

# 불완전 인용정보 하에서의 특허의 기술적 중요도 평가 모형\*

김 현  
백석대학교 경상학부  
(hkim@bu.ac.kr)

백동현  
한양대학교 경상대학 경영학부  
(estarbaek@hanyang.ac.kr)

신민주  
삼양 데이터 시스템 e-Biz  
(shinmj@syds.com)

.....

국내 연구관리 기관에서는 연구비 지원에 대한 결과물의 하나로 출원된 특허 기술을 관리하고 있으나, 체계적인 방법의 부족으로 특허의 질적인 수준을 평가하는 데 어려움을 겪고 있다. 특히, 국내 특허의 경우 인용정보를 기록하는 것이 의무화되어 있지 않아 인용정보를 이용한 특허 가치 평가가 어려운 상황이다. 본 연구는 인용정보가 없거나 미비한 상황에서 특허의 기술적 중요도를 평가할 수 있는 방법론을 제안한다. 제안된 방법론은 평가대상 특허 선정, 관련 특허군 수집, 주제어 출현빈도 행렬 작성, 특허군의 군집화, 그리고 기술적 중요도 평가 등의 5 단계를 거쳐 평가대상 특허의 가치를 기술적 중요성 측면에서 객관적이고 정량적으로 평가할 수 있도록 한다. 제안된 방법론을 '휴대단말기를 이용한 대금결제'를 위한 사용자 인증방법' 특허기술에 적용하여 그 기술적 중요도를 평가하였다.

.....

논문접수일 : 2007년 07월      게재확정일 : 2008년 06월      교신저자 : 백동현

## 1. 서론

특허는 발명을 장려·보호·육성함으로써 기술의 진보와 발전을 도모하고 국가산업의 발전에 기여하는 제도이자 기존의 기술 변화에 대한 연구과정에서 기술의 본질이나 특성을 파악할 수 있는 객관적이고 표준적인 기술정보이다. 이러한 특허의 특성을 활용하기 위해서 선진국에서는 특허 간의 인용관계를 통하여 특허간의 기술적 연관관계뿐만 아니라 특허의 상대적인 중요도 등을 분석하고 있다. 특허 기술의 가치는 매우 높은 수준의 편향성, 즉 소수의 핵심기술이 매우 높은 가치를 가지는 반면 대다수의 많은 기술들은 별다른 가치를

갖지 못한다는 특성을 나타내고 있다(Swann, 1993). 따라서 특허 기술의 가치라는 관점에서 기술 확산을 평가하고 예측하기 위해서는 단순하게 특허의 빈도만을 이용해서는 정확한 평가가 어려우며 더욱이 개별 기술을 분석 대상으로 삼는 경우에는 동일하게 개별 기술 단위에서 측정될 수 있는 변수의 도입이 요구되고 있다.

Harhoff et al.(2003)은 특허기술의 중요성이나 영향력 등을 고려하여 특허기술의 가치를 평가하는 것이 이론적으로 타당함을 제시하고 있다. 그의 연구에서는 특허기술의 가치, 또는 가치변화에 영향을 주는 요인에 대한 선행 연구요인들이 제시되고 있는데, 기술의 특허 영역, 특허 인용빈도, 특허

\* 이 논문은 한양대학교 일반연구비 지원으로 연구되었음(HY-2007-G).

의 범위, 특허 청구 범위의 채택 정도 등은 모두 특허의 기술적 가치와 정의 상관관계를 가지는 것으로 검증되고 있다.

특허정보를 위한 다양한 분석기법이 사용되고 있지만 인용 분석은 가장 빈번하게 채택되는 기법 중의 하나이다(Narin, 1994). 인용분석은 특정 연구자 또는 연구기관의 연구 성과가 후속연구에서 얼마나 많이 인용되었는가 하는 정도를 측정하여 연구자 또는 연구 성과의 영향력이나 중요성을 평가하는 것이다. 즉, 특정 논문이나 기술이 장점이 많고 영향력이 높을수록 더 많이 인용되리라는 것이다(Albert et al., 1991; Mcmillan and Hamilton, 2000).

이와 같이 특허 데이터를 이용한 분석, 특히 인용정보를 이용한 특허 가치평가에 대한 연구가 선진국을 중심으로 활발히 진행되고 있으나 국내의 경우 특허의 인용정보를 관리하는 것이 의무화 되어 있지 않아 인용정보를 이용한 특허 가치평가가 어려운 상황이다.

본 연구에서는 특허 간의 인용정보가 없거나 미비한 상황에서 특허의 기술적 중요도를 객관적이고 정량적으로 평가하기 위한 방법론을 개발하였다. 제안된 방법론은 평가대상 특허선정, 관련 특허군 수집, 주제어 출현빈도 행렬 작성, 특허군의 군집화, 기술 중요도 평가 등의 5단계로 구성되어 있다. 제안된 방법론을 '휴대단말기를 이용한 대금 결제를 위한 사용자 인증방법' 특허기술에 적용하여 그 기술적 중요도를 평가하였다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서 특허의 기술적 중요도의 개념과 특허의 기술적 중요도와 관련된 기존 연구를 정리한다. 제 3장에서는 본 연구에서 제안하는 특허의 기술적 중요도 평가 모형을 제시하였으며, 제 4장에서는 개발한 모형을 토대로 실제 사례를 적용한 결과에 대해

설명한다. 마지막으로 제 5장에는 본 논문의 결론과 향후 연구 과제를 정리하였다.

## 2. 이론적 배경

인용분석은 많은 비판에도 불구하고 특허 데이터의 기술적·경제적 분석에 있어 그 우수성을 많은 연구자들이 제시하고 있다. Kutznets(1962)와 OECD(1994)의 연구에서도 데이터의 확보 가능성, 데이터 범위의 포괄성, 그리고 개별 기술에 대한 정보량 등 특허 데이터가 다른 지표에 비하여 상대적으로 유용하다는 주장이 제기되었다. 특히, 인용분석을 통해 기술영향력지수와 특허기술의 수명주기(Hirschey and Richardson, 2001), 기업 간 교섭력(Mowery et al., 1998), 혁신활동의 경제적 가치, 그리고 국가 간 기술연계와 지식이전(Tijssen, 2001) 등과 같은 지표들이 도출 가능하며, 이러한 지표들은 기술자산의 질적 수준을 측정할 수 있는 척도로 사용할 수 있다는 점이 연구자들에 의해 제시되고 있다.

