Technological Forecasting & Social Change 97 (2015) 181–192

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining

Hyun Joung No [a,1], Yoonjung An [b], Yongtae Park [b,*]

[a] School of Mechanical and Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-744, Republic of Korea
[b] Department of Industrial Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-744, Republic of Korea

ARTICLE INFO

Article history:
Received 1 January 2014
Received in revised form 13 April 2014
Accepted 16 April 2014
Available online 13 May 2014

Keywords:
Knowledge flow
Business method
Patent citation analysis
Text mining
Technological classes

ABSTRACT

With information and communication technology (ICT) as an enabling platform, diversified new business methods (BMs) have been developed. These new technology-based BMs have played an important role in knowledge flow as they became a patentable subject matter. However, there are not many studies on knowledge flow through the technology-based BMs or BM patents in spite of its importance. As an attempt to provide a deeper understanding of technology-based BMs with regard to knowledge flow, this paper explores knowledge flows driven by the technology-based BMs through investigating both cited and citing patents. In order to explore the knowledge flows, this paper proposes an algorithm that utilizes both the citation and textual information of BM patents. In addition to citation information, text data in patent documents are used to measure the degree of knowledge flow in a more accurate way. A case study is conducted with the BM patents related to postage metering system and the analysis result is presented in a positioning map that shows different knowledge flow patterns of technological classes. Moreover, the technology-based BM patents as knowledge flow drivers are classified based on the amount of knowledge exchanged between the base BM patents and their patent citations.

1. Introduction

Business environment has been changing rapidly since the 1990s and companies have to constantly develop new business methods/business models to keep up with such changes and survive in the market. The most noticeable change among all is the substantial application of ICT in the business environment; ICT has become a new enabler for business communication and processing commercial transactions [1–6]. With ICT as a critical part of new BMs, BMs are now one of the patentable subject matters that gain special attention, and a growing number of companies try to seek patent protection for the new BMs [3]. As a result of that, the number of BM patents grew rapidly over a very short period [6,7].

Despite their importance in the business environment, it is difficult to find previous studies discussing about technology-based BMs with respect to their relationships with technologies, technology-based BMs of other kinds or knowledge flow. Recently, some researchers made attempts to study such topics. Kim et al. [8] identified between technology-based services, which are represented by the BMs, and ICTs. Some studies aimed to explore technology diffusion of BMs [1] and identify internal technological relationships among the BM patents and patterns of business model evolution [9]. Other studies merely focused on explaining the theoretical background of impact of technology on BMs/business model innovation [2–4,6].

* Corresponding author. Tel.: +82 2 880 8358; fax: +82 2 889 8560.
E-mail addresses: anesno1@snu.ac.kr (H.J. No), parkyt1@snu.ac.kr (Y. Park).
[1] Tel.: +82 10 7148 1978, fax: +82 2 878 3511.

http://dx.doi.org/10.1016/j.techfore.2014.04.007
0040-1625/

There are different terminologies used for BMs that include ICTs, but there is no specific terminology or definition that is widely agreed upon among researchers or practitioners. Hunt [10] called such BMs "computer-implemented business methods." Wu [6] used two different terms which are "software-embodied business methods" and "internet business methods." Those BMs are also simply called "business methods" [11] and "business methods based on Internet technologies" [1]. In this paper, they will be called "technology-based business methods" and narrowly defined as "type of business methods limited to patentable subject matter classified in USPTO Class 705 which only includes business methods based on technologies."

The role of knowledge exchange is especially important in a knowledge and technology driven economy because it allows better penetration and diffusion of innovation and stimulates cooperation in R&D [12,13]. There have been extensive studies emphasizing the importance of knowledge flow/spillover. Glaeser et al. [14] suggest that knowledge flow/spillover is directly linked to three factors of economic growth, which are specialization, competition and diversity, and they are characterized by a higher intensity of intra-industry knowledge spillover, inter-firm innovation flows and inter-industry knowledge exchange, respectively. Also, Huggins and Johnston [15] argue that knowledge exchange through networking with various partners in different domains can open opportunities for novel combination and recombination of ideas or best-of-breed solutions that originate from different resource bases and knowledge bases. Such knowledge networks are thus an important aspect of the innovation process [15–17].

