생성하며, 그 결과는 다음의 표2와 같다.

표 2. IPC코드 기반 분석 Matrix
Table 2. Analysis matrix based on IPC Code

Patent	IPC1	IPC2	IPC3
Patent 1	H01L	G02F	G01N
Patent 2	B23K	G02F	
Patent 3	H05H	H01L	
Patent 4	B23K	B23K	B28D
:	:	:	:
:	:	:	:

다음으로는 생성된 Matrix에 대하여 선행개발과 후행개발의 관계를 파악하기 위하여 연관규칙마ining(Association Rule Mining)을 실시하여 연관맵을 생성한다. 연관맵을 생성하기 위해서는 신뢰도 기반의 규칙 생성이 필요하다. 규칙의 개수는 지지도와 신뢰도를 통하여 결정되므로 실험의 반복을 통하여 C社에 의미 있는 기술의 그룹핑 가능하도록 임계값을 선정한다. 구체적인 실험방법은 다음의 표3과 같이 진행되며, 실험결과 최적의 임계값을 선정하여 연관규칙을 생성하였다.

표 3. 반복실험을 통한 임계값 선정
Table 3. Threshold decision by repeated experiment

Repeat	Confidence	Support	Rule
Experiment 1	0,001	0.5	31
Experiment 2	0,003	0.6	26
Experiment 3	0,003	0.7	24
Experiment 4	0,0045	0.7	24
Experiment 5	0,005	0.7	1

표3의 실험1에서는 Confidence 값이 0.001이고, Support 값이 0.5

일 경우 31개의 연관규칙이 발생하였음을 확인 할 수 있으며, 실험 2에서는 Confidence 값이 0.003이고, Support 값이 0.6일 경우 26개의 연관규칙이 발생하였음을 확인 할 수 있었다. 이렇게 실험을 반복하여 C社에 의미 있는 기술의 그룹핑이 가능한 임계값을 선정하였다.

생성된 규칙을 바탕으로 아래의 그림6과같이 구조화된 연관맵을 작성한 후, 기술군집 작업을 실시한다. 각각의 기술군집은 관련성이 깊은 기술들에 모여 있으므로 이를 활용하면 C社의 기술현황을 파악할 수 있으며, 이를 기반으로 향후 기술개발 전략수립에 이용할 수 있는 자료로 활용될 수 있다.



그림 6. 연관맵
Fig. 6. Association map

실험결과 4개의 기술이 군집되어 있는 것을 확인할 수 있다. 이를 통하여 C社의 연구개발은 4개의 기술군 중심으로 수행되었음을 알 수 있다. 4개의 기술군집 중에서 가장 넓게 분포되어있는 H01L(반도체 기술 관련 전기적/반도체 특성 이용 장치), B23K(디스플레이 기술 관련 특수 기계제조업)의 기술이 C社에서 활발하게 개발되고 있는 기술임을 알 수 있다. 또한, B22F(기타 금속 가공제품 제조업) 기술도 여러 기술과 연관되어 있음을 확인 할 수 있으며, 특히 연관 특허 건수가 많음을 의미하는 노드 굵기를 통하여 서로 많은 특허가 연관되어 있음을 확인 할 수 있다. 따라서, C社는 향후에도 반도체/디스플레이 분야 관련 H01L, B23K 및 B22F 기술을 활발하게 개발 할 것으로 예측할 수 있다.

5. 결론 및 향후 연구

본 논문에서는 C社의 특허를 전수조사하여 특허를 확보하고 IPC 코드 기반의 연관규칙마이닝(Association Rule Mining: ARM) 신뢰성 Matrix를 생성 후, 실험의 반복을 통하여 해당회사에 대한 최적의 임계값을 선정 및 적용하여 구조화된 연관맵을 형성함으로써 기술을 군집화 하였다. 이렇게 본 논문의 실험을 통하여 C社 특허 기술의 선후 관계를 파악할 수 있는 연관맵을 얻을 수 있었다. 이를 활용하여 해당 기업의 기술개발 현황을 파악할 수 있으며, 또한, 이를 기반으로 향후 기술개발 동향 트렌드와 기술개발 전략 수립 시 참고할 수 있는 기반 자료로 사용할 수 있다.