이처럼 특허 데이터의 분석에 있어 인용분석은 가장 빈번하게 채택되는 분석기법 중의 하나이지만 전체특허의 연결구조를 파악하기 어렵고, 2개 특허 간의 연관성만으로는 분석의 범위와 깊이가 제한적이며, 인용의 빈도와 같은 피상적인 분석으로는 특허 기술 간의 내재적인 관계를 분석하기 어려우며, 인용분석을 위해서는 자료 수집 및 분석에 많은 시간이 소요된다는 단점을 가지고 있다(Yoon and Park, 2004).

인용분석 이외의 방법으로 특허가치를 평가한 것으로, Scotchmer(1991, 1996)의 연구에서는 특허의 영역(patent scope)이 특허기술의 가치에 미치는 영향이 크다고 하였으며, 지표로는 IPC 분류 코드에 따른 특허 기술의 범위가 넓을수록 가치가

높을 것이라는 연구 결과가 있다. Lanjouw(1998)의 연구에서는 특허등록 국가의 수가 많을수록 특허 기술의 가치에 영향을 미칠 것이라는 연구 결과가 있다. Yoon and Park(2004)의 연구는 인용분석의 단점을 보완하기 위해 상호 간의 인용관계로 구성되는 인용관계 네트워크에 대한 구조를 파악하여 어떤 기술이 가장 중요한 기술인지 평가할 수 있도록 하는 네트워크 분석 방법론을 제안하였다. 네트워크 분석은 객체간의 관계(혹은 연결망)에 대한 분석을 통해 전체 구조를 연구하는 방법을 말한다. 네트워크 분석 방법은 그동안 특허의 인용관계를 분석함에 있어 개별 특허기술의 인용빈도, 인용간격 등의 산술적인 특성의 계산에 머무르던 분석의 수준을 높여줄 수 있는 효율적인 기법을 제공하고, 특허기술 간의 연관성을 시각적으로 제시함으로써 특허기술의 구조를 직관적으로 이해할 수 있도록 한다. 또한 단순한 인용빈도의 계산 이외에 다양한 지표를 산출할 수 있으므로 보다 풍부한 분석이 가능하다.

인용정보가 없거나 불완전하여 인용정보를 이용한 특허 가치평가가 어려운 경우 텍스트 마이닝 기법을 활용하여 특허 간의 연관성을 파악한 뒤 이로부터 특허의 기술적 중요도를 평가하는 접근이 가능하다. 텍스트 마이닝 기법은 거대한 텍스트 집합으로부터 이해할 수 없는 지식을 확실히하고 잠재적으로도 유용하게 획득하기 위한 기법으로 인터넷 탐색 엔진, 데이터베이스 관리 도구 등의 예를 들 수 있다. 텍스트 마이닝 기법을 이용하여 주어진 문제에 적절하게 적용함으로써 대량의 텍스트 문서들을 대상으로 서로 유사성이 높은 문서들끼리 그룹화한다거나 빈번하게 출현하는 단어들을 특성화 하는 등의 지식을 발견해 낼 수 있다(Larsen and Aone, 1999).

텍스트 마이닝 분석기법은 그 방법에 따라 문서

의 범주화, 군집화, 요약의 세 가지로 분류할 수 있다. 문서의 범주화는 사전에 정의된 고정된 범주 항목으로 문서를 분류하는 것으로 문서들의 계층적 구조를 형성한다. 문서의 군집화는 문서들 간의 유사도를 측정하여 유사하다고 판단되는 문서들끼리 그룹화 시킨다. 유사도에 따라 문서들을 트리로 표현할 수 있으며 비교적 일정한 유사도로 그룹화 된 텍스트 집합의 독립적인 무리를 형성하고 있는 그룹을 얻을 수 있다. 문서 요약은 문서의 전체 내용을 대표할 수 있는 내용으로 문서의 길이를 줄이는 작업을 말한다. 본 연구에서는 특허기술의 주제어 출현빈도 행렬을 이용하여 특허들 간의 유사도를 측정하여 유사한 특허들 간의 그룹화를 통해 관계구조를 타당성 있게 도출하였다. 국내의 경우, 특허와 같이 인용관계에 관련된 별도의 정보가 없는 경우 매우 유용한 방법으로 사용될 수 있다. 인터넷 환경 하에서 텍스트 마이닝은 대량의 웹 페이지 데이터로부터 유용하고 이해 가능한 정보를 추출하기 위해 많이 활용되고 있다(김경재와 김병국, 2005; 이장희 외, 2002).

### 3. 특허의 기술적 중요도 평가 모형

본 연구에서 제안하는 특허의 기술적 중요도를 평가하는 방법론은 <그림 1>에 설명된 것과 같이 평가대상 특허 선정, 평가대상 특허 관련 특허군(S<sub>1</sub>) 수집, 주제어 출현빈도 행렬 작성, 특허군의 군집화, 그리고 기술적 중요도 평가 등 5단계로 구분되어 있다.

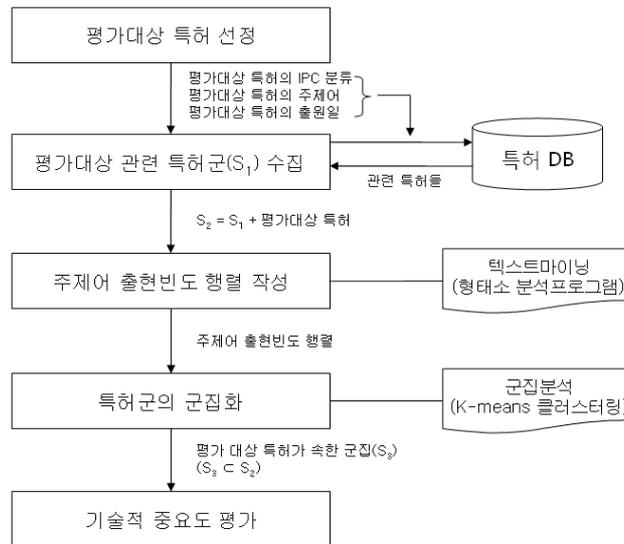
먼저, 기술적 중요도를 평가하고자 하는 평가대상 특허를 선정하고, 선정된 평가대상 특허의 IPC 분류, 주제어, 출원일 등의 정보를 이용하여 평가대상 특허와 관련성이 높은 관련 특허군 S<sub>1</sub>을 특허 DB로부터 구한다. 다음으로, 형태소 분석프로

그램을 이용해 관련 특허군  $S_1$ 과 평가대상 특허로 구성된 특허군  $S_2$ 에 대해 주제어 출현빈도 행렬을 작성한다. 다음으로, 주제어 출현빈도 행렬에 K-means 클러스터링 기법을 적용하여 특허들은 유사도가 높은 몇 개의 군집으로 군집화한다. 마지막으로 평가대상 특허가 속한 군집  $S_3$ 를 대상으로 특허의 기술 중요도를 구한다. 이하에서는 각 단계에 대해 자세히 설명한다.