Although it is not difficult to conceptualize a phenomenon of knowledge flow, it is very difficult to measure the degree of knowledge flow [12]. The two main methods are direct and indirect [18]. The main direct method is to use information in patent citations. The indirect method of measuring knowledge flows typically regresses total factor productivity (TFP) growth on factors thought to be potentially causing information flows, such as the presence of multinational enterprises (MNEs) or international-trade status. [18]. They both have advantages and disadvantages, but we decide to use patent citation as a measure of knowledge flow due to the following reasons: patent citation is a certified evidence of previous knowledge used by the inventor [19], data can be obtained easily and International Patent Classification (IPC) corresponds with the purpose of our study.

Since BMs have been disclosed to the public in the form of a patent, meaning they are more exposed to knowledge flows, it is worth studying important implications of the BM patents in terms of knowledge flow. The BM patents enable effective measurement of knowledge flow of BMs, with citation and other information. Knowledge flow is stimulated by the BM patents through active cited (backward citation) and citing (forward citation) patents. There exist some previous studies showing the empirical evidence that the BM patents not only cite a significant number of previous patents but are also cited by a substantial number of subsequent patents [11,20].

Both cited and citing patents represent knowledge flow in a similar manner, but the underlying economic rationales of these two processes differ [21,22]. Cited patents (backward citations) have been used to measure technological knowledge acquired by the patenting entity and thus regarded as knowledge utilization; on the other hand, citing patents (forward citations) have been interpreted as a measure of the knowledge diffusing outward from the patenting entity and thus regarded as knowledge dissemination [21,22]. The more frequently a patent is cited by patenting entities, the greater the related technology may have influenced, implying that the technological knowledge is more widely disseminated. Since the BM patents have a substantial number of both cited and citing patents, it can be interpreted that they play an important role in utilizing and disseminating knowledge.

Although there are numerous studies using patent citation information as a proxy for knowledge flow between technologies or actors, and patent citations encapsulate important information about knowledge flow, there are still some drawbacks to use citation information. Patent citations, which are linked to the patenting procedure itself, capture only the knowledge flows, thus underestimating the actual extent of knowledge flows [23]. Also, they could be biased by incorrect citing of sources; thus supplementary investigation is required to allow citation information to be confidently applied [21].

In order to overcome the drawback of citation based approach, text mining, using textual data to discover useful pattern, can be applied along with citation analysis. Co-word analysis is mainly utilized to explore the concept network in different fields since the nature of words, on which co-word analysis is based, can act as the important carrier of knowledge [24]. Words and co-occurrences of words cover a much broader domain than citations [25]. Words occur not only as indicators of links among documents but also internally within documents. Thus, the text data can be used to measure a degree/amount of knowledge transferred by measuring text similarities between patents while patent citation is used to measure a path of knowledge flows.

As an attempt to provide a deeper understanding of technology-based BMs with regard to knowledge flow, the paper proposes a framework for exploring knowledge flows driven by technology-based BMs from their utilized technologies to disseminated technologies, by investigating both cited and citing patents. The proposed approach integrates patent citation analysis and text mining to explore the knowledge flow through technology-based BMs. First, knowledge flow path is traced using citation links between technology-based BM patents and their cited patents, which represent utilized knowledge sources, and between technology-based BM patents and their citing patents which represent disseminated knowledge. Then, co-word analysis as a text mining is integrated with the citation analysis to verify the degree of knowledge transferred between BM patents and cited/citing patents by measuring the text similarity between BM patents and their utilized/disseminated knowledge source. The integrated approach will lead to a better measurement of knowledge flow in terms of the degree of knowledge flow.

2. Proposed approach: integrating patent citation analysis and text Mining

2.1. Overall research process

In addition to the fact that patent information is better protected from data disruption than other database, citation information provides citation links which can be used to analyze technological diffusion, valuation or impact, among various patents. In many studies [26–29], patent citation information has been frequently used to construct the knowledge flow matrices for measuring knowledge flows.

In this paper, patent citation and text data are integrated to classify technology-based BM patents as knowledge flow drivers, and to measure the degree of knowledge flow driven by technology-based BMs.

The proposed approach consists of four steps as shown in Fig. 1. First, technology-based BM patents, which are base patents in the research, and their cited/citing patents are collected from the USPTO database. In the research, the base patents are the subject of analysis, which act as a mediator of knowledge flow through cited and citing patents. For the base patents, all the patents belonging to research-related patent class (USPTO classes, in our case) are collected. Second, keywords are extracted from the abstracts of base patents and citing/cited patents to measure the degree of knowledge flow. The lists of descriptors are standardized to delete a variant of the same word. For the third step, after constructing the keyword matrix of base patents, and of cited/citing patents, textual similarities between the base BM patents and the cited/citing patents which are clustered as patent classes (i.e. USPTO classes or subclasses) are computed to measure what degree of knowledge is actually exchanged between them. Cosine similarity is used as a similarity measure since it is easy to interpret and simple to compute for long and sparse vectors [30–32]. Lastly, patterns of knowledge flows driven by technology-based BMs are identified in terms of patent classes. In other words, the patent classes are categorized into different groups based on the degree of knowledge the base BM patents utilized or disseminated.