하지만, 본 논문에서는 해당회사의 기술현황을 파악할 수 있었지만, 기술현황과 시계열적 흐름을 동시에 파악할 수는 없었다. 디스플레이/반도체 등의 IT(Information Technology) 기술은 제품수명주기가 매우 빠르게 변화하기 때문에, IT기술의 과거 기반기술 현황과 현재 개발되고 있는 기술을 시계열적으로 동시에 파악할 수 있다고 한다면, IT 분야 등 제품수명주기가 빠른 기술에 대한 기술개발 전략 수립에 유용한 자료로 활용 될 것이다. 또한, 이를 확장하여 IT융합 기술에도 적용하여 향후 기술전략 수립에 유용한 자료로 활용될 수 있을 것이다.

향후에는 C社의 유관 디스플레이/반도체 경쟁기업들에 대해서도 연관맵을 구축하고, 과거의 기술과 현재의 기술개발 흐름을 파악할 수 있도록 시계열적으로 추가 반영되는 연관맵 구축 연구가 추가적으로 진행되어야 할 것이다.

References

- [1] J. Lee, J. Lee, G. Kim, S. Park, D. Jang, "Establishment of Strategy for Management of Technology Using Data Mining Technique", *Journal of Korean Institute of Intelligent Systems*, Vol. 25, no. 2, pp. 126-132, 2015.
- [2] T. Montecchi, D. Russo, Y. Liu, "Searching in Cooperative Patent Classification: Comparison between keyword and concept-based search", *Advanced Engineering Informatics*, Vol. 27, pp. 335-345, 2013.
- [3] S. Jun, S. Lee, J. Ryu, S. Park, "A Novel Method of IP R&D using Patent Analysis and Expert Survey", *Queen Mary Journal of Intellectual Property*, Vol. 5, no. 4, pp. 474-494, 2015.
- [4] W. Lee, "Analysis of technology industry linkage and korean firms' patent strategy utilizing information from patent database" *Science & Technology policy institute*, pp. 16-17,

- 2004.
- [5] J. Kim, "Structured Association Map for Visualizing Association Rules", *The Korea Contents Society Conference*, pp. 325-326, 2015.
 - [6] S. Jun, S. Park, D. Jang, "Patent analysis & technology forecasting", *Kyowoo publishing company*, pp. 76-78, 2014.
 - [7] J. Park, H. Chung, "An Effective Reduction of Association Rules using a T-Algorithm", *Korean Institute of Intelligent System*, Vol. 19, no. 2, pp. 285-290, 2009.
 - [8] J. Lee, H. Chung, J. Kim, "Implementation of purchasing pattern classification system using neural network and association rules", *Korean Institute of Intelligent System*, Vol. 13, no. 5, pp. 530-538, 2003.
 - [9] S. Jun, "A big data learning for patent analysis", *Korean Institute of Intelligent System*, Vol. 23, no. 5, pp. 406-411, 2013.
 - [10] K. Lee, K. Lee, "Mining of subspace contrasting sample groups in microarray data", *Korean Institute of Intelligent System*, Vol. 21, no. 5, pp. 569-574, 2011.

저 자 소 개



송원호(Wonho Song)

2005년 : 단국대학교 전자공학 공학사
 2015년~현재 : 고려대학교 공학대학원
 지식재산학 석사과정

관심분야 : Patent Analysis, Technology Evaluation, Patent Management, Data Mining
 Phone : +82-2-3290-4124
 E-mail : sky7312114@korea.ac.kr



이준석(Junseok Lee)

2013년 : Zhejiang University, Mechanical Engineering and Automation, 공학사
 2014년~현재 : 고려대학교 대학원
 산업경영공학부 석·박사
 통합과정

관심분야 : Patent Analysis, Data Mining, Management of Technology
 Phone : +82-2-3290-3900
 E-mail : jxli12@korea.ac.kr



박상성(Sangsung Park)

2006년 : 고려대학교 대학원
 산업시스템정보학과 공학박사
 2006년~2014년 : 고려대학교 산업경영공학부
 연구교수
 2014년~현재 : 한국지식재산전략원 평가위원

2015년~현재 : 고려대학교 기술경영전문대학원 조교수
 2016년~현재 : 지식재산창조협의회 운영위원

관심분야 : Patent Analysis, Data Mining, Management of Technology, Technology Evaluation
 Phone : +82-2-3290-4618
 E-mail : hanyul@korea.ac.kr