### 3.1 평가대상 특허 선정

이 단계에서는 기술적 중요도를 평가하고자 하

는 ‘평가대상 특허’를 선정하고, 선정된 평가대상 특허의 IPC 분류, 주제어, 출원일 등의 정보를 획득한다. IPC(international patent classification)란 국제 특허분류를 말하며, 출원된 발명의 기술내용을 파악하여 해당 기술이 속하는 기술 분류를 명확히 정하도록 한다. IPC 분류는 <그림 2>에 있는 것처럼 A~H까지 섹션을 분류하고 각 섹션별로 클래스로 세부 분류를 하고 각각의 클래스는 서브클래스로 분류된다. 또한, 각 서브클래스는 “그룹”이라고 칭하는 세부 전개항목으로 전개된다. “그룹”은 메인그룹(즉, 분류계층구조의 4번째) 또는 서



<그림 1> 기술적 중요도 평가 방법론

A 섹션-첫 번째	01 클래스-2번째	B 서브클래스-3번째	33/00 or 33/08 그룹	메인그룹-4번째 서브그룹-최하위
--------------	---------------	----------------	----------------------------	----------------------

<그림 2> IPC 분류체계

브그룹(즉, 메인그룹 계층의 분류보다 더 낮은 분류계층)으로 이루어진다.

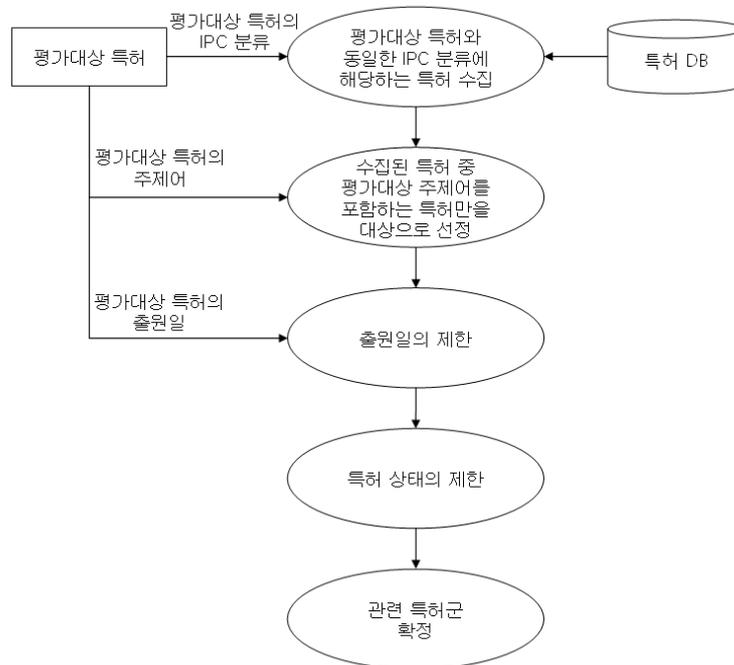
관련 특허군을 수집하기 위해서는 선정된 평가대상 특허로부터 주제어를 추출하여야 하는데, 본 연구에서는 평가대상 특허의 제목과 초록 부분에서 명사를 대상으로 주제어를 선정한다. 주제어 선정은 제목과 초록만으로 한정된 이유는 해당 특허를 대표할 수 있는 내용이 제목과 초록에 대부분 포함되었을 것으로 진제하였기 때문이다. 주제어는 제목과 초록을 검토하여 분석자가 선정하되 일상적으로 쓰이거나 내용이 포괄적인 명사는 제외한다.

### 3.2 관련 특허군 수집

다음은 평가대상 특허와 관련이 있을 것으로 예

상되는 관련 특허군을 수집하는 단계이다. 특허들간의 인용정보가 완전하게 존재한다면 그 정보를 이용하여 비교 대상이 될 특허들을 수집할 수 있지만, 그렇지 않은 상황에서는 비교 대상이 될 특허군을 합리적인 방법으로 수집하여야 한다. 본 연구에서는 <그림 3>과 같은 일련의 절차에 의해서 관련 특허군을 수집한다.

먼저, 평가대상 특허와 동일한 IPC 분류에 해당하는 특허들을 수집한다. 본 연구에서는 IPC 분류를 적용함에 있어 첫 4자리 코드(예 : A01B)만을 이용하는데, 특허 DB로부터 평가대상 특허와 동일한 첫 4자리 코드를 갖는 특허들을 수집한다. IPC분류 코드의 첫 4자리 코드만을 이용하는 이유는 섹션(A)이나 클래스(A01) 수준만을 이용하는 경우 대상이 되는 특허가 너무 많을 뿐만 아니라



<그림 3> 관련 특허군 수집 절차

평가대상 특허와의 관련성도 적게 된다. 반면, 전 체크코드(예 : A01B 33/08)를 이용할 경우 너무 세분화되는 경향이 있으며 특허 기술의 형식(회로, 도면 등)까지 포함하기 때문에 적합하지 않다.

다음은, IPC 분류를 이용해 수집된 특허들에 주제를 적용하여 그 대상을 한정한다. IPC 분류의 첫 4자리를 이용한다 하더라도 수집된 특허들의 수는 여전히 많으며 평가대상 특허와 연관성이 높다고 말하기 어렵다. 따라서 수집된 특허들 중에 제목과 초록 부분에 평가대상 특허의 주제를 포함하는 특허만으로 그 대상으로 한정한다. 제목과 초록을 대상으로 하는 것은 대표성을 가지는 부분이기 때문이다. 특허 DB에 있는 색인어도 대표성을 가진다고 할 수 있으나, 색인어 등록이 의무로 되어있지 않으므로, 경우에 따라서는 색인어가 등록되어 있는 않은 특허가 존재하여 주제어 검색 대상으로 선정할 수 없다.

다음은, 앞의 두 단계를 거쳐 수집된 특허들을 대상으로 출원일 기준을 적용한다. 오래된 특허일수록 중요성이나 영향력이 과대평가되는 경향이 있으므로 평가대상 특허와의 관계를 고려함에 있어서 오래된 특허를 제외하여 평가대상 특허가 과소평가되는 것을 막는다. 본 연구에서는 평가대상 특허의 출원일 기준 5년 이내에 출원된 특허부터 현재까지의 특허로 출원일을 제한한다. 예를 들어, 평가대상 특허가 2003년에 출원되었다면 앞의 두 단계에서 수집된 특허들 중에서 1999년부터 현재 시점까지의 출원일을 가지는 특허만을 선정하게 된다.