2.2. Integrated approach by combining patent citation analysis and text mining

The idea of studying the text information combining with bibliometric methods and vice versa is not new to the literature. There are several studies combining bibliometric information with textual content to obtain improved performance in clustering [33], classification [34] and bibliometric mapping [35]. However, most studies focus on combining co-citation with word analysis in the context of evaluative bibliometrics in order to improve efficiency of co-citation clustering and bibliometric mapping [36]. Since text-based approach usually is based on rather rich vocabularies and peculiarities of natural language, the relationship between documents is somewhat fuzzy and not always reliable. On the other hand, if strict citation-based criteria are applied, that is, if non-periodical references and occasional coupling links are removed, the resulting citations-by-document matrix becomes extremely sparse. Combining two techniques helps to improve the reliability of relationship and the clustering algorithm as well [37].

The text information and citation information are combined and configured, as shown in Fig. 2, to measure the similarity among base BM patents and their cited/citing patents in order to explore the knowledge flows through technology-based business methods.

2.3. Similarity measure

The issue of document similarity attained more attention with increasing interest in information retrieval. It is a measure used to compare two objects and to determine if they are

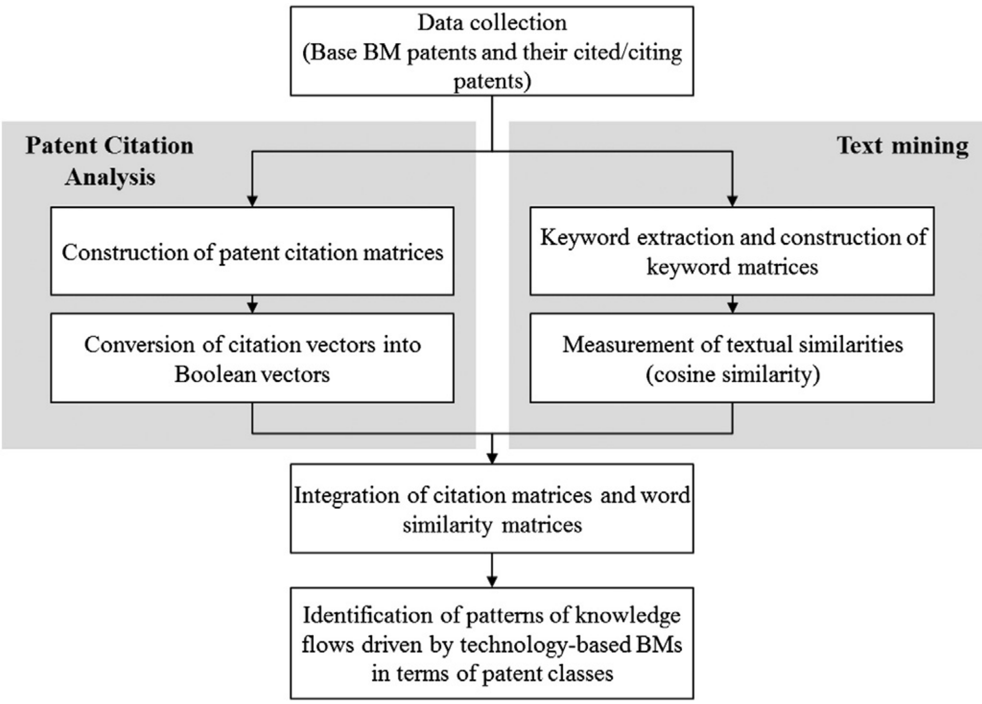


Fig. 1. Overall process for exploring knowledge flows driven by technology-based BMs.

related to the same topics [38]. Similarity measures yield an indication of the relevance of an object, a document, to a given standard, the query. According to McGill et al. [39], there are more than 60 different similarity measures; dice coefficient, cosine coefficient, overlap coefficient [32], and the spreading activation similarity measure. Each similarity measure has its strengths and weaknesses in practice.

The distance-based similarity measure and the angle-based cosine measure are the most popular measures. However, the distance-based similarity measure takes only the impact of the distance into account. Thus, documents with the same distance to the reference point shall have the same similarity regardless of the direction of the document [40]. The cosine measure can effectively identify documents in a vector document space that have the same indexing term distribution within each document [41]. That is, the same proportion of weights is given for any pair of indexing terms between two documents if they have the same indexing terms. The most widely used measure is still the cosine similarity in the vector space model and the cosine similarity is easy to interpret and simple to compute for long and sparse vectors [30,32,42].

