Technological Forecasting & Social Change 92 (2015) 332–345

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO

Jeeeun Kim, Sungjoo Lee *

Department of Industrial Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon, Gyeonggi-do 443-749, South Korea

ARTICLE INFO

Article history:
Received 6 September 2012
Received in revised form 18 July 2014
Accepted 17 January 2015
Available online 7 February 2015

Keywords:
Patent database
Innovation studies
USPTO
EPO
JPO
KIPO

ABSTRACT

Patents, which are regarded as important sources of innovation, are closely related to innovation strategies, and data from patent registrations are frequently used in innovation studies. Generally, such studies collect patent information from a patent database and analyze it to identify innovation trends and strategies at the firm, industry, and national levels. Therefore, it is important not only to develop and apply a suitable analysis method, but also to select a patent database that is appropriate for such studies' research objectives. But few previous studies on patent analysis have carefully examined the suitability of the databases they have used.

To help in the use of patent databases for innovation studies, this paper investigates the characteristics of four databases: the USPTO, EPO, JPO, and KIPO. For this purpose, all patents registered at each database over the three years 2008–2010 were extracted for cross-sectional analysis, and the number of annual patents registered in two patent classes, A61 and G06, during the last 20 years (1992 to 2011), was collected for longitudinal analysis. The analysis perspectives were set in the context of comparison between databases. After integrating the results of this comparison, this study summarized key features of each patent database in terms of the necessary conditions to be good knowledge sources of innovation studies. The research results are expected to contribute to a better understanding of the characteristics of various patent databases, and to further support the validity of patent analyses.

1. Introduction

New technologies serve as key drivers of firm growth and success in fast-changing competitive markets (Martino, 1993). Therefore, identifying new technology opportunities by monitoring technological changes and analyzing technological innovation patterns becomes an essential part of establishing a successful organizational strategy (Nosella et al., 2008). Among the various methods used for achieving this, patent analysis is the most commonly applied. Generally, 80% of patent documents are made up of technological information (Teichert and Mittermayer, 2002), and technologies that can be commercialized are first published as patents intended 'for industrial use'. In other words, the correct investigation of patent information makes it possible to evaluate a technology's originality, progressiveness, and commercial potential (Kuznets, 1962). Thus, patent analysis has been used for various purposes — both long-term and short-term — such as monitoring technology trends, analyzing technology innovation patterns or developing technology strategies. Patent information can also be stored in well-structured databases (DBs) that give excellent data accessibility because they are open to the public. Therefore, patents are used as objective and mature indicators of innovation, and for quantitative analyses of technology innovations (Chen et al., 2012).

Since it is necessary to select patent DBs in order to implement such studies, the selection significantly affects the study results. The DBs most commonly used in previous studies include those from the United States Patent and Trademarks

* Corresponding author. Tel.: +82 31 219 2419; fax: +82 31 219 1610.
E-mail addresses: mistylake2357@ajou.ac.kr, topdogje@naver.com (J. Kim), sungjoo@ajou.ac.kr (S. Lee).

http://dx.doi.org/10.1016/j.techfore.2015.01.009

Office (USPTO), the European Patent Office (EPO), the Japanese Patent Office (JPO), and so on. The USPTO DB, which contains the most information and is easy to use, is regarded as a representative patent DB (Archibugi and Planta, 1996). The EPO DB, meanwhile, has been regarded as a valuable information source for analyzing the progress of technology innovation in European economies (Abraham and Moitra, 2001), while innovation studies also utilize the JPO DB because the Organization for Economic Co-operation and Development (OECD) has highlighted the importance of patents registered with the JPO under the concept of 'triadic patent families', which refer to the set of patents filed at these three major patent offices (the USPTO, the EPO, and the JPO) (Hicks et al., 2001). However, it is important to reiterate that analysis results (and implications drawn from them) will differ depending on which DB is used for analysis. For instance, not all developers file their patents at the USPTO, even though the US is the world's largest technology market.

It can sometimes be a better choice for firms to file at the EPO for various reasons; possibly because the firm's products or technologies are expected to be marketed in Europe rather than in the USA; the size of markets adopting those technologies (along with their supply chains) is larger in Europe than in the USA; or products in technological fields close to those of the patents are mostly manufactured in Europe. Moreover, European small-and medium-sized enterprises (SMEs) which offer end-user product/services only for domestic market may be satisfied with filing in their own country, unless they plan to globalize their business, or they supply their products to domestic arms of global companies. As international applications can be very expensive, a deliberate choice about where to file a patent application is an essential part of a firm's patenting strategy. Thus the innovation patterns derived from analyses based on the USPTO DB and the EPO DB are likely to differ, even when they concern the same technologies: so when adopting a patent DB for an innovation study, it is desirable to consider each of their characteristics fully, and select the one that will suit the study's objectives. However, most previous studies have selected patent DBs based on a vague notion of their characteristics or a subjective idea that a particular patent DB is superior to the others — but there have been few attempts to provide meaningful information for selecting patent DBs for innovation studies.

The object of this study, therefore, is to analyze the characteristics of the major DBs used in technology innovation studies, so as to produce some 'pointers' for the selection of appropriate DBs for such studies. For this purpose, the study was implemented in three stages. First, bibliographic information about all the patents registered during the 2008–2010 period was collected from the three major DBs — USPTO, EPO, and JPO — which are predominantly used to analyze global technology innovations, as well as from the Korean Intellectual Patent Office (KIPO) DB, which has seldom been used to study global technology innovations: it is regarded as representing the innovation activities limited to a specific country. We also collected the annual numbers of patents over the last 20 years (1992–2011) from these four sources in two patent classes — A61 and G06 — which are those in which the most number of patents have been registered during the 2008–2010 period in the four patent DBs. Second, three perspectives were determined for analyzing the characteristics of the four patent DBs, along with detailed indexes for each perspective. Under the first perspective, the *'degree of innovation'* was analyzed by evaluating *'how actively innovation results — i.e., that is, patents — are registered with each DB'*. Under the second perspective — *'innovation participants'* — the study analyzed *'how many different innovation participants are involved in registering innovation results with each DB, and who are the major assignees'*. Under the third — *'innovation targets'* — the study examined *'how many innovation results are registered in each DB through in many different technologies, and which of these are the major technologies'*. Finally, using the results of these analyses, the characteristics of these four patent DBs were integrated in order to derive insights. Based on the insights, this study discusses the strength and weakness of each patent DB in the context of innovation studies. The research results are expected to verify the characteristics of the patent DBs that are most frequently mentioned in existing innovation studies, and so help clarify the selection of patent DBs for future patent-based studies.

The remainder of this paper is structured as follows. Section 2 examines the theoretical background of this research, based on existing technology innovation studies that have used patent information. Section 3 presents this study's overall research framework, along with detailed indexes for analyzing the characteristics of patent DBs on the basis of this framework. The results of the proposed indexes are derived in Section 4, and important implications are discussed in Section 5. Section 6 describes this study's contributions and limitations, and offers some conclusions.

2. Theoretical background

2.1. Patents as proxies of innovation activities

In previous innovation studies, innovation activities have been measured from various perspectives, largely grouped into two types — of input (e.g. R&D capital stock, R&D staff) and of output (e.g. new products, patents), and using a wide range of measurements (Cruz-Cázares et al., 2013). These studies have been especially interested in the relationships between innovation and firm performance, some emphasizing the short-term direct effect of innovation inputs on firm performance (Georg et al., 2002), and others the long-term indirect effect through the innovations achieved (Balkin et al., 2000).

Earlier studies tended to take the input perspective, and many have used R&D expenditure as the principal measure of innovation (O'Regan et al., 2006), although other different types of innovation inputs have been used, such as R&D intensity (Hitt et al., 1997) and R&D manpower (Wang and Huang, 2007). However, it has been argued that innovation inputs cannot directly measure innovation activity (OECD/Eurostat, 1997). For example, R&D expenditure cannot encompass all corporate innovation efforts, such as learning by doing, acquisition of new machinery, or investment in human capital (Hashi and Stojčić, 2013).

On the other hand, innovation can manifest itself as innovation output, so a variety of such outputs, such as product innovations (Li, 2000), process innovations (Akgüne et al., 2009) and patents (Zahra and Nielsen, 2002) have been suggested as measures of innovation activity. Among them, the most commonly employed measures include the proportion of sales generated by new products, new product announcements, and numbers of patents (Acs and Audretsch, 1987; Klomp and Van Leeuwen, 2001; Loof and Heshmati, 2002). The proportion of sales attributable to new products and the frequency of new product announcements are useful in that they connect the innovation activities directly to their technological and commercial success (Hashi and Stojčić, 2013). However, in spite of their usefulness, by sticking to these measures it is likely that process innovations are neglected (Kemp et al., 2003); moreover, collecting the relevant data is not easy. For instance, new product announcements may be made in various sources and in different countries, so the selection of the relevant sources in which new products are announced should be carefully scrutinized.