마지막으로, 수집된 특허들의 특허 상태를 참조한다. 특허의 상태는 공개, 등록, 거절, 포기 4가지가 있으며, 본 연구에서는 공개 또는 등록 상태인 특허를 대상으로 수집하며, 거절 또는 포기 상태인 특허는 분석에서 제외시킨다.

### 3.3 주제어 출현빈도 행렬 작성

다음 단계는 평가대상 관련 특허군  $S_1$ 과 평가대상 특허로 구성된 특허군  $S_2(S_2 = S_1 + \text{평가대상 특허})$ 에 대해 주제어 출현빈도 행렬을 작성한다. 이 단계에서는 특허군  $S_2$ 에 속한 각 특허에 대해 주제를 추출하고 표준화시킨 뒤 주제어 출현 빈도 행렬을 작성한다. 주제어 추출 대상은 제목, 요약, 초록, 그리고 발명의 상세한 설명을 나타내는 발명의 목적, 발명이 이루고자하는 기술적 과제, 발명의 구성 및 작용, 발명의 효과 등에서 명사를 대상으로 추출한다. 청구의 범위(청구항)와 색인어는 추출대상에서 제외한다. 청구의 범위(청구항)의 경우 특허기술을 나타낼 수 있다고 할 수 없으므로 추출대상에서 제외하고, 색인어의 경우 필수 등록 항목이 아니고, 다수의 특허에서 색인어가 등록되어 있지 않은 경우가 많아 추출 대상에서 제외한다.

형태소 분석기 프로그램을 이용하면 이러한 주제어들의 추출을 편리하게 수행할 수 있다. 형태소란 단어(어절)를 구성하고 있는 더 이상 별개의 의미나 기능으로 분리할 수 없는 최소한의 언어 요소라고 할 수 있고, 형태소 분석기 프로그램은 형태소를 분석해내는 프로그램을 뜻한다.

일반적으로 각 특허마다 추출되는 주제어들의 수는 매우 많아 이를 모두 이용할 경우 분석이 매우 복잡하게 된다. 따라서 본 연구에서는 각 특허에서 추출된 주제어들 중 상위빈도 3개의 주제어(단어)만을 선정하며, 이 과정에서 일반명사는 제외하고 복합명사는 분리하여 분석한다. 특허군  $S_2$ 로부터 추출된 주제어들(각 특허별로 3개씩)은 표준화 과정을 통해 동일한 의미이지만 다르게 사용되는 경우(예, 데이터, 데이타, data)를 하나로(예, 데이터) 표준화시킨다. 만약 특허군  $S_2$ 에 있는 특허들의 수가 100개라면 각 특허별로 3개씩의 주제

어들이 추출되므로 총 300개(100×3)의 주제어들이 존재할 수 있지만 이러한 표준화 과정을 통해 300개보다 적은 수의 주제어들만 남게 된다.

다음은, 추출된 주제어들을 이용해 출현빈도 행렬을 작성한다. <표 1>은 주제어와 특허간의 출현빈도 행렬을 나타낸다. <표 1>에서 특허 1에는 주제어 1은 35번, 주제어 2는 32번, 주제어 k는 14번 나타나고 있음을 보여준다. 이를 각각의 특허와 주제어별로 출현 빈도에 따라 행렬로 나타낸다. 출현빈도 행렬에서 5% 이하의 특허(100개의 특허라면 5개 이하의 특허)에만 출현하는 주제어들은 제거하여 출현 빈도 행렬을 작성한다. 그 이유는 제거된 주제어의 경우 특정 특허에만 관련되어 있는 특이성이 높은 단어일 가능성이 높기 때문이다.

### 3.4 특허군의 군집화

군집분석(clustering)이란 관측 대상이 지니고 있는 다양한 특성의 유사성을 바탕으로 동질적인 군집으로 결합하거나 여러 대상들을 몇 개의 동질

적인 군집으로 구분하는 통계적 기법이다. 군집분석에서는 분석에 사용할 변수를 선정하고, 분석대상을 변수에 근거하여 군집화 하는 것으로 분석대상 사이의 거리를 측정하여 서로 동질적인 집단으로 묶어주는 방법이다. 거리 측정 방법으로는 유클리디안 거리(euclidian distance)와 유클리디안 제곱 거리 등이 사용된다.

평가대상과 관련되어 있는 특허들을 수집함에 있어서 <그림 3>의 절차에 따라 특허들을 수집하지만, 좀 더 관련성이 높은 특허들을 대상으로 하기 위하여 군집분석을 실시한다. 이 때, 군집분석에 사용된 변수는 각각의 특허에서 추출된 주제어를 이용한다. <표 2>는 주제어 출현빈도 행렬에 대하여 군집분석에서 많이 사용되는 K-means 클러스터링 방법을 적용하여 특허들을 몇 개의 군집으로 구분한 결과이다.

<표 2>는 각각 2개, 3개, 4개의 군집으로 나뉘었을 때의 결과를 보여 주고 있으며 표 안에 있는 숫자는 해당 군집에 속한 특허들의 수를 나타낸다. 예를 들어, 2집단 군집은 평가대상 특허와 관련 특

<표 1> 주제어 출현빈도 행렬

	주제어 1	주제어 2	...	주제어 k
특허 1	35	32	...	14
특허 2	11	0	...	3
...	...	...	...	...
특허 n	0	2	...	6

<표 2> 군집분석 결과

	2집단 군집	3집단 군집	4집단 군집
군집 1	16개	15개	26개
군집 2	N-16개	7개	16개
군집 3		N-23개	27개
군집 4			N-69개
합계	N개	N개	N개

주) ■ : 평가 대상 특허가 포함된 군집.

허군 전체의 집단을 2개의 군집으로 나눈 결과로 군집 1에는 16개의 특허가 포함되어 있고, 군집 2에 포함된 특허의 수는 전체 특허의 수 N에서 군집 1에 포함된 16개의 특허를 제외한 나머지(N-16)개의 특허가 포함되어 있다. 3집단 군집의 경우 평가대상 특허와 관련 특허군 전체의 집단을 3개의 군집으로 나눈 것이고, 4집단 군집은 4개의 군집으로 나눈 것이다. 또한, 음영으로 표시된 부분은 평가대상 특허가 포함된 군집을 의미한다. 몇 개의 군집으로 나눌 것인가는 분석자가 그 수를 지정하거나 군집분석 프로그램이 자동으로 군집의 수를 결정할 수 있다. 이렇게 군집이 형성되면, 모든 특허들을 대상으로 하기보다 평가대상 특허가 포함된 군집에 포함된 특허만을 대상으로 이하의 분석을 수행하는 것이 더욱 의미 있는 결과를 얻을 수 있다.