In the paper, keyword-based similarity values are measured, from the 'Patents by keywords' matrix, using the concept of the cosine measure of 'Salton & McGill' [32] which is defined as the cosine of the angle enclosed between two vectors x and y. The cosine measure of 'Salton & McGill' has an advantage over the Pearson correlation in that the similarity is insensitive to the number of zeros as the cosine is not based on the mean of the distribution [42].

In this research, the similarity between a base BM patent and cited/citing citation, at patent-class (USPTO) level, is defined as

$$Sim(P,C) = \frac{\sum_{K=1}^{n} P_k C_k}{\sqrt{\sum_{K=1}^{n} {P_k}^2}\sqrt{\sum_{K=1}^{n} {C_k}^2}}$$

where P_k is the frequency of keyword k in base patent P and C_K is the frequency of keyword k in cited/citing patents at patent-class level. The boolean case of $P_k = 0$ when there is no keyword existing in the corresponding patent, or $P_k = 1$ when there is a keyword. The boolean case of $C_k = 0$ when there is no keyword existing in corresponding citation patent, or $C_k = 1$ when there is a keyword. n is the total number of patent-class for base patents and citation patents.

Since the cosine formula normalizes for the length of the word-profiles of both object (base patent) and query (citation

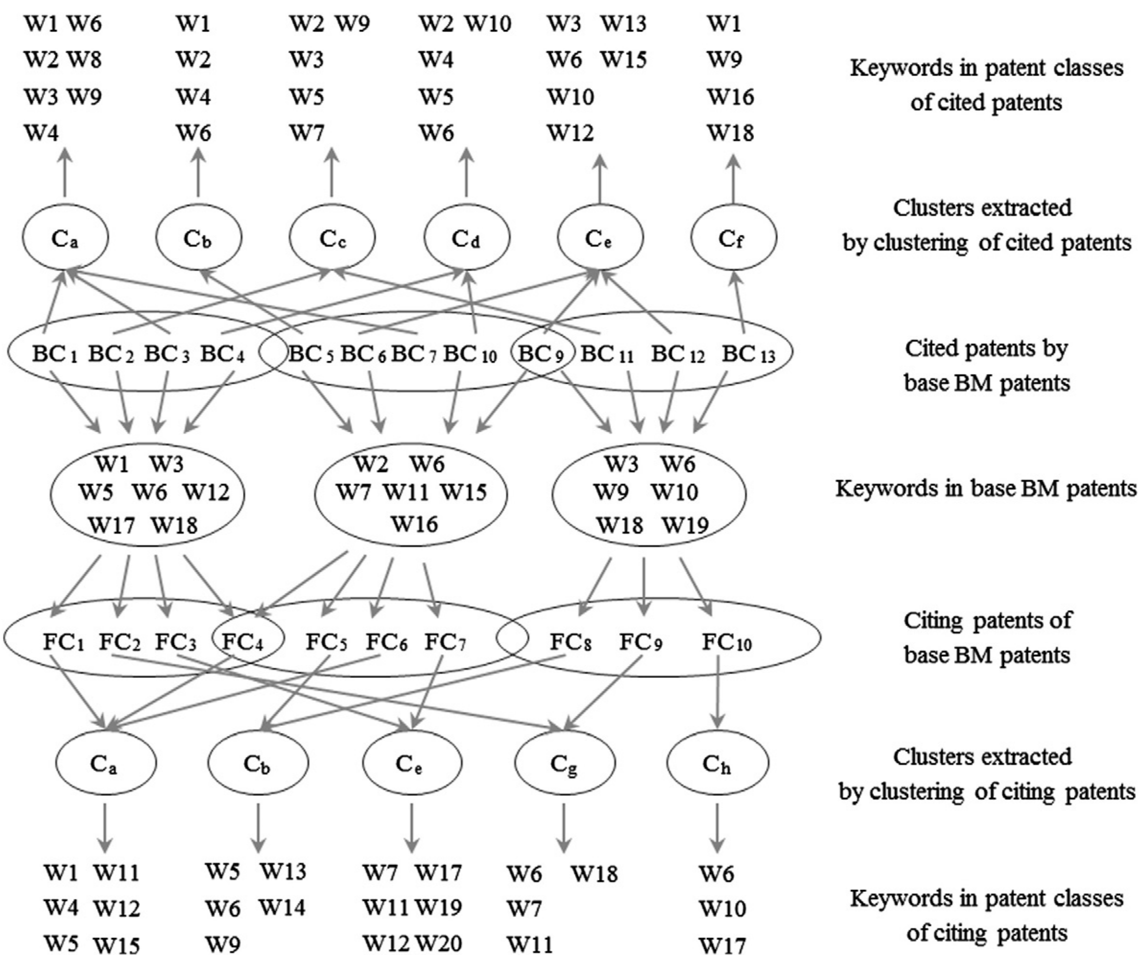


Fig. 2. Citation analysis and text mining in an integrated approach.

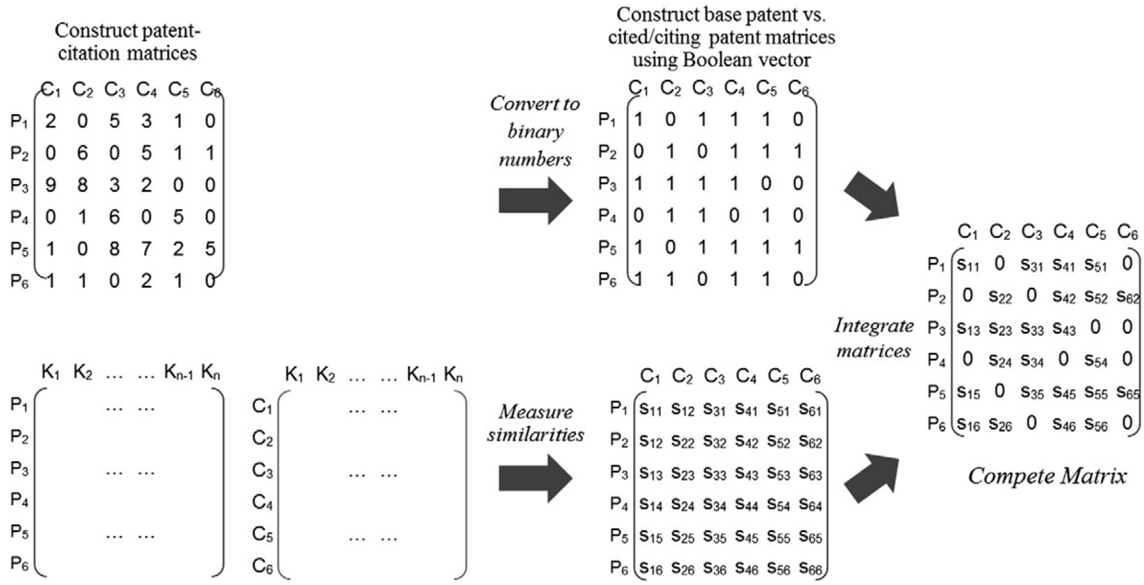


Fig. 3. Integration of patent citation and text mining.

patents), objects with long word-profiles can be penalized for their 'representational richness' if this does not correspond to the richness in the query's representation [43].

The keyword-based matrices for base patents and for cited/citing patents at patent-class level are constructed, then the keyword-based similarity matrices for "base patents by cited/citing patents" are constructed based on the value of the keyword-based similarity. As seen in Fig. 3, there are similarity values if there is a direct citation relationship and 0 otherwise, by using boolean vectors of 'base patent by cited/citing patents' as a weight.

3. Case study: postage metering system

3.1. Data collection

As explained in the research framework, all the data is retrieved from the USPTO database. The postage metering system related BM patents are selected as the base BM patents. Under the USPTO classification scheme, Class 705 is further divided into subclasses based on subjects, such as operations research, POS terminal or electronic cash register, electronic shopping and finance. Among the various subjects, postage metering system (060–062, 400–411) is selected as the base BM patents for our empirical study. The postage metering system BM patents belong to two different subclasses: cryptography and cost/price. Subclasses 060–062, belonging to cryptography, are defined as the subject matter wherein a charge for mailing an article is determined, markings representing this charge are affixed to the article, and respective modifications to an account balance are made. Subclasses 400–411, belonging to cost/price, are defined as the subject matter wherein the data processing or calculating computer comprises means for determining and printing cost required for mailing an article.