Another commonly used indicator is numbers of patents. A patent, itself, cannot guarantee the commercial success of innovation efforts, but rather presents an intermediate measure of innovation output (Kemp et al., 2003). Nevertheless, innovation is defined as all scientific, technological, organizational, financial, and commercial activities that lead to, or are intended to lead to, the implementation of technologically new or improved products or services (OECD/Eurostat, 1997), and if we restrict our focus to 'technological innovation', patents are invaluable sources that can record firms' scientific and technological activities. Patent data are attractive for innovation research especially for the following reasons. First, they are rich and full, as patent documents are a source of well-grounded and rich descriptions of technology, and such data have been collected for dozens of years around the world. In addition, they have multi-facet information about 'potential innovations', such as inventors, the point an invention was made, possible industrial applications, and technological novelty compared to prior solutions. Second, the large volumes of patent data are maintained in structured and standardized formats, and so are readily available for analysis: so, naturally, bibliometric analysis of patent data has gained much research interest over recent decades. Third, patent data are easy to access — they are (and need to be) open to the public and are up-to-date. Consequently, they have been widely utilized to investigate innovation activities and, more specifically, technology advances.

At the same time, patent data suffer from several shortcomings when used as measures of innovation activity. First, not all inventions are patented, and the propensity of patenting is sensitive to firm and industry specific characteristics, as well as to changes in legislation (Arundel and Kabla, 1998; Bouwer and Kleinknecht, 1999). SMEs may not patent their innovations due to their limited budgets: such firms may prefer to use trade secret mechanisms rather than patents to protect their technologies. In a similar vein, process innovation can be protected effectively by trade secrets or know-how. Thus, in an industry where process innovation is more important than product innovation, patenting may not be the most appropriate form of gaining competitive advantage. Second, whether an invention can be patented or not may be subjected to patent examiners. And so, not all patents have the same value. Third, patent data are dispersed, as each country has established its own patent offices — so therefore, the selection of the appropriate patent DBs for analysis is critical to obtain reliable results. However, the amount and accessibility of patent data generally offset these drawbacks, so they have been commonly used to measure innovation activities, particularly in terms of technological knowledge generation, at technology, sector, and national levels.

2.2. Innovation studies using patent data

Innovation studies using patent data can be classified into three different categories on the basis of nation, industry and technology, and firm (Hinze and Grupp, 1992). Research conducted at the national level can again be sub-divided into three categories. Studies in the first category attempt to use patent data to identify the nation-specific innovation capacity or innovation patterns (Jacobsson and Philipson, 1996; Ma and Lee, 2008; Lee et al., 2007; Breschi and Catalini, 2010): in other words, they determine trends in changing innovation activities in order to analyze national innovation systems (Wang et al., 2010), or a nation's innovation capacities and specialized technologies. These studies generally use the bibliometric data in patents to measure the degree of industrial development in a specific nation (Ehrnberg and Jacobsson, 1997; Lee et al., 2008), either to determine the most technologically innovative industries in specific nations (Breschi and Catalini, 2010) or to perform comparative analyses of technology innovation trends between nations (Storto, 2006; Schwartz et al., 2012). The studies in the second category analyze the patterns of relationships between nations during innovation processes. The major data source for these studies is patent information about assignees, developers, and citations. Studies belonging to the third category propose methodologies to support decision-making for establishing national innovation policies (Ernst, 1995). Various patent-analysis methods have been developed for national level technology planning (Li, 2009).

At the industrial and technological level, there are, again, three general research streams. First, there are a group of studies that use patent citation data to analyze knowledge networks, such as those between 'innovation participants' in a particular industry (or technology), or networks between industries (or technologies) (Hung and Tang, 2008; Subramanian et al., 2011). Representative studies focus on analyzing networks between scientists and technicians, who are the innovation participants (Hung and Tang, 2008), or knowledge flows between technologically developed and under-developed countries (Godinho and Ferreira, 2012) in the context of particular industries (or technologies). The second group of studies focuses on monitoring technologies, with the objective of either to verify the discontinuous technological changes using the bibliometric data in patents or to derive emerging patterns of such changes (Paci et al., 1997). Recent efforts have applied text-mining methods to patents' descriptive data, enabling information that is particularly valuable for monitoring technologies to be extracted (Li, 2009). Studies in the third research stream are aimed at supporting innovation policies in specific industries (or technologies). In these studies, the patent analysis is combined with other technology management techniques (such as technology roadmaps and TRIZ), either to analyze industrial and technological trends (Li, 2009; Subramanian et al., 2011; Goto and Motohashi, 2007), or to identify the factors that lead to successful industrial and technological innovations (Hicks et al., 2001; Hinze and Grupp, 1992; Janodia et al., 2009).

Firm-level research can be further classified into two categories. Studies in the first category investigate the relationships between a firm's innovation activities and its performance.

They focus on the differences in the performance of firms with and without active innovation strategies, using patent data as a proxy measure of those activities (Abraham and Moitra, 2001), to determine which technological factors facilitate successful performance (Wu and Mathews, 2012). Those in the second category adopt the patent-analysis method to support firms' innovation policies, and generally use patent information to propose patent maps to support firm decision-making, and to analyze their innovation capacities or the success of their innovation policies (Hicks et al., 2001; Hinze and Grupp, 1992; Janodia et al., 2009).

A review of the previous studies on technology innovation using patent data reveals the existence of numerous studies on different topics across various fields. These studies appear to derive one common conclusion and suggestion — the need to collect patent information and then to analyze it. However, most studies have either failed to consider the logical issues regarding the selection of specific DBs seriously, or have merely neglected this aspect. However, since the use of different DBs can lead to totally different analysis results, it is absolutely necessary to establish a process for verifying DBs' characteristics, and then selecting the most appropriate according to the study's objectives. In addition, it is necessary for a study to provide a logical basis for selecting a particular DB as its data source.

3. Methodology

3.1. Research framework

Fig. 1 presents the overall research process. First, the relevant patent data for analysis were obtained. We collected all the patents registered with the USPTO, EPO, JPO, and KIPO between Jan. 1, 2008 and Dec. 31, 2010 for cross-sectional analysis, and also searched the number of annual patents in each DB for the last 20 years, from 1992 to 2011, regarding only two selected patent classes for longitudinal analysis. The patent data collection was performed using WIPS (www2.wips.co.kr), an online patent retrieval system which offers a full database of patents, thus eliminating the need to visit the individual patent and trademark offices' homepages to access their patent DBs. Second, the collected data were analyzed to answer the following research questions, which are essential to consider in selecting a patent DB for innovation studies:

- Does the DB have information enough to show various aspects of innovation patterns now and to suggest future trends?
- Is the information provided by suppliers in different regions enough to show global patterns of innovation? Is there any dominant nation as a key supplier?
- Does the DB cover a range of technological fields so that it can be used for the analysis of innovation patterns in any technology? Is there any dominant technology in the DB?
- Can different patterns of innovation be observed in different patent DBs over time, for both the short- and long-term periods?

In order to answer for these questions, we derived three different perspectives (innovation activities, innovation participants, and innovation targets (i.e. main technological fields)) from which to analyze the characteristics of DBs, based on the rich literature review, and defined eight indexes to measure these different perspectives. Following this, after calculating the index values, we performed a cross-sectional analysis of the characteristics of each DB. It was also necessary to determine, from the longitudinal perspective, if a growing trend in a specific technology was similarly represented across the various DBs.