### 3.5 기술적 중요도 평가

마지막 단계는, 평가대상 특허가 속한 군집  $S_3$ 를 대상으로 특허의 기술적 중요도를 구한다. 군집  $S_3$ 는 바로 앞 단계인 ‘특허군의 군집화’를 통해 특허군( $S_2$ )을 동질성을 갖는 몇 개의 군집으로 나누었을 때 평가대상 특허가 속한 군집이다. 기술적 중요도 평가는 주제어 출현빈도를 이용한 특허간 유클리디안 거리 계산, 유사도 행렬 작성, 기술적 중요도 계산 등의 단계를 거친다.

먼저, 특허  $i$ 와 특허  $j$ 의 주제어 출현빈도 행렬 ( $n_{ij}$ )이 <그림 4>와 같다면 식 (1)을 이용해 두 특허 간 유클리디안 거리( $d_{ij}$ )를 구할 수 있다.

임의의 두 특허  $i$ 와  $j$ 간의 유클리디안 거리를 구하였으면 식 (2)를 이용해 두 특허 간 유사도( $a_{ij}$ )를 계산할 수 있다. 유사도 값은 0에서 1사이의 값을 갖도록 표준화하였으며, 값이 클수록 두 특허 간의 유사도가 크음을 의미한다. <표 3>은 이러한 과정을 통해 구한 유사도 행렬의 예이다.

	주제어 1	주제어 2	...	주제어 k
특허 $i$	$n_{i1}$	$n_{i2}$	...	$n_{ik}$
특허 $j$	$n_{j1}$	$n_{j2}$	...	$n_{jk}$

<그림 4> 특허  $i$ 와 특허  $j$ 의 주제어 출현빈도 행렬

$$d_{ij} = \sqrt{(n_{j1} - n_{i1})^2 + (n_{j2} - n_{i2})^2 + \dots + (n_{jk} - n_{ik})^2} \quad (1)$$

$$a_{ij} = \frac{\max_{k, l \in S_3}(d_{kl}) - d_{ij}}{\max_{k, l \in S_3}(d_{kl}) - \min_{k, l \in S_3}(d_{kl})} \quad (2)$$

<표 3> 유사도 행렬

	특허 $i$	특허 $j$	특허 $k$
특허 $i$	1	$a_{ij}$	$a_{ik}$
특허 $j$	$a_{ji}$	1	$a_{jk}$
특허 $k$	$a_{ki}$	$a_{kj}$	1

유사도 행렬로부터 임의의 특허  $i$ 의 기술적 중요도( $TI(i)$ )를 계산할 수 있다. 특허  $i$ 의 기술적 중요도는 식 (3)을 이용해 구할 수 있으며 그 값이 클수록 기술적 중요도가 높다.

$$TI(i) = \frac{\sum_{j=1}^k a_{ij} (j \neq i)}{k-1} \quad (3)$$

이렇게 계산된 기술적 중요도는 시간적 선후관계를 고려하지 않은 채 특허군 내에서 특정 특허가 차지하는 중요도의 크기를 나타낸다. 그러나 그 특허가 후속 특허들에 미치는 영향의 정도를 파악하려면 해당 특허 이후에 출원된 특허에 대한 영향력을 측정할 수 있어야 한다. 특허의 경우 출원일이라는 시간 정보를 가지고 있으므로 해당 특허의 출

원일의 이후에 출원된 특허들만을 고려한 기술적 중요도를 계산할 수 있다. 이런 분석을 통하여 평가대상 특허가 이후 출원된 특허에 어느 정도의 영향력(과급력)을 가지고 있는가를 측정할 수 있다. 이에 대한 자세한 설명은 제 4장에서 하기로 한다.

#### 4. 평가 모형의 적용

본 연구에서 제안한 방법론을 실제 특허 사례에 적용하여 검증하였다. 이하에서 각 단계별로 자세히 설명하기로 한다.

##### 4.1 평가대상 특허 선정 및 관련 특허군 수집

본 연구에서는 ‘휴대단말기를 이용한 대금결제 를 위한 사용자 인증방법’을 평가대상 특허로 선정 하였으며, 추출된 주제어는 휴대단말기, 대금결제, 사용자인증, 휴대폰결제이다(<표 4> 참조). 주제어를 선정할 때 인터넷 쇼핑물과 같은 일반적인 단어는 제외하였다.

관련 특허군 수집을 위해서 한국특허정보원 특허기술정보서비스(<http://www.kipris.or.kr/>) DB에

<표 4> 평가대상 특허

제목(특허명)	휴대단말기를 이용한 대금결제를 위한 사용자 인증방법
출원일	2003년 6월 17일
IPC분류코드	H04Q 7/24
추출된 주제어	휴대단말기, 대금결제, 사용자인증, 휴대폰결제

서 IPC코드 앞 4자리(H04Q)를 적용하여 14,313개의 특허를 수집하였다. 14,313개의 특허 중 제목과 초록에(휴대단말기 or 대금결제 or 사용자인증 or 휴대폰결제)의 조건을 추가하고, 출원일을 1999년 1월 1일~2005년 3월 16일까지로 제한하고, 특허의 상태도 공개 또는 등록으로 제한하여 71개 특허를 관련 특허군으로 선정하였다. 관련 특허군 71개와 평가대상 특허로 구성된 72개의 특허에 대해 제목, 요약, 발명의 상세한 설명, 출원일, IPC 코드, 상태 등을 정보를 획득하였다.

##### 4.2 주제어 출현빈도 행렬 작성

이렇게 얻어진 72개 특허의 제목, 요약, 발명의 상세한 설명 부분을 대상으로 형태소 분석을 실시

<표 5> 각 특허의 형태소 분석 결과

특허 1		특허 2		특허 3		특허 4		...
주제어	빈도	주제어	빈도	주제어	빈도	주제어	빈도	...
데이터	35	결제	39	상품	44	RF	44	...
다운	33	번호	26	정보	40	시스템	42	...
로드	33	사업자	22	방송	38	카드	40	...
다운로드	32	대금	20	상기	35	교통	34	...
특허	23	고안자	20	구매	33	버스	25	...
단말기	21	계좌	20	상품정보	27	금액	21	...
방법	20	계좌이체	19	이동	25	정보	20	...
이미지	19	은행	19	단말기	25	RF교통	19	...
전화	17	신용	19	관련	25	상기	17	...
한국	17	고객	18	대금	25	청구	14	...
전체 322개 단어		전체 322개 단어		전체 185개 단어		전체 316개 단어		...