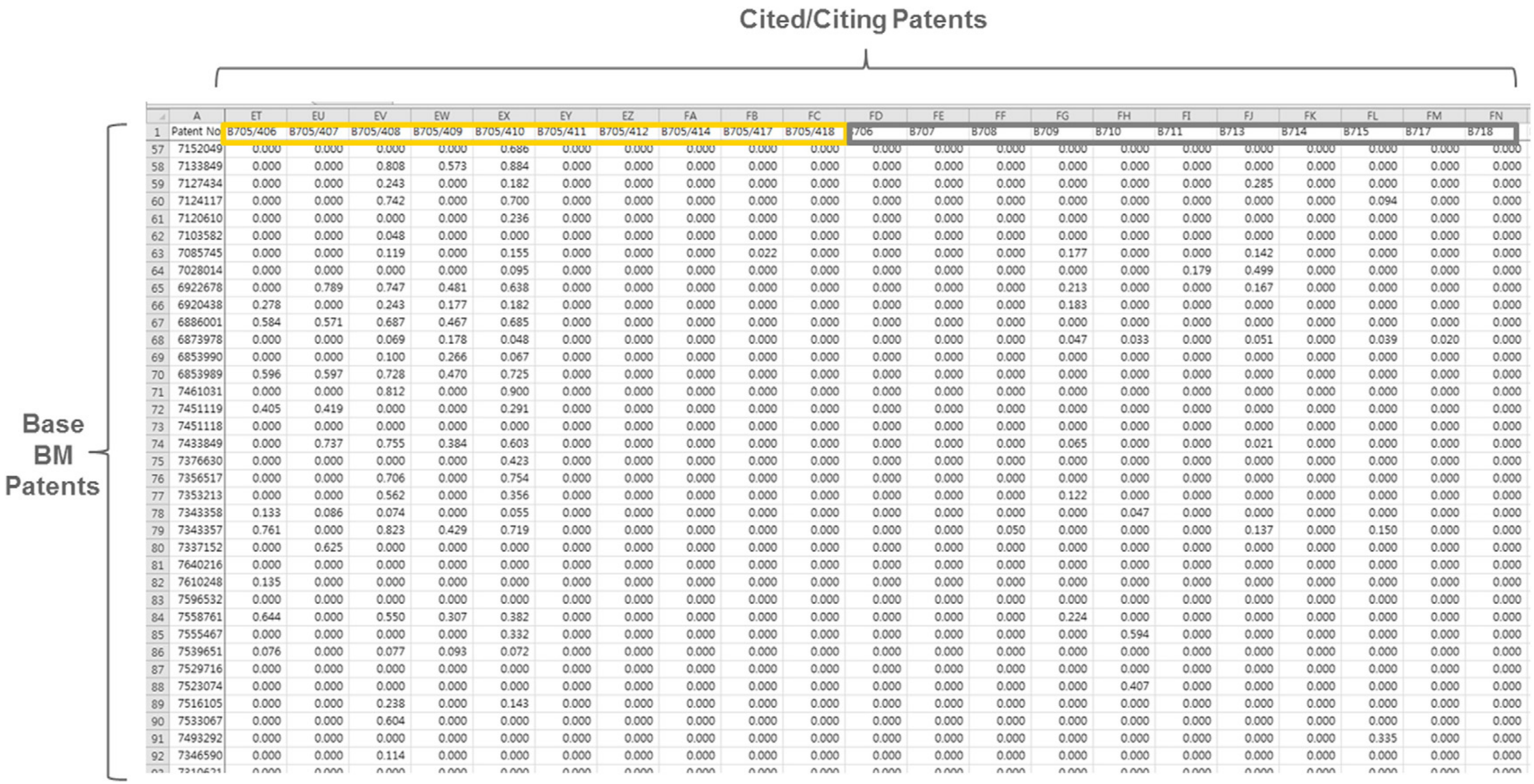
In terms of a time span for building the dataset, the patents issued from 2004 to 2009 are specifically chosen because there needs to be a certain period of time for the base patents to have a sufficient number of both cited and citing patents. The number

Table 1
Base patents (postage metering system BMs) and cited/citing citations.

Base/citation	Number of patents	Descriptions
Base patents	149	Issued from 2004 to 2009—need a certain period of time to have citing patents (forward citations)
Backward citations	3494	Eliminated those that have not been issued and those that do not include abstract
Forward citations	643	Eliminated those that do not include abstract

Table 2
Subclasses of base patents.