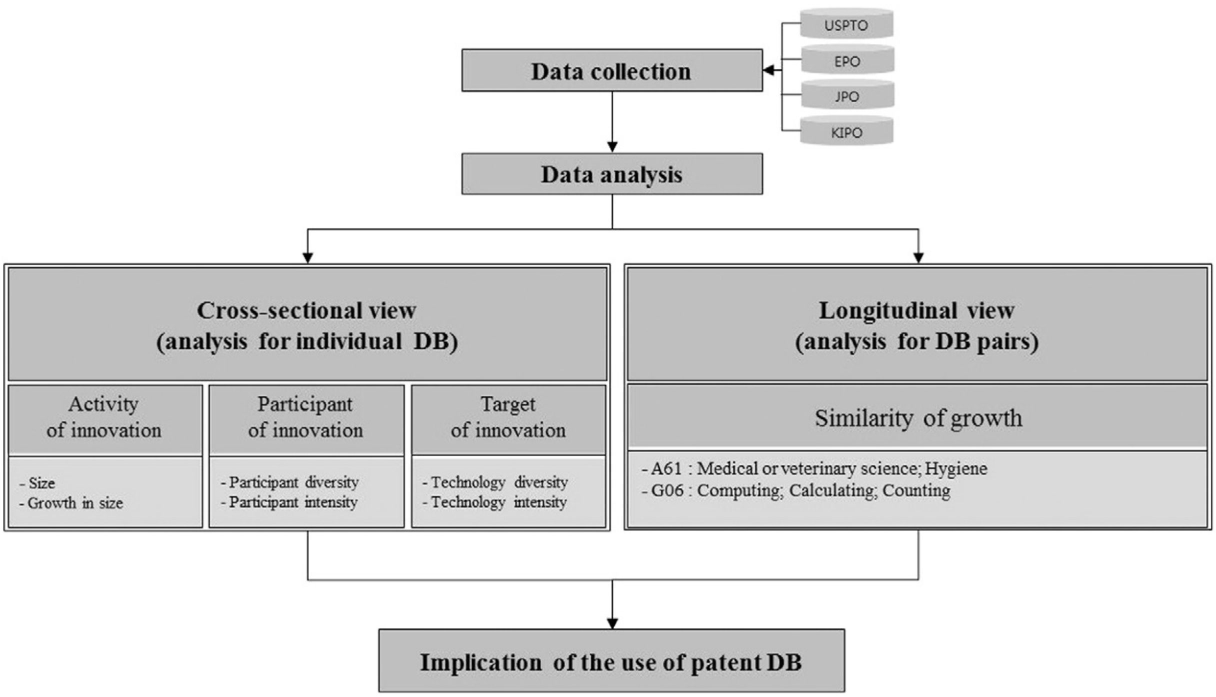

Fig. 1. Overall research framework.

To verify this, we decided to examine the similarities in patent registration patterns and obtained correlations for the numbers of annual patent applications in the different DBs in two technology fields, that is, in two patent classes. Finally, after integrating the results of all these analyses, we could summarize the strength and weakness of each patent DB as a knowledge source of innovation studies. The following sections contain detailed explanations of the analysis methods.

3.2. Use of patent databases for innovation studies

As noted above, the major international sources of patent information for technology innovation studies are the USPTO, EPO, and JPO DBs. To determine how frequently each DB was sourced, we examined the top journals — Research policy, Technovation, R&D Management, Technology Forecasting and Social Change — from among 50 technology innovation journals (Lee and Yoon, 2012) using the key words 'patent analysis'. This search was performed on May 1, 2013, through each journal's online services. Thereafter, we restricted our focus to these journals and performed a simple analysis to determine which patent DBs were used in their papers. The analysis results are summarized in Table 1. We excluded from our analysis those papers whose full manuscripts were inaccessible.

In line with our expectations, a substantial number of papers in the search had used the USPTO DB, which is known to contain a huge amount of technological information and is regarded as a representative DB for patent analysis (Archibugi and Planta, 1996). Many firms, in particular, tend to file patents in the USPTO, especially when they involve drastic improvements to existing products or processes, to claim the exclusive rights to use the innovation in the US, which is the world's largest commercial and technology market and manufacturing economy. This further increases international firms' patent application at the USPTO, which makes its DB exceptional in terms of both quality and quantity of information. In addition, the USPTO DB provides citation information, which supports various patent analyses, and so is most commonly regarded as an appropriate data source for innovation studies (Goto and Motohashi, 2007).

Table 1
The number of papers using each patent DB for innovation studies.

DB	Journal			
(keywords: "patent analysis")	Research policy	Technovation	R&D Management	Technological foresight and social science
Total	62	48	28	112
USPTO	11	12	3	31
EPO	3	3	0	3
JPO	1	0	0	1
KIPO	0	1	0	0
Others[a]	16	9	5	10
Duplicated[b]	9	5	0	6
Unrelated[c]	22	18	16	60

[a] Others: Innovation studies using other DBs apart from the USPTO, EPO, JPO, and KIPO DBs.
[b] Duplicated: Innovation studies using more than two DBs from among the USPTO, EPO, JPO, and KIPO DBs.
[c] Unrelated: Although the keyword "patent analysis" has been presented in a particular paper, it is simply mentioned as a term in the paper or the study objective is not to perform a patent analysis but to describe it theoretically.

The second most popular DB is the EPO DB: while it contains less data than the USPTO DB, it contains rich information regarding technological innovations mostly in and related to European countries (Abraham and Moitra, 2001). Particularly, if we regard the European market as an integrated market, in terms of its scale, rather than as a single market, analyzing the EPO DB would be significant for investigating global technology trends.

Although innovation studies do not commonly use the JPO DB (possibly due to its data accessibility drawbacks), it is also an important DB. OECD regards the patents filed in the JPO, USPTO, and EPO DBs as belonging to the 'triadic patent families', indicating that these three major regions lead the world's patenting activities. The Japanese market represents a high level of technology and has played a major role in providing parts and materials in the global value chain, so patents registered with the JPO have, in fact, played an important role in evaluating technology levels or identifying core patents (Ernst, 1995).

Finally, innovation studies may use several other DBs, including KIPO — we will call them as local DB, because we expect them to have the region-specific rather than globally common characteristics compared to the three major DBs — apart from the three major DBs of the USPTO, EPO and JPO DBs (as shown in Table 1). Although it is not possible to perform a global analysis to arrive at a global view of trends using such local DBs, they can enable the analysis of technological innovation patterns in different regions. Among various local DBs, we selected the KIPO for this research because Korea is one of the places where patenting activities are most active, and KIPO's patent data are easily accessible online, and therefore concluded that its patent DB was worth analyzing. In fact, the IMD (International Institute for Management Development) World Competitiveness Rankings ranked Korea as second in patent applications per capita, fourth in overall patent applications, fourth in patents granted to residents, and sixth in number of patents in force, while the Global Competitiveness Report 2012–2013 published by the WEF (World Economic Forum) ranked Korea ninth in numbers of PCT (Patent Cooperation Treaty) patents, followed by Sweden, Switzerland, Finland, Israel, Japan, Denmark, Germany, and the Netherlands. Consequently, we concluded that analysis of the local KIPO patent DB, as well as the major patent DBs — USPTO, EPO and JPO DBs — was justified.

This study therefore performed a comparative analysis of the characteristics of these four patent DBs in order to arrive at a clear understanding of their differences and similarities. As the USPTO DB has been used in many studies and is widely regarded as a highly reliable data source, it was essential to verify the USPTO's representativeness. It would also be valuable to verify the characteristics of the DBs of other two members of the OECD's triadic patent family — the EPO and JPO. The KIPO DB represents regional trends, and the difference between the KIPO and these major DBs can help indicate whether or not a local DB (such as the KIPO DB) can be used to investigate global trends in innovation.

3.3. Focus of the cross-sectional analysis

Patent statistics have long been used as important indicators for monitoring technological changes (Thongpapanl, 2012). Researchers have designed a number of indexes based on diverse patent information in order to gain significant implications of technological innovations (Hirschey et al., 2001). However, to ensure the validity of these research findings, it is crucial for a patent DB to meet several conditions before it is used for analysis. For example, the higher the number of patents that are registered every year, and the more plentiful is the patent information on various nations and industrial fields, the greater the value a DB will have as a source of technological information to analyze global patterns of innovation.

This study introduces three views of the general innovation process, conceptualizing it as one in which 'innovation activities' are performed by 'innovation participants' to achieve 'innovation targets', which result in the granting of a patent. Thus, the major objects to be analyzed through patent information are the innovation activities, innovation participants, and innovation targets. We can analyze the characteristics of patent DBs from these three perspectives and measure the characteristics using eight indexes, as shown in Table 2.

First, the perspective of innovation activities involves determining which DB contains the largest number of innovation outputs, in other word, which patent DB can provide the largest amount of information. This perspective involves measuring the number of patent registrations and the annual growth in those numbers for each DB. A large number of patent registrations in a particular patent DB indicate that the DB contains substantial information about potential innovations. Although every country has its own patent laws, the basic principles for granting patent rights are similar: the 'novelty' of an invention with respect to the state of the art, its 'inventive-step' and its 'industrial applicability'. The novelty condition requires that an invention should not be in public in any way, anywhere in the world, before the patent application is lodged. The inventive-step condition is satisfied when the invention is not obvious to someone with a good knowledge of the relevant subject, while the final condition concerns the capability of the claims to being made or used in or by an industry. Thus, the degree of patent registration can represent the degree of technological innovation, which supports the argument that the more patents are registered in a DB, the more information regarding technological innovation activities the DB captures. Regarding the amount of information — as both the absolute and relative values are important — this study measured the absolute numbers of patents registered in each DB (*size*) and the rate of increase or decrease in this number (*the growth in size*) between 2008 and 2009 and 2009 and 2010.