하여 각 특허의 주제어들을 추출하였다. 주제어 추출을 위해서 국민대학교에서 개발한 형태소 분석기 프로그램 HAM 7.0을 사용하였다. 주제어 추출은 명사를 대상으로 하였으며 일반적으로 사용되는 명사는 제외하였고, 복합명사는 분리하여 추출하였다. <표 5>는 각 특허별로 추출된 주제어와 그 주제어의 출현빈도를 정리한 표의 일부를 나타내고 있으며, <표 6>은 <표 5>의 결과를 토대로 각 특허별 상위 빈도 3개 주제어로 구성된 216개의 주제어들을 정리한 결과이다. 특허 1의 경우 상위빈도 3개는 ‘데이터’, ‘다운로드’, ‘단말기’이다.

특허군 내의 각 특허에서 얻어진 216개 주제어를 대상으로 표준화(중복단어 제거, 복합명사 분리)를 실시하였고, 최종적으로 84개 주제어를 선정하였다. 전체 주제어 84개와 특허 72개를 이용한 주제어 출현빈도 행렬은 <그림 5>와 같다.

<표 6> 각 특허의 주제어

특허	특허1	특허2	특허3	특허4	...
주제어	데이터	결제	상품	RF	...
	다운로드	사업자	방송	카드	...
	단말기	대금	구매	교통	...

전체 특허 중에서 5%(3개) 이하의 특허에서 출현한 주제어는 제외하여(57개 주제어×72개 특허) 행렬을 작성하였다.

### 4.3 특허군의 군집화

앞에서 작성된 주제어 출현빈도 행렬에 대해 SPSS 11.0을 사용하여 K-means 클러스터링 분석을 실시하였다. 클러스터링은 <표 7>에 정리된 것과 같이 2집단, 3집단, 4집단 등 3가지 경우로 수행



<그림 5> 주제어 출현빈도 행렬

<표 7> 특허 72개의 군집분석 결과

7	2집단 군집	3집단 군집	4집단 군집
군집 1	16개	15개	26개
군집 2	56개	7개	16개
군집 3		50개	27개
군집 4			3개
합계	72개	72개	72개

주) ■ : 평가 대상 특허가 포함된 군집.

하였으며, 이중 3집단 군집이 군집 결과가 가장 좋았다. 3집단 군집에는 15개, 7개, 50개의 특허가 각각의 군집을 형성하였다. 같은 군집에 속해있는 특허들은 유사성을 바탕으로 결합된 동질적인 군집이라 할 수 있다. 따라서 이후 분석에서는 특허들 간의 관련성을 증가시키기 위해 3집단 군집에서 평가대상 특허가 포함된 군집의 50개 특허만을 분석의 대상으로 한정하였다.

#### 4.4 기술적 중심도 평가

특허들 간의 유클리디안 거리를 구한 뒤 유사도 행렬을 작성한 결과는 <그림 6>과 같다.

각 특허에 대한 기술적 중요도를 평가한 결과는 <표 8>과 같다. <표 8>에는 각 특허에 대한 기술적 중요도와 기술적 중요도의 순위를 전체 특허를 고려한 경우와 후행 특허만 고려한 경우를 정리한 것이다. 후행 특허만 고려한 경우의 기술적 중요도는 해당 특허보다 출원일이 늦은 특허만을 대상으로 하여 기술적 중요도 값을 계산한 것이다. 즉, 해당 특허보다 출원일이 늦은 특허와의 유사도 값을 평균한 것이다.

본 연구의 평가대상 특허인 p35(Patent No.35)의 경우 전체 특허를 고려한 경우 기술적 중요도는 0.4484로 전체 50개 특허 중 5번째로 큰 값을 가졌다. 그러나 후행 특허만을 고려한 경우는 그

	p01	p02	p05	p06	p07	p08	p09	p10	p12	p17	p18	p20	p22
p01	1.00	0.29	0.47	0.27	0.28	0.35	0.57	0.21	0.59	0.44	0.50	0.35	
p02	0.29	1.00	0.39	0.15	0.27	0.44	0.31	0.46	0.37	0.41	0.36	0.13	0.27
p05	0.47	0.39	1.00	0.34	0.46	0.41	0.36	0.31	0.41	0.41	0.39	0.17	0.41
p06	0.27	0.15	0.34	1.00	0.39	0.34	0.33	0.33	0.39	0.36	0.31	0.41	0.40
p07	0.28	0.27	0.46	0.39	1.00	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
p08	0.35	0.44	0.41	0.33	0.33	1.00	0.33	0.33	0.33	0.33	0.33	0.33	0.33
p09	0.57	0.41	0.36	0.33	0.33	0.33	1.00	0.33	0.33	0.33	0.33	0.33	0.33
p10	0.21	0.13	0.17	0.07	0.07	0.17	0.17	1.00	0.33	0.33	0.33	0.33	0.33
p12	0.59	0.46	0.57	0.35	0.35	0.35	0.35	0.33	1.00	0.33	0.33	0.33	0.33
p17	0.44	0.37	0.62	0.31	0.20	0.31	0.52	0.17	0.55	1.00	0.50	0.50	0.38
p18	0.50	0.41	0.49	0.33	0.34	0.33	0.29	0.81	0.52	0.66	1.00	0.48	0.48
p20	0.35	0.36	0.43	0.36	0.43	0.36	0.43	0.36	0.43	0.36	0.43	1.00	0.48
p22	0.41	0.31	0.59	0.28	0.19	0.22	0.44	0.09	0.41	0.38	0.48	0.48	1.00

<그림 6> 특허들 간의 유사도 행렬

<표 8> 기술적 중요도 평가 결과

특허 No.	전체 특허를 고려한 경우		후행 특허만 고려한 경우	
	기술적 중요도	순위	기술적 중요도	순위
p01	0.1525	35	0.2312	24
p02	0.3021	19	0.2592	21
...	...	...	...	...
p25	0.3124	16	0.4112	4
...	...	...	...	...
p35	0.4484	5	0.1224	34
...	...	...	...	...
p72	0.0812	43	0.1252	32

값이 0.1224로 34번째에 해당하는 작을 값을 가졌다. 이러한 결과로부터 평가대상 특허  $p35$ 의 경우 전체 특허 내에서 기술적으로 중요한 역할을 차지하고 있으나 후행특허에 미치는 영향이 크지 않다고 추측할 수 있다. 반면, 특허  $p25$ (Patent No.25)의 경우 전체 특허를 고려한 경우 기술적 중요도는 0.3124로 기술적 중요도가 크다고 할 수는 없으나 후행 특허만 고려한 경우는 0.4112로 전체 특허 중 4번째로 큰 값을 가져 후행특허에 미치는 영향이 크다고 판단할 수 있다.