Subclass (14)	Title	No. ofbase patents
705/060	Postage metering system (Cryptography)	28
705/061	Reloading/recharging	3
705/062	Having printing detail (e.g., verification of mark)	16
705/401	Postage metering system (Cost/Price)	41
705/402	Special service or fee (e.g., discount, surcharge, adjustment, etc.)	11
705/403	Recharging	2
705/404	Record keeping	7
705/405	Data protection	1
705/406	With specific mail handling means	3
705/407	Including mailed item weight	8
705/408	Specific printing	20
705/409	Rate updating	1
705/410	Specialized function performed	7
705/411	Display controlling	1



	A	ET	EU	EV	EW	EX	EY	EZ	FA	FB	FC	FD	FE	FF	FG	FH	FI	FJ	FK	FL	FM	FN
1	Patent No	B705/406	B705/407	B705/408	B705/409	B705/410	B705/411	B705/412	B705/414	B705/417	B705/418	706	B707	B708	B709	B710	B711	B713	B714	B715	B717	B718
57	7152049	0.000	0.000	0.000	0.000	0.686	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
58	7133849	0.000	0.000	0.808	0.573	0.884	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
59	7127434	0.000	0.000	0.243	0.000	0.182	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.285	0.000	0.000	0.000	0.000
60	7124117	0.000	0.000	0.742	0.000	0.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.094	0.000	0.000
61	7120610	0.000	0.000	0.000	0.000	0.236	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
62	7103582	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
63	7085745	0.000	0.000	0.119	0.000	0.155	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.177	0.000	0.000	0.142	0.000	0.000	0.000	0.000
64	7028014	0.000	0.000	0.000	0.000	0.095	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.179	0.499	0.000	0.000	0.000	0.000
65	6922678	0.000	0.789	0.747	0.481	0.638	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.213	0.000	0.000	0.167	0.000	0.000	0.000	0.000
66	6920438	0.278	0.000	0.243	0.177	0.182	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.183	0.000	0.000	0.000	0.000	0.000	0.000	0.000
67	6886001	0.584	0.571	0.687	0.467	0.685	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
68	6873978	0.000	0.000	0.069	0.178	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.047	0.033	0.000	0.051	0.000	0.039	0.020	0.000
69	6853990	0.000	0.000	0.100	0.266	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
70	6853989	0.596	0.597	0.728	0.470	0.725	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
71	7461031	0.000	0.000	0.812	0.000	0.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
72	7451119	0.405	0.419	0.000	0.000	0.291	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
73	7451118	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
74	7433849	0.000	0.737	0.755	0.384	0.603	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.021	0.000	0.000	0.000	0.000
75	7376630	0.000	0.000	0.000	0.000	0.423	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
76	7356517	0.000	0.000	0.706	0.000	0.754	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
77	7353213	0.000	0.000	0.562	0.000	0.356	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.122	0.000	0.000	0.000	0.000	0.000	0.000	0.000
78	7343358	0.133	0.086	0.074	0.000	0.055	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.047	0.000	0.000	0.000	0.000	0.000	0.000
79	7343357	0.761	0.000	0.823	0.429	0.719	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.137	0.000	0.150	0.000	0.000
80	7337152	0.000	0.625	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
81	7640216	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
82	7610248	0.135	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
83	7596532	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
84	7558761	0.644	0.000	0.550	0.307	0.382	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.224	0.000	0.000	0.000	0.000	0.000	0.000	0.000
85	7555467	0.000	0.000	0.000	0.000	0.332	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.594	0.000	0.000	0.000	0.000	0.000	0.000
86	7539651	0.076	0.000	0.077	0.093	0.072	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
87	7529716	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
88	7523074	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.407	0.000	0.000	0.000	0.000	0.000	0.000
89	7516105	0.000	0.000	0.238	0.000	0.143	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
90	7533067	0.000	0.000	0.604	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
91	7493292	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.335	0.000	0.000
92	7346590	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Fig. 4. Integrated matrix of “Boolean citation matrix” and “Word-similarity matrix”.