The second perspective — innovation participants — is used to verify how many different countries registered patents in each DB: we used information about the nationality of patent assignees, and analyzed the diversity and intensity of their different countries. We calculated the number of countries that have ever applied to each patent office for one or more patents during the two years 2008–2010. The fact that a particular patent office registers patents to assignees from various countries means that the office has greater value in diversity and its DB is significant in indicating global innovation trends. The analysis of participant intensity, on the other hand, is performed to investigate whether the patent application activities are limited to a few specific countries or are performed generally across various different countries. It is common that the patents registered in each country show high assignee rates for the country's own national and domestic firms. However, if this rate is very high, it lowers the value of the country's patent DB as a way of analyzing global innovation patterns, though it can still be used to analyze regional innovation patterns. In analyzing participant intensity, we used the concepts of the five-firm concentration ratio (5CR) and the Herfindahl–Hirschman index (HHI). The 5CR is the market share of the five largest firms in the industry — high values means the industry is strongly concentrated and controlled by only a few firms (Wu and Mathews, 2012). In this study, the 5CR was defined as the share of patents filed by the top five countries, based on the nationalities of patent assignees in a specific DB. The HHI is also frequently used to represent market concentration, and is calculated as the sum of squaring the market shares of the participating firms (Comanor and Scherer, 1969). In this study, we modified the original HHI equation (as noted in Table 1) to analyze the concentration of the assignees' countries. Again higher values indicate greater concentration — so we can

Table 2
Perspectives of analysis and the detailed contents.

Perspectives	Indicators		Operational definition
Innovation activities (Application dates)	Degree of activities Growth rate of activities		The number of issued patents (Size of activities) Average growth rates of patents issued from 2008 to 2009 and from 2009 to 2010
Innovation participants (Nationality of applicants)	Participant diversity		The number of applicants' countries (Size of actors)
	Participant intensity	5CR	Concentration ratio for top 5 countries of applicants
		HHI	Concentration ratio for nationality of applicants $\text{HHI} = \sum_{i=1}^{n}\left[\frac{TEU_i}{\sum_{i=1}^{n} TEU_i}\right]^2$ *HHI = concentration ratio for nationality of applicants *TEUi = total number of patents issued in a patent DB *n = total number of different countries of applicants
Innovation targets (International patent classification)	Technology diversity		The number of IPCs (Size of IPC)
	Technology intensity	5CR	Concentration ratio for top 5 sections of IPC
		HHI	Concentration ratio for IPC $\text{HHI} = \sum_{i=1}^{n}\left[\frac{TEU_i}{\sum_{i=1}^{n} TEU_i}\right]^2$ *HHI = concentration ratio for size of IPC codes *TEUi = total number of patents issued in a patent DB *n = total number of different IPC codes

understand lower values of our two adapted ratios as indicating lower concentration, and this greater likelihood that a patent DB will be suitable for global innovation studies.

Finally, the perspective of innovation targets is proposed to verify the number of different technologies that are registered in each country using class level (i.e., two-digit), International Patent Classifications (IPC) by analyzing each country's technology diversity and intensity. The reason we adopted two-digit IPCs for this analysis are two-fold. First, the pilot analysis indicated that patents are classified into hundreds of technological fields when two-digit IPCs are used, which we judged as an appropriate number to use to define technologies for our analysis. If we use lower levels classification to define technology (e.g., three or four-digit IPCs), the size (i.e., numbers) of patents in each technological field may vary greatly, which could distort the research results. On the contrary, if we use a higher level classification (one-digit IPCs), the research results will become too general to yield meaningful implications. Secondly and more importantly, most previous research on patent analysis for innovation studies has used class-level (two-digit) units to define technology: since our research purposes to support the selection of patent DBs for innovation studies, we adopted the same classification level as used in the majority of previous studies (Lee et al., 2010; Geum et al., 2013).

Technology diversity is measured based on the diversity of the IPC (i.e., the diversity of the innovation targets), that relate to the patents filed during 2008–2010. Thus, we can calculate the number of IPC codes to which at least one patent among those registered in each patent office during 2008–2010 belongs. A large number of IPCs indicates that the patents registered in that patent office cover various technology fields, and so present plenty of information for innovation studies. The technology intensity analysis — like the participant intensity analysis — also uses the values of 5CR and HHI, but the subjects of the analysis change from assignee countries to technologies (IPCs), as shown in Table 2. The technology fields in which innovations occur may differ between countries because of their different market and competitive advantages conditions — and so the technology fields in which more (or less) patents are registered may also vary by country. Here, the higher the values of 5CR and HHI, the more the registered patents are limited to a smaller number of specific technology fields. Moreover, these values, can significantly reflect the regional characteristics of different countries.

3.4. Focus of the longitudinal analysis

While the analysis in the previous section takes a cross-sectional view of patent DB characteristics during 2008–2010, the analysis in this section employs a longitudinal view. Patent information is usually used for time-series analyses. In particular, the 'numbers of registered patents by year' constitute some of the most frequently used variables in innovation studies, as proxy measures for the degree of innovation at national, industrial, and firm levels.

This research aims to examine whether the degree of innovation at the technology level yields similar results in all patent DBs (or at least in those used to analyze global trends). Thus, we first selected the top two two-digit IPCs from the analysis results referred to in Section 3.3, and collected the numbers of annual patents in each of those IPCs for each patent DB, for the last 10 and 20 years, respectively, for the reference year 2011. We then performed a correlation analysis of the annual numbers of patents for each DB pair and for each IPC. There are two reasons why two among the top five IPCs showing high patent shares were adopted for the analysis. First, IPCs are assigned to a patent after its contents have been carefully examined by a patent examiner at the time of the application, so they can be taken as a reliable technology classification system — so it is reasonable to use IPCs to represent technology fields. The IPC system — which is a consistent and internationally recognized technology classification system — is, of course, regularly upgraded to enhance the retrieval, recognition, and delivery of patent information as technologies evolve — but compared to other regional patent classification systems, it is a stable system. Second, an IPC with a high patent share can be regarded as a technology field that experiences active innovations, regardless of countries, so there is a relatively low possibility of presenting biases due to regional characteristics. If the analysis results reveal relatively high correlations between patent DBs across all two IPCs, any of them can be used to investigate global innovation trends (or, at least, to analyze the growth patterns of the major technology fields). But if those correlations are weak, the patent DBs should be carefully selected to guarantee the validity of research results based on its data.

4. Results

4.1. Innovation activities

Fig. 2 shows the results of the analysis of innovation activities, showing that the USPTO has been the most active in registering patents during the 2008–2010 period, followed by the JPO. Both countries have significant numbers of leading global companies, and large markets in both B2B and B2C sectors. Moreover, Japan has strengths in parts and materials technologies, and numerous important suppliers along global production supply chains, so that technology protection has become essential for its firms to gain and retain core competencies and competitive advantages in these increasingly global markets. Interestingly, the number of patent registrations in the KIPO exceeds that in the EPO, although the KIPO DB, as expected, has not been used as frequently for innovation studies as the EPO DB. Though the Korean market size is comparatively small, companies decide to file patents in Korea not only because of its market, but also to gain protection along its firm's manufacturing supply chains. Possibly as Korea is one of key suppliers to global ICT industries, the number of registrations in the KIPO in this sector exceeds that in the EPO. The level of active patent registration in Korea reveals that patents are regarded as an important technology protection tool in that country, as well as a weapon for technology competition in the global market. In the case of EPO, while firms can register their patents with the EPO, they can also register them with patent offices in each individual country. When applying patents to countries adopting a non-substantive examination system, such as France, Italy, Spain and Switzerland or only few of European countries, it may be quick and easy to register a patent directed to patent offices in each individual country. So the numbers of patents registered with the EPO may be less than expected.

Regarding the growth rate of patent registrations, the USPTO and JPO DBs show a continuous increase in the numbers

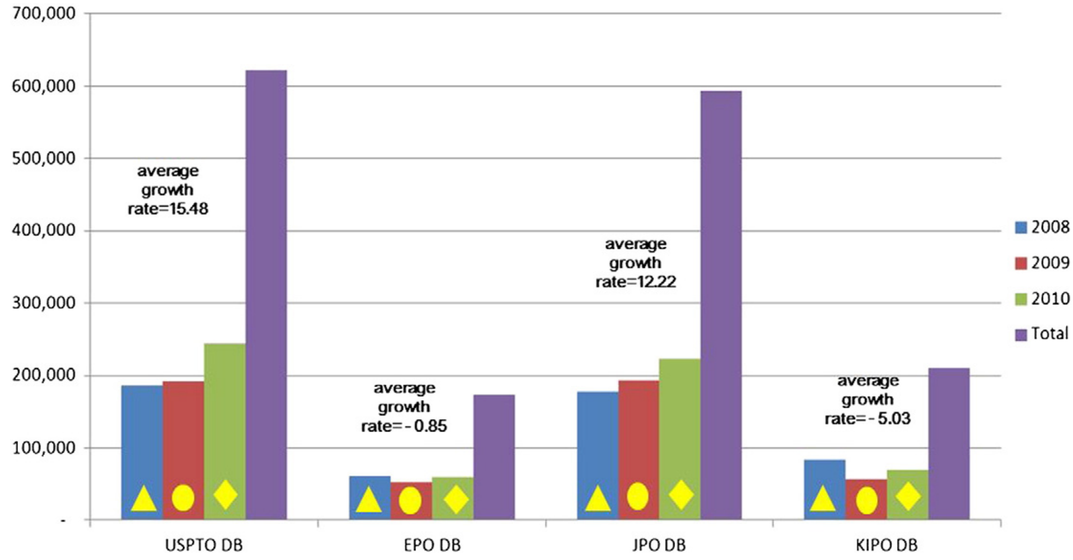

Fig. 2. The number of patents in the four databases.

of patents registered over the 2008–2010 period, whereas the EPO and KIPO DB numbers decline temporarily between 2008 and 2009. As the US and Japan are two of the most technologically advanced countries in the world, companies will certainly endeavor to protect their own technologies and obtain exclusive rights to use them, while preparing for infringement risks and gain advantageous positions in their relationships with their partners by patenting on those countries. The analysis also indicates some fluctuations in patenting activities in the EPO and KIPO.