일반적인 인용 분석에서는 특정 기술의 중요성이나 영향력을 평가하기 위해서 인용빈도를 사용하고 있으나, 단순히 인용빈도만을 이용하는 경우 출원된 지 오래된 기술일수록 중요성이나 영향력이 과대평가될 수 있으며 반대로 최근에 출원된 기술에 대한 정확한 평가가 어렵다는 문제점을 가진다. 특허출원의 시간적 선후관계에 무관하게 주제의 유사성을 이용하여 기술적 중요도를 측정할 경우, 만약 상대적으로 늦게 출원된 특정한 기술이 선행 특허기술과의 유사성이 매우 높다면 그 특허기술은 매우 중요한 기술로 평가받게 되는데 이는 합리적인 결과로 보기 어렵다. 왜냐하면 특허 기술의 기술적 가치는 그 기술이 후속기술개발의 기반이 되느냐에 달려 있고, 이는 후속기술에 의해 인용된 빈도에 의해 추정될 수 있기 때문이다 (Wartburg, 2005). 따라서 본 연구에서는 평가하고자 하는 특허가 후행특허와 가지는 인접성도 같이 고려하여 기술의 중요성을 평가하는 것이 합리적이라고 판단하였다.

## 5. 결론

본 연구에서는 특허간의 인용정보가 없거나 미비한 상황에서 특허의 기술적 중요도를 평가할 수

있는 방법론을 제안하였다. 제안된 방법론은 평가대상 특허 선정, 관련 특허군 수집, 주제어 출현빈도 행렬 작성, 특허군의 군집화, 그리고 기술적 중요도 평가 등의 5단계를 거쳐 평가대상 특허의 가치를 기술적 중요성 측면에서 객관적이고 정량적으로 평가할 수 있도록 하고 있다.

관련 특허군 선정을 위해서, 우선 평가대상 특허가 갖는 IPC 분류코드의 첫 4자리와 동일한 IPC 분류코드를 갖는 특허들을 수집하였으며, 이중 제목과 초록 부분에 평가대상 특허의 주제어를 포함하는 특허만으로 그 대상으로 한정하였으며, 특허출원일과 특허 상태를 이용해 추가적으로 그 대상을 좁혀갔다. 주제어 출현빈도 행렬 작성은 형태소 분석 프로그램을 이용해 각 특허별로 주제어 출현빈도를 산출하였다. 특허군의 군집화는 주제어 출현빈도에 대하여 K-means 클러스터링 기법을 적용하여 특허들을 몇 개의 군집으로 구분하였는데, 이는 모든 특허들을 대상으로 하기보다 평가대상 특허가 포함된 군집에 포함된 특허만을 대상으로 이후의 분석을 수행함으로써 보다 의미 있는 결과를 얻기 위함이다. 마지막으로 각 특허의 기술적 중요도를 계산하기 위해 유클리디안 거리를 이용한 특허 간 유사도 행렬을 구하고 이로부터 기술적 중요도를 계산하였다. 또한, 본 연구에서는 시간적 순서를 고려하여 특허의 기술적 중요도를 평가할 수 있도록 설계되어 있다. 본 연구는 ‘휴대단말기를 이용한 대금결제를 위한 사용자 인증방법’을 평가대상 특허로 선정하여 제안된 방법론의 적용 과정을 단계적으로 설명하고 그 적용 가능성을 확인하였다.

본 논문의 한계점은 다음과 같다. 첫째, 특허 기술의 서지정보를 이용하여 인용관계를 파악하는 방법에 대한 비판이 있을 수 있다. 둘째, 기술적 중요도를 도출하는 대상 기술군의 구성에 따라 동일

한 기술의 기술적 중요도가 달라질 수 있어 척도의 신뢰성을 확보하기 어렵다. 셋째, 기술의 중요도를 분석하기 위한 양적방법론으로서의 자료수집과 분석에 지나치게 많은 자원투입이 요구된다는 점이다.

미국, 유럽 특허와는 달리 국내 특허에서는 선출원된 특허기술의 청구항목에 대한 권리를 인정하는 수단으로서 인용정보를 요구하지 않고 있다. 따라서 본 논문에서는 국내 특허에 대해서 특허서지정보를 이용하여 기술적 연관성을 도출하고 있다. 그러므로 본 논문에서의 방법론은 특허의 권리관계를 기반으로 하는 인용관계보다는 기술적 유사성의 정도를 측정하여 이를 기술적 영향력 관계로 파악하는 것이라고 볼 수 있다. 물론 개별 특허기술의 차원에서 보면 특정 특허기술과의 유사성이 곧 영향관계를 의미한다고 보기는 어려우나, 군집화된 유사한 기술군 내에서는 기술적인 유사성이 높은 특허기술이 영향력이 높은 기술이 될 가능성이 높다고 판단하였다.

위에서 제시된 한계점은 대상기술군의 구성에 따라 기술적 중요도가 달라지는 두 번째 한계점과도 밀접하게 연관되어 있다. 만약 기술적 중요도를 절대적인 척도로 이해한다면 본 논문에서 사용된 방법은 척도의 신뢰성을 확보하기 어렵다. 반면에 상대적인 기술적 중요도를 이용함으로써 기술적 중요도가 제한된 기술군 내에서만 의미를 갖게 되는 한계가 있지만 척도의 신뢰성 문제를 해결할 수 있다. 동시에 기술의 유사성을 기술적 영향력으로 해석하는데 발생할 수 있는 오류를 줄일 수 있다. 왜냐하면 기술군이 동질적일수록 특허기술의 서지정보의 우연한 유사성이 기술적 영향관계로 해석될 확률이 낮아지기 때문이다.

향후 연구과제로, 본 연구에서 제안된 방법론의 철저한 검증은 위해서는 일정 개수 이상의 사례에

대해 추가적인 분석이 요구되며, 기술적 중요도와 기업의 성과(예 : 재무적 성과)와의 연관성에 대한 분석이 보충되어야 할 것이다. 또한 분석 과정의 전산화 및 자동화가 필요하며, 기술적 중요도에 대한 시계열 추이를 포함하는 분석 방법에 대한 연구가 필요하다. 아울러 분석대상 기술군의 크기가 증가하면 데이터 분석량이 기하급수적으로 증가하기 때문에 이를 줄일 수 있는 합리적인 기법의 개발이 요구된다.

## 참고문헌

- 김경재, 김병국, “데이터 마이닝을 이용한 인터넷 쇼핑물 상품추천시스템”, *한국지능정보시스템학회논문지*, 11권 1호(2005), 191~205.
- 이장희, 유성진, 박상찬, “성공적인 e-Business를 위한 인공지능 기법 기반 웹 마이닝”, *한국지능정보시스템학회논문지*, 8권 2호(2002), 159~175.
- Albert, M. B., D. Avery, F. Narin, and P. Mcalister, “Direct Validation of Citation Counts as Indicators of Industrially Important Patents”, *Research Policy*, Vol.20, No.3(1991), 251~259.
- Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel, “Citation frequency and the value of patented innovation”, *The Review of Economics and Statistics*, Vol.81, No.3(1999), 511~515.
- Harhoff, D., F. M. Scherer, and K. Vopel, “Citations, family size, opposition and the value of patent rights”, *Research Policy*, Vol.32, No.8(2003), 1343~1363.
- Hirschey, M. and V. J. Richardson, “Valuation effects of patent quality : A comparison for Japanese and U.S. firms”, *Pacific-Basin Finance Journal*, Vol.9, No.1(2001), 65~82.

- Karki, M. M. S., "Patent Citation Analysis : A Policy Analysis Tool", *World patent Information*, Vol.19, No.4(1997), 269~272.
- Kutznets, S., "Innovation Activity : Problems of definition and measurement, In R. Nelson (ed.)", *The rate and direction of inventive activity*, Princeton, NJ : Princeton University Press, 1962.
- Lanjouw, J. O., "Patent protection in the shadow of infringement : simulation estimations of patent value", *The Review of Economic Studies*, Vol.65, No.4(1998), 671~710.
- Larsen, B., and C. Aone, "Fast and Effective Text Mining Using Linear-Time Document Clustering", in *Proceedings of the Fifth ACM SIGKDD International Conference on KDD*, 1999.
- McMillan, G. S. and R. D. III. Hamilton, "Using Bibliometrics to Measure Firm Knowledge : An Analysis of the US Pharmaceutical Industry", *Technology Analysis and Strategic Management*, Vol.12, No.4(2000), 465~475.
- Meyer, M., "Does science push technology? Patents citing scientific literature", *Research Policy*, Vol.29, No.3(2000), 409~434.
- Mowery, D. C., J. E. Oxley, and B. S. Silverman, "Technological overlap and interfirm cooperation : implications for the resource-based view of the firm", *Research Policy*, Vol.27, No.5(1998), 507~523.
- Narin, F., "Patent bibliometrics", *Scientometrics*, Vol.30, No.1(1994), 147~156.
- OECD, "The measurement of scientific and technological activities : Using as science and technology indicators-Patent manual", Paris, 1994.
- Putnam, J., "The value of international patent rights", Ph. D. Thesis. Yale University, Yale, 1996.
- Schmoch, U., "Tracing the knowledge transfer from science to technology as reflected in patent indicators", *Scientometrics*, Vol.26, No.1(1993), 193~211.
- Scotchmer, S., "Protecting early innovators : should second-generation products be patentable?", *The RAND Journal of Economics*, Vol.22, No.2(1996), 322~331.
- Scotchmer, S., "Standing on the shoulders of giants : cumulative research and the patent law", *Journal of Economic Perspectives*, Vol.5, No.1(1991), 193~211.
- Swann, P., "Inference from mixed bags : The economic value of patent counts, Innovation counts and the lake", London Business School, 1993.
- Tijssen. R. J. W., "Global and domestic utilization of industrial relevant science : Patent citation analysis of science-technology interactions and knowledge flows", *Research Policy*, Vol.30, No.1(2001), 35~54.
- Trajtenberg, M., "A penny for your quotes : patent citations and the value of inventions", *The RAND Journal of Economics*, Vol.21, No.1(1990), 172~187.
- Van der Drift, J., "Statistics of European patents on legal status and granting data", *World Patent Information*, Vol.10, No.4 (1988), 243~249.
- Wartburg, I. V., T. Teichert, and K. Rost, "Inventive progress measured by multi-stage patent citation analysis", *Research Policy*, Vol.34(2005), 1591~1607.
- Yoon, B. and Y. T. Park, "A text-mining based patent network : Analytical tool for high-technology trend", *Journal of High Technology Management Research*, Vol.15, No.1 (2004), 37~50.

Abstract

## A Model for Evaluating Technology Importance of Patents under Incomplete Citation

Heon Kim<sup>\*</sup> · Dong-hyun Baek<sup>\*\*</sup> · Min-ju Shin<sup>\*\*\*</sup> · Dong-seok Han<sup>\*\*\*</sup>

Although domestic research funding organizations require patented technologies as an outcome of financial aids, they have much difficulty in evaluating qualitative value of the patented technology due to lack of systematic methods. Especially, because citation data is not essential to patent application in Korea, it is very difficult to evaluate a patent using the incomplete citation data. This study proposes a method for evaluating technology importance of a patent when there is no or insufficient citation data in patents. The technology importance of a patent can be evaluated objectively and quantitatively by the proposed method which consists of 5 steps such as selection of a target patent, collection of related patents, preparation of key word vector, clustering patents, and technological importance assessment. The method was applied to a patent on 'user identification method for payment using mobile terminal' in order to evaluate technology importance and demonstrate how the method works.

**Key Words** : Patents, Text Mining, Key Word Vector, Clustering, Technology Importance

---

\* Division of Business and Commerce, Baekseok University

\*\* Department of Business Administration, Hanyang University

\*\*\* E-Biz Team, Samyang Data System

## 저 자 소개



**김 현**

연세대학교 경영학과를 졸업하고, 동 대학원에서 경영학 석사 및 박사학위를 취득하였다. LG경제연구소에서 근무하였으며, 현재 백석대학교 경상학부에 재직 중이다. 주요 관심분야는 기술경영, 기술가치평가, 생산전략 등이다.



**백동현**

한양대학교 산업공학과를 졸업하고, 한국과학기술원 산업공학과에서 석사 및 박사학위를 취득하였다. 하이닉스 반도체에서 근무하였으며, 현재 한양대학교 경상대학 경영학부 및 일반대학원 e-business경영학과 부교수로 재직 중이다. 주요 관심분야는 경영정보시스템, 기술가치평가, 시스템공학 등이다.



**신민주**

백석대학교 경상학부를 졸업하고 한양대학교 일반대학원 e-business경영학과에서 석사학위를 취득하였다. 현재 삼양 데이터 시스템 e-Biz 팀에 재직 중이며, 주요 관심분야는 경영정보시스템, 인적오류 분석 등이다.