Here, it should be noted that some of the fluctuations might be attributed to the economic condition. Under the poor economic background, patent applications are reduced with the decrease in R&D investment and deterioration of financial market. Actually, when investigating the top 100 companies in the PCT applications, their growth has slowed down during 2007 and 2008 and most of those in automobile, construction, consumer goods have cut their R&D budgets from 2008 to 2009 (WIPO, 2010). Especially, many of these firms are originated from Europe and may have affected the patent numbers in the EPO more than the other patent DBs. Moreover, the economic condition was unstable especially in Europe during the periods. Among main contributors to the EPO DB including Germany, France and the UK, both Germany and France show a continuous decrease in GDP from 2008 to 2010 and the UK shows a decrease in GDP from 2008 to 2009 by 18.4% and then a small increase from 2009 to 2010 by 2.7% according to World Bank. The same patterns are observed in patent application. According to the WIPO, both Germany and France show a decrease in patent registration from 2008 to 2010 by 20% and 8% respectively; only the UK represents a small increase, 4%, in patent registration during the period. Therefore, it can be said that the growth in patents was very small for the past three years in the EPO DB partly due to the poor economic background. The EPO DB was more sensitive to the economic condition than the other DBs.

4.2. Innovation participants

Table 3 presents the analysis results with regard to innovation participants. Since the JPO DB did not provide information regarding the nationality of assignees, the JPO DB was excluded from this analysis.

First, regarding the participant diversity, patent assignees from 128 countries registered patents with the USPTO over the three years 2008–2010, assignees from 95 (121 in the case of EPO (2)) and 78 countries registered patents with the EPO and KIPO, respectively. In the case of the EPO DB, we measured the values on participant diversity and participant intensity in two ways: EPO (1) shows the analysis results when treating the 28 EU countries as a single country, while the EPO (2) results were obtained considering them as individual ones. The analysis results indicate that patent applications from various countries are included in the USPTO and EPO DBs, which are widely used in global innovation studies.

Second, with respect to the 5CR results, regardless of DBs, these were generally over 70%, so it appears that patenting activities can be characterized as a 'highly concentrated' market, i.e., one that is dominated by a few countries. The EPO (2) DB has the lowest value (73.77%), showing a relatively low concentration ratio, while the 5CR values in the KPO exceed 95%, showing that Korean assignees constituted the largest share of domestic patent assignees among the top five DBs. In terms of the major assignee countries, the core countries for the USPTO are the US, Japan, and Germany; those for the EPO are Germany, the US, and Japan (in that order); and those for the KIPO are Korea, Japan, and the US. Thus, regardless of the patent DBs, three countries — the US, Germany, and Japan — appear to have registered the most substantial numbers of foreign patents. Germany, which ranks third and fourth in the USPTO and KIPO DBs, respectively, ranks first in the larger (EPO (2)) version of the EPO DB, while Taiwan, China, and the Netherlands only appear in the top five country rankings in the USPTO, EPO, and KIPO DBs, respectively.

Table 3
The intensity and diversity of applicants.

		USPTO DB	EPO DB		KIPO DB
			EPO (1)	EPO (2)	
Diversity	Numbers of applicants' countries	128 countries	95 countries	121 countries	78 countries
Intensity	5CR (%)	85.13	94.56	73.77	95.71
	Top 5 countries (%)	1. USA (49.04)	1. EU (48.56)	1. Germany (22.05)	1. Korea (73.74)
		2. Japan (21.37)	2. USA (21.54)	2. USA (21.54)	2. Japan (13.07)
		3. Germany (5.42)	3. Japan (18.18)	3. Japan (18.18)	3. USA (6.31)
		4. Korea (5.39)	4. China (4.12)	4. France (7.88)	4. Germany (1.78)
		5. Taiwan (3.92)	5. Korea (2.15)	5. China (4.12)	5. Netherlands (0.81)
	HHI	2950	3180	1410	5650

Regarding the HHI measurement, the EPO (2) DB shows the lowest value, as it represents a cluster of countries rather than a single country. But, again, the KIPO DB exhibits a much higher value than the other two DBs, because the proportion of patents registered by Korean assignees (which rank first in the KIPO DB) is significantly higher than that in the others. Generally in the marketing field, an HHI value higher than 1800 stands for a highly concentrated state, while values between 1000 and 1800 and below 1000 represent less concentrated and non-concentrated states, respectively. Based on these criteria, Table 3 shows that all the DBs (apart from the EPO (2) DB) show highly concentrated states.

4.3. Innovation targets

Table 4 shows the results of the innovation target analysis (the IPC relevant classifications are detailed in Appendix A). First, with respect to technology diversity, the technology fields covered by the EPO, JPO, and KIPO DBs correspond to about 120 IPCs, based on the patents registered with each of them between 2008 and 2010. The USPTO DB, however, covers 258 IPCs — more than twice the number covered by the others — showing that the patents registered with the USPTO capture more information regarding various technology fields than those registered with the other three DBs, enabling the USPTO DB to provide more abundant information to technology innovation studies.

With regard to technology intensity, however, the results differ considerably from our expectations. We had anticipated that the USPTO, which showed the highest diversity level, would have the lowest intensity level. However (contrary to our expectations) the USPTO DB showed the highest technology intensity level, followed by the KIPO, EPO, and JPO DBs (in that order) when measured by both the 5CR and HHI methods. The 5CR value of the USPTO is slightly greater than 50%, representing a fairly concentrated state, while the corresponding values of the EPO and JPO DBs are quite low, at less than 40% and 20%, respectively.

When investigating the individual technology fields with active patent registrations, we find that G06 (Computing; Calculating; Counting) ranks first in the USPTO DB, reflecting the fact that the US figures conspicuously in the computer and digital industries. In the case of the EPO DB, A61 (Medical or veterinary science; Hygiene) ranks first — this class includes hygiene-related, medical, and veterinary sciences as well as subclasses relating to medical devices, facilities, and methodologies for treating patients. Since Europe is the second-largest market for medical systems after the US, it is not surprising that these technology fields show active innovation. H01 (Basic electric elements) and H04 (Electric communication technique) occupy the top ranks in both the JPO and KIPO DBs: these countries both have competitive presences in these technology fields, which include several semiconductor-related technologies, whose development they lead. Consequently, domestic firms in these countries are actively conducting R&D activities, and foreign firms working in these sectors tend to register patents with the JPO and KIPO to make more effective use of the Japanese and Korean markets and to prevent their competitors from using their technologies. With regard to the top five IPCs, apart from B60 (Vehicles in general) in the EPO DB and G02 (Optics) in the KIPO DB, the same IPC codes ranked at different levels in all the four DBs. As these technology fields generally exhibit short technology life cycles, active minor innovations, and relevant technologies that are easy to protect through patents, they show active levels of patent applications, leading to high registration rates across all DBs.

In terms of the HHI, all DBs show values below 1000, indicating low concentration levels, particularly low in the case of the JPO DB. This indicates that the JPO DB registers patents that are evenly distributed across various fields, possibly because

Table 4
The intensity and diversity of IPC codes.

		USPTO DB	EPO DB	JPO DB	KIPO DB
Diversity	Numbers of IPCs	258	121	121	122
Intensity	5CR (%)	58.40	37.90	17.29	42.67
	Top 5 IPCs (%)	G06 (18.42)	A61 (11.49)	H01 (5.17)	H01 (15.37)
		H01 (13.24)	H04 (10.02)	H04 (4.58)	H04 (13.99)
		H04 (11.74)	H01 (5.76)	G06 (2.59)	G06 (4.80)
		A61 (7.80)	B60 (5.48)	A61 (2.55)	A61 (4.62)
		G01 (7.21)	G01 (5.15)	G01 (2.41)	G02 (3.89)
	HHI	860	480	160	640

Japan possesses advanced technologies in many fields. However, in both the USPTO and the KIPO DBs, patent registrations seem to be more concentrated in some promising technology fields.

4.4. Similarities in innovation patterns revealed by comparing DBs

Tables 5 and 6 show the correlation analysis results. Among the IPCs ranking in the top five patent types in the USPTO, EPO, JPO, and KIPO DBs, the most common are A61 (Medical or veterinary science; Hygiene) and G06 (Computing; Calculating; Counting). After obtaining the number of patents registered annually for these two IPCs in each DB, we calculated the Pearson correlation coefficients for the last 10 and 20 years.

Table 5 reveals no significant correlations between the DBs over the last 10 years. The only DB pair that presents significant correlations is the USPTO and JPO DBs in G06. However, when the analysis period is extended to the last 20 years, the growth patterns of these two technology fields exhibit very strong and significant correlations across the four DBs, as Table 6 shows. In the G06 technology field, in particular, all the correlation coefficient values are greater than 0.77 and are significant at the level of 0.01. On the other hand, the growth patterns for the A61 technology, as seen in the USPTO DB, are not significantly correlated with those in the EPO and KIPO DBs. This indicates that the use of different DBs may lead to different results when time-series analyses are carried out to identify innovation patterns in different technologies. Consequently, it should be noted that it is necessary to select carefully the most appropriate DB to analyze the innovation patterns, according to the technology and perspective involved. In specific, studies designed to examine short-term innovation patterns should be more careful in their DB selection than those exploring long-term innovation patterns. However, local DBs may also be used for innovation studies when the scope of the analysis is limited to technologies with global strength in that particular region and long-term innovation patterns are analyzed.

5. Implications and discussion

5.1. The characteristics of patent DBs

Table 7 summarizes the characteristics of the USPTO, EPO, JPO, and KIPO DBs, which reveal several issues that need to be addressed concerning their use for innovation analysis.

First, the USPTO DB can be regarded as a representative DB that exhibits global innovation patterns, and as being generally suitable for implementing innovation studies. The USPTO DB issues substantial numbers of patents, and these numbers have increased continuously, so its DB that provides the most abundant information on technologies, which again is expected to increase in the future. From the innovation participants' perspective, numerous assignees from different countries have registered their patents with the USPTO, and from that of innovation targets, the USPTO DB contains patents related to many different technologies. However, the USPTO DB shows a slightly higher intensity, at both national and technological levels, than the EPO DB. Nevertheless, considering the large number of patents and the diversity of the assignees' countries and technologies in the USPTO, a slightly higher level of intensity — which is caused by active innovation activities conducted by a few leading assignees in a few promising technology fields — does not decrease the representativeness of its DB. In particular, the USPTO DB offers excellent data accessibility and provides citation information that is unavailable in the other DBs (although these characteristics are not explicitly considered in this study), increasing its applicability for use in various innovation studies.

Second, with regard to the characteristics of other DBs, the EPO DB has the fewest issued patents among the four studied here, but our results indicate that it offers a high level of objectivity in terms of diversity and intensity of innovation participants and innovation targets. Although the EPO DB contains less information, it still provides relatively significant information for studying global innovation trends. The JPO DB contains the second largest amount of information, and is distinguished by its rather low level of technology intensity. But we faced difficulties in collecting data pertaining to innovation participants, suggesting that data accessibility is an important factor affecting the selection of this DB for innovation studies. In contrast, even though it issues large numbers of patents, the KIPO DB, shows assignees from fewer countries, and a higher proportion of patents granted to domestic assignees. This factor most differentiates it from the USPTO, EPO, and JPO DBs, which are commonly used to analyze innovation trends from the global perspective.

Third, synthesizing our research results, we find that the USPTO, EPO, and JPO DBs satisfy the conditions for use in innovation studies — in fact, these three patent DBs, which fulfill the OECD's concept of triadic patent families, can be recognized as representative DBs that appropriately exhibit global trends in technology development. The validity and utility of the DBs are also supported by their characteristics as identified in this study, so we can reasonably accept the results of previous studies that have used one of them, implicitly accepting their suitability for conducting global level innovation studies.

Finally, the four DBs show some common findings in the top two technology fields, while they exhibit differences in terms of the "technology fields covered" and the "degree of innovation in a specific technology for a short-term period of time." These differences are based on the fact that these major markets have different characteristics, and that each country has different technologies that provide it with a competitive advantage. Hence, it is essential either to analyze the various aspects of the major DBs simultaneously, or to carefully select a representative DB, especially when attempting to investigate short-term global technology innovation patterns, or in deriving core patents in specific technology fields. Table 8 summarizes the findings from our cross-sectional and longitudinal analysis results, and discusses each DB as a data source of innovation studies.

Table 5
Results of the correlation analysis for the last 10 years (from 2002 to 2011).

	A61				G06			
	USPTO	EPO	JPO	KIPO	USPTO	EPO	JPO	KIPO
USPTO	1.00	−0.37	0.09	−0.53	1.00	0.31	0.95**	0.57
EPO		1.00	0.46	0.54		1.00	0.24	0.19
JPO			1.00	0.39			1.00	0.49
KIPO				1.00				1.00

** $p < 0.01$.

Table 6
Results of the correlation analysis for the last 20 years (from 1992 to 2011).

	A61				G06			
	USPTO	EPO	JPO	KIPO	USPTO	EPO	JPO	KIPO
USPTO	1.00	0.22	0.58**	0.32	1.00	0.80**	0.93**	0.86**
EPO		1.00	0.71**	0.81**		1.00	0.79**	0.77**
JPO			1.00	0.76**			1.00	0.77**
KIPO				1.00				1.00

** $p < 0.01$.

5.2. Possible future trends of patent DBs

It is worth pointing out some possible future trends of patent DBs. In general, the numbers of patents can be expected to grow, given the fierce competition between global firms and the increasingly complex global business eco-systems, so that managing patent data on a regional basis may no longer be effective. Actually, the greatly increasing costs of patenting have prompted some international organizations (such as WIPO) to aim to integrate all the different patent DBs around the world into one system. When these combined DBs become available, patent analysis on global innovation patterns can be expected to become more valid. At the moment, information service providers such as Thomson gain patent data from worldwide DBs and provide them to users in the same format: using these DBs can be an option for global analysis, even though they were originally published in different countries and under different patent policies.

Second, the use of patent data may change in different industries, so the value of patent analysis for innovation studies may also differ across industries. For example, biotechnology has emerged over recent decades as a vital source of new technologies for the pharmaceutical and agricultural chemical industries (Gans and Stern, 2003). Over 55% of all new products approved by the FDA by 1997 have (mostly or partly) been based on discoveries using biotechnologies which corresponds to the huge number of patent applications in the relevant technological field (A61) over recent years. Actually, biotechnology has led to dramatic improvements in human health, and to compelling value proposition for health care and agricultural consumers (Gans and Stern, 2003). In spite of the early and widespread commercial success of biotechnology in the 1970s, the progress of its commercialization has become more debatable as the technology advances, because of possible side effects. Various other factors besides technological success need to be considered in bringing biotechnologies to market, such as regulations, policy, and social norms. As a broad definition of innovation includes successful commercialization, the value of lessons from the innovation patterns revealed in biotechnology-patent data may be limited.

Again, patent analysis is not a recommendable approach to disclosing innovation patterns in military sectors, as few inventions in such sectors are patented. However, with the increasing interest in dual-use technologies and emphasis on patent management, more inventions will be patented in this sector, as in other civil sectors, so that patent analysis may also provide researchers and developers in the military sector with valuable information.

6. Conclusions

The objective of this study was to analyze the major patent DBs used in innovation studies. To this end, we collected the patents registered with the USPTO, EPO, JPO, and KIPO DBs during the three years (2008–2010). Based on the previous studies that used patent DBs, we determined three different perspectives from which to consider patent DBs in innovation studies, and defined eight indexes to analyze these perspectives. The analysis results verified that the USPTO, EPO, and JPO DBs provide valuable information for analyzing technology innovation patterns from the global perspective. Among these DBs, the USPTO showed the best performance in the 'innovation activity (i.e. degree of innovation)' and 'innovation participants' perspectives, and, in the 'innovation targets' perspective, covered many more IPC codes, and patents corresponding to each code, than the other DBs. So we conclude that the USPTO DB is the most appropriate DB for the purpose of innovation studies,

Table 7
Summary of the characteristics of DBs.

Perspectives of analysis			USPTO DB	EPO DB[a]	JPO DB	KIPO DB
Innovation activities	Number of patents (2008–2010)		622,244	173,043	592,987	209,663
	Growth rate of patents	2008–2009 (%)	3.6	−13.2	9.3	−31.7
		2009–2010 (%)	27.3	11.5	15.2	21.6
		Average (%)	15.5	−0.9	12.2	−5.0
Innovation participants	Diversity of nations	Number of nations	128	95[a](121)	–	73
	Intensity of nations	5CR (%)	85.1	94.6[a](77.6)	–	95.7
		HHI	2950	3180[a](1410)	–	5650
Innovation targets	Diversity of IPCs	Number of IPCs	258	121	121	122
	Intensity of IPCs	5CR (%)	58.4	37.9	17.3	42.7
		HHI	860	480	160	640

[a] The results when the 28 countries belonging to the EU are treated as a single country.

Table 8
Patent DBs for innovation studies: discussions based on the cross-sectional and longitudinal analysis results.

Source	Characteristics	For innovation studies
USPTO DB	▪ (degree) It has the most abundant information on technological innovation and the amount of information has increased rapidly. ▪ (participants) The ratio of applications by residents to non-residents is slightly less than 50%, which is higher than that of the EPO, but does not lower the representativeness as a source of innovation studies, considering a number of global companies in the US and the diversity of assignees' nations. ▪ (targets) Patent filings are concentrated on a few dominant technologies compared to EPO and JPO DBs, with most observed in the G06 (computing; calculation; counting, physics) IPC — but the scope of information is wide, covering 256 IPCs. ▪ (patterns) Its long-term correlations of innovation patterns with the other DBs are significant and high for G06 — but not with the EPO and KIPO for A61.	The USPTO DB is a representative DB for innovation studies, considering its amount, variety, and diverse sources of information; is the most suitable for investigating global innovation patterns in spite of its relatively concentrated information in terms of applicants and technological areas.
EPO DB	▪ (degree) Its amount of information is comparatively small and the annual growth rate in the number of patents is almost zero over the study period. ▪ (participants) The ratio of applicants by residents to non-residents is the lowest and the diversity of assignees' nations is high, implying that the information in the DB is provided by worldwide innovators. ▪ (targets) The coverage of technological knowledge is similar to the other DBs (except the USPTO) but its intensity of IPCs is relatively low. ▪ (patterns) Its long-term correlations of innovation patterns with other DBs are generally significant and high, but not with the USPTO DB for A61.	The EPO DB has information provided by global innovators, although its amount its limited; can be used for innovation studies to investigate global trends but with limited data.
JPO DB	▪ (degree) It has the second largest amount of information, which has showed steady and continuous growth in the past three years. ▪ (targets) The coverage of technological knowledge is similar to the other DBs (except the USPTO) but its low intensity of IPCs is distinguishing. ▪ (patterns) Its long-term correlations of innovation patterns with the other DBs are significant for both A61 and G06.	The JPO DB has strength in the amount of information and decentralization of information across different technological areas; can be used for innovation studies especially when focusing on various technologies.
KIPO DB	▪ (degree) Its amount of information is slightly greater than that of the EPO DB, but it has experienced negative growth in its patent numbers over the study period. ▪ (participants) The ratio of applicants by residents to non-residents is very high and the diversity of assignees' nations is relatively low, indicating that this DB represents regional patterns of innovation. ▪ (targets) The coverage of technological knowledge is similar to the other DBs (except USPTO), but this DB has relatively high intensity of IPCs, implying that its information is concentrated in a few technological areas. ▪ (patterns) Its long-term correlations of innovation patterns are high and significant with other global DBs in general (expect with the USPTO DB for A61).	The KIPO DB is more suitable for investigating regional patterns of innovation, but its top IPCs show similar long-term innovation patterns to other DBs; can be used for studying innovation in its leading technological areas and for long-term analysis.

considering its amount, variety and diverse sources of information as a whole. We also find that the EPO and JPO DBs are superior to the USPTO DB in the *intensity* of nations and IPCs but are inferior in the *coverage* of nations and IPCs as a data source of innovation studies. In addition the EPO DB has limited amount of data and the JPO DB has limited data accessibility. Thus, we argue that it is also reasonable to use the EPO and JPO DB for innovation studies, taking account of their merits and demerits compared to the USPTO DB. Unlike the three DBs, the KIPO DB is inferior to the USPTO in all aspects to be a global DB. In particular, the KIPO DB shows a high intensity of applicant nations and technological areas, which makes it inappropriate to use the DB for studying innovation patterns from the global perspectives.

A large proportion of its patents are granted to domestic assignees, so it is likely that innovation patterns observed in the KIPO DB show innovations by Korean companies (that is, regional patterns) rather than those of global companies. Of course, in some technology areas such as ICT, Korea leads global trends so that, if the focus were restricted to only to those areas and to long-term analysis, the KIPO DB analysis results based on the KIPO DB could be used as a proxy of global innovation patterns.

This study is one of the first attempts to analyze the characteristics of major patent DBs, and is expected to provide a basis for selecting DBs for various studies that seek to identify technology innovation patterns based on patent information. Unlike the previous studies which have implicitly assumed that any patent DB could represent global patterns of innovation for their studies without any explicit proof of that assumption, this study has taken a quantitative approach to prove that DBs most frequently used for innovation studies have been worth analyzing as data sources for innovation studies. According to the analysis, the implicit assumption made in the previous studies has been proved from actual data analysis, so the validity using patent data from the USPTO, EPO and JPO DBs in previous studies investigating the global patterns of innovation has been strongly confirmed.

Despite all these meaningful contributions, this research has some limitations that indicate the need for future research. The first limitation pertains to data access. In this study, we were unable to collect information regarding the assignees' countries from the JPO DB: this information was not freely available from either the JPO website or the WIPS, (either in English or in Korean). Not all patent offices provide the same level of format of information about their patents, so we could not fully investigate the characteristics of the four DBs from all three innovation perspective. Further studies will need to find a way to determine assignees' countries or develop a new proxy measure to analyze "innovation targets" using the available data. Considering other local DBs may be promising for the future research. In particular, it would be interesting to see the analysis results from SIPO (the State Intellectual Property Office of the People's Republic of China), as there has been a dramatic rise in the number of patent applications in China. Comparing the patterns of innovation in different regional patent DBs would help identify the characteristics of innovation in each region, though such differences may emanate either from regional innovation activities, or from the characteristics of local patent systems.

Second, in this study, we analyzed the characteristics of patent DBs based on three different perspectives: 'innovation activities', 'innovation participants', and 'innovation targets'. But, to arrive at a deeper understanding of patent DBs, it is necessary to consider various other perspectives and to design related indexes. For example, the data accessibility or the scope of information provided by DBs can be an important factor affecting the selection of patent DBs for innovation studies. The USPTO DB is well known for its ease of data access and richness of data content. One noticeable feature of this DB is that it provides citation data, so it can also be used to assess influences or economic values of specific technology or to calculate direct measures of technological knowledge flows. The citation data are invaluable for innovation studies, and have been used in many ways and for a variety of purposes. On the other hand, although KIPO started to provide citation data very recently, it is not possible to conduct citation analysis with data from KIPO because of its small data size. The JPO DB has distinguishing and attractive information element — the F-term — which is a Japanese patent classification system that classifies patents according to the technical features of their inventions to help searching for patents from various viewpoints, and so enables an in-depth analysis of technological trends. Since these characteristics affect the decisions on which patent DBs to use, they are worth being considered in future research.

Finally, this research used patent statistics as a proxy of innovation. Both patent applications and the registered patents (the focus of this research) are worth analyzing, as application data can also show patterns of innovation activities. Analyzing the differences of patent application patterns in different DBs — including USPTO, EPO, and PCT will be meaningful in helping understand the selection of patent DBs for innovation studies.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. NRF-2013R1A2A2A03016904).

Appendix A

Table A1
Definition of IPC codes *(Source: http://www.patent.go.kr/jsp/ka/menu/support/main/SupportMain0601.jsp.)*

Section	Class	Classification
A	A27	Furniture (A47)
Human necessities	A42	Headwear
	A61	Medical or veterinary science; hygiene
B	B02	Crushing, pulverizing, or disintegrating; preparatory treatment of grain for milling
Performing operations; transporting	B60	Vehicles in general
G	G01	Measuring; testing
Physics	G02	Optics
	G06	Computing; calculating; counting
H	H01	Basic electric elements
Electricity	H04	Electric communication technique
	H05	Electric techniques not otherwise provided for

References

Abraham, B.P., Moitra, S.D., 2001. Innovation assessment through patent analysis. Technovation 21 (4), 245–252.

Acs, Z., Audretsch, D., 1987. Innovation market structure and firm size. Rev. Econ. Stat. 69 (4), 567–574.

Akgüne, A.E., Keskin, H., Byrne, J., 2009. Organizational emotional capability, product and process innovation, and firm performance: an empirical analysis. J. Eng. Technol. Manag. 26 (3), 103–130.

Archibugi, D., Planta, M., 1996. Measuring technological change through patents and innovation surveys. Technovation 16 (9), 451–468.

Arundel, A., Kabla, I., 1998. What percentage of innovations are patented? Empirical estimates for European firms. Res. Policy 27 (2), 127–141.

Balkin, D.B., Markman, G.D., Gomez-Mejia, L.R., 2000. Is CEO pay in high technology firms related to innovation? Acad. Manag. J. 43 (6), 1118–1129.

Bouwer, E., Kleinknecht, A., 1999. Innovative output, and a firm's propensity to patent. An exploration of CIS micro data. Res. Policy 28 (6), 615–624.

Breschi, S., Catalini, C., 2010. Tracing the links between science and technology: an exploratory analysis of scientists' and inventors' networks. Res. Policy 39 (1), 14–26.

Chen, Y., Yang, Z., Shu, F., Hu, Z., Meyer, M., 2012. A patent based evaluation of technological innovation capability in eight economic regions in PR China. World Patent Inf. 31 (2), 104–110.

Comanor, W.S., Scherer, F.M., 1969. Patent statics as a measure of technical change. J. Polit. Econ. 77 (3), 392–398.

Cruz-Cázares, C., Bayona-Sáez, C., García-Marco, T., 2013. You can't manage right what you can't measure well: technological innovation efficiency. Res. Policy 42 (6–7), 1239–1250.

Ehrnberg, E., Jacobsson, S., 1997. Indicators of discontinuous technological change: an exploratory study of two discontinuities in the machine tool industry. R&D Manag. 27 (2), 107–126.

Ernst, H., 1995. Patenting strategies in the German mechanical engineering industry and their relationship to company performance. Technovation 15 (4), 225–240.

Gans, J., Stern, S., 2003. The product market and the market for ideas: commercialization strategies for technology entrepreneurs. Res. Policy 32 (2), 333–350.

Georg, G., Zahra, S., Wood, R., 2002. The effects of business university alliances on innovative outout and financial performance: a study of publicly traded biotechnology companies. J. Bus. Ventur. 17 (16), 577–609.

Geum, Y., Kim, C., Lee, S., Kim, M., 2013. Technological convergence of IT and BT: evidence form patent analysis. ETRI J. 34 (3), 439–449.

Godinho, M.M., Ferreira, V., 2012. Analyzing the evidence of an IPR take-off in China and India. Res. Policy 41 (3), 499–511.

Goto, A., Motohashi, K., 2007. Construction of a Japanese patent database and a first look at Japanese patenting activities. Res. Policy 36 (9), 1431–1442.

Hashi, I., Stojčić, N., 2013. The impact of innovation activities on firm performance using a multi-stage model: evidence from the community innovation survey. Res. Policy 42 (2), 353–366.

Hicks, D., Breitzman, T., Olivastro, D., Hamilton, K., 2001. The changing composition of innovative activity in the US — a portrait based on patent analysis. Res. Policy 30 (4), 681–703.

Hinze, S., Grupp, H., 1992. Applied research and industrial development in East Germany: international comparison by performance indicators. Technovation 12 (4), 257–278.

Hirschey, M., Richardson, V.J., Scholz, S., 2001. Value relevance of nonfinancial information: the case of patent data. Rev. Quant. Finan. Acc. 17 (3), 223–235.

Hitt, M.A., Hoskisson, R.E., Kim, H., 1997. International divercification: effects on innovation and firm performance in product-diversified firms. Acad. Manag. J. 40 (4), 767–798.

Hung, S., Tang, R., 2008. Factors affecting the choice of technology acquisition mode: an empirical analysis of the electronic firms of Japan, Korea and Taiwan. Technovation 28 (9), 551–563.

Jacobsson, S., Philipson, J., 1996. Sweden's technological profile: what can R&D and patents tell and what do they fail to tell us. Technovation 26 (5), 245–253.

Janodia, M., Rao, V., Pandey, S., Sreedhar, D., Ligade, V., Udupa, N., 2009. Impact of patents on India pharma industry's growth and competency: a perspective of pharmaceutical companies in India. J. Intellect. Prop. Rights 14, 432–436.

Kemp, R.G.M., Folkeringa, M., de Jong, J.P.J., Wubben, E.F.M., 2003. Innovation and firm performance. Scales Research Reports. EIM business and policy research, Zoetermeer (downloaded from http://www.ondernemerschap.nl/pdf-ez/H200207.pdf).

Klomp, L., Van Leeuwen, G., 2001. Linking innovation and firm performance: a new approach. Int. J. Econ. Bus. 8 (3), 343–364.

Kuznets, S., 1962. Inventive activity: problems of definition and measurement. The Rate and Direction of Inventive Activity: Economic and Social Factors. Princeton University Press for the National Bureau of Economic Research, Princeton.

Lee, S., Yoon, B., 2012. Applicability of patent information in technology forecasting: a sector-specific approach. J. Intellect. Prop. Rights 17 (1), 37–45.

Lee, S., Kang, S., Park, Y.S., Park, Y., 2007. Technology roadmapping for R&D planning: the case of the Korean parts and materials industry. Technovation 27 (8), 433–445.

Lee, S., Lee, S., Seol, H., Park, Y., 2008. Using patent information for designing new product and technology: keyword based technology roadmapping. R&D Manag. 38 (2), 169–188.

Lee, S.K., Mogi, G., Lee, S.K., Kim, J.W., 2010. Econometric analysis of the R&D performance in the national hydrogen energy technology development for measuring relative efficiency: the fuzzy AHP/DEA integrated model approach. Int. J. Hydrog. Energy 35 (6), 2236–2246.

Li, L.X., 2000. An analysis of sources of competitiveness and performance of Chinese manufacturers. Int. J. Oper. Prod. Manag. 20 (3), 299–315.

Li, Y.R., 2009. The technological roadmap of Cisco's business ecosystem. Technovation 29 (5), 379–386.

Loof, H., Heshmati, A., 2002. Knowledge capital and performance heterogeneity: a firm level innovation study. Int. J. Prod. Econ. 76 (1), 61–85.

Ma, Z., Lee, Y., 2008. Patent application and technological collaboration in inventive activities: 1980–2005. Technovation 28 (6), 379–390.

Martino, J.P., 1993. Technological Forecasting for Decision Making. McGraw-Hill, Inc., NY.

Nosella, A., Petroni, G., Salandra, R., 2008. Technological change and technology monitoring process: evidence from four Italian case studies. J. Eng. Technol. Manag. 25 (4), 321–337.

O'Regan, N., Ghobadian, A., Galler, D., 2006. Is search of the drivers of high growth in manufacturing SMEs. Technovation 26 (1), 30–41.

OECD/Eurostat, 1997. OECD Proposed Guidelines for Collecting and Interpreting Technological Innovation Data—Oslo Manual. OECD, Paris.

Paci, R., Sassu, A., Usai, S., 1997. International patenting and national technological specialization. Technovation 17 (1), 25–38.

Schwartz, M., Peglow, F., Fritsch, M., Günther, J., 2012. What drives innovation output from subsidized R&D cooperation? — Project-level evidence from Germany. Technovation 32 (6), 358–369.

Storto, C.L., 2006. A method based on patent analysis for the investigation of technological innovation strategies: the European medical prostheses industry. Technovation 26 (8), 932–942.

Subramanian, A.M., Chai, K.H., Mu, S., 2011. Capability reconfiguration of incumbent firms: Nintendo in the video game industry. Technovation 31 (5–6), 228–239.

Teichert, T., Mittermayer, M.A., 2002. Text mining for technology monitoring. Proceedings of IEEE International Conference on, Engineering Management (IEMC), pp. 596–601.

Thongpapanl, N.T., 2012. The changing landscape of technology and innovation management: an updated ranking of journals in the field. Technovation 32 (5), 257–271.

Wang, E.C., Huang, W., 2007. Relative efficiency of R&D activities: as cross county study accounting for environmental factors in DEA approach. Res. Policy 36 (2), 260–273.

Wang, M.Y., Chang, D.S., Kao, C.H., 2010. Identifying technology trends for R&D planning using TRIZ and text mining. R&D Manag. 40 (5), 491–509.

WIPO, 2010. World Intellectual Property Indicator. WIPO, Geneva.

Wu, C.Y., Mathews, J.A., 2012. Knowledge flows in the solar photovoltaic industry: insights from patenting by Taiwan, Korea, and China. Res. Policy 41 (3), 524–540.

Zahra, S.A., Nielsen, A.P., 2002. Sources of capabilities, integration and technological commercialization. Strateg. Manag. J. 23 (5), 377–398.

Jeeun Kim is a PhD candidate on industrial engineering of at Ajou University, Suwon, Rep. of Korea. Her main research interests include technology planning, patent analysis, and service engineering. She has authored many articles related to the technology management, which have been presented in various international conferences.

Sungjoo Lee received her BS and PhD in technology management from Seoul National University, Seoul, Rep. of Korea, in 2002 and 2007, respectively. After spending six months as a senior researcher at the Ubiquitous Computing Innovation Center, she moved to the UK to work as a visiting scholar at the University of Cambridge. She is currently an assistant professor at Ajou University, Suwon, Rep. of Korea. Her research interests include patent analysis, technology roadmapping, and intellectual property management.