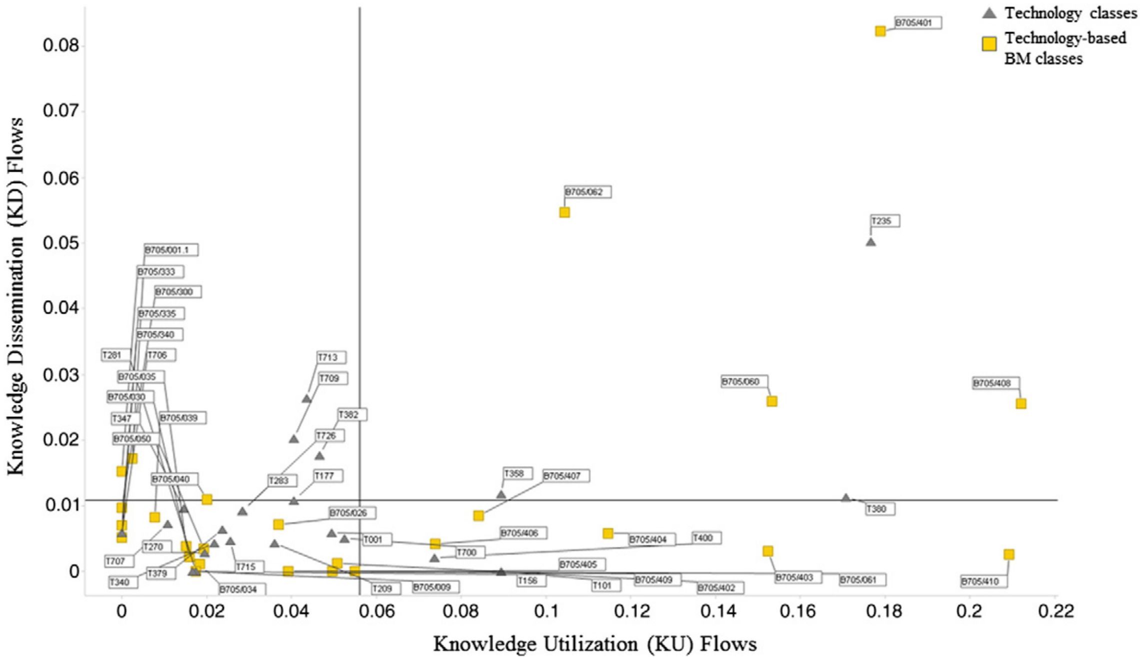


Fig. 5. Positioning of technological classes with relation to knowledge flow driven by technology-based BMs.

of base patents is 149 and that of cited and citing patents is 3494 and 643, respectively. The detailed information about the base patents and citations are listed in Tables 1 and 2.

3.2. Construction and integration of matrices

The dataset for base BM patents includes backward and citations and their text information as well. It is extended by the following operations:

1) Build the dataset for cited patents (backward citations) of base BM patents
2) Build the dataset for citing patents (forward citations) of base BM patents
3) Construct the Boolean matrices of "base patent vs. cited/citing patents"
4) Build the dataset for keywords of base BM patents and cited/citing patents
5) Construct the keyword matrix of "base BM patents vs. keywords" and "patent class vs. keywords" for cited/citing patents
6) Measure similarities based on the keyword matrix of "base BM patents vs. keywords" and "patent class vs. keywords" for cited/citing patents
7) Integrate the Boolean matrices of "base BM patents vs. cited/citing patents" and keyword similarity matrices of "base BM patents vs. cited/citing patents" (seen in Fig. 4).

Table 3
Patterns of knowledge flows through technology-based BMs in terms of technological classes.

Knowledge flow pattern	Description	Affiliated technological classes		
		Category	Characteristics	Examples
High KU–high KD	High degree of both knowledge inflow and outflow through technology-based BMs	Technology	ICTs mainly related to subject (cryptography & cost/price) specific technologies	Register, cryptography, etc.
		Technology-based BM	Postage metering system BMs that mainly include subject (cryptography & cost/price) specific technologies	Postage meter system (cryptography & cost/price), etc.
High KU–low KD	High degree of knowledge inflow and low degree of knowledge outflow through technology-based BMs	Technology	General technologies related to printing	Printing, typewriting machines
		Technology-based BM	Postage metering system BMs that include supporting technologies	Specialized function performed, recharging, record keeping, etc.
Low KU–high KD	Low degree of knowledge inflow and high degree of knowledge outflow through technology-based BMs	Technology	ICTs related to data/image processing	Electrical computers and digital processing systems, image analysis
		Technology-based BM	BMs that include business system infra technologies	Special goods or handling procedure, automated electrical financial or business practice or management arrangement

Table 4
The average degree of knowledge flows between technology-based BM patents and their cited/citing patents.

Technology-based BM patents (Subclass)	Cited/citing patents from technology patent classes		Cited/citing patents from BM patent classes	
	Degree of knowledge flow between cited patents and BM patents	Degree of knowledge flow between BM patents citing patents	Degree of knowledge flow between cited patents and BM patents	Degree of knowledge flow between BM patents and citing patents
705/060	1.08	0.25	1.47	0.30
705/061	1.26	0.59	1.49	0.36
705/062	1.75	0.21	2.23	0.21
705/401	1.41	0.16	2.27	0.19
705/402	1.33	0.19	1.57	0.34
705/403	0.93	0.00	2.79	0.15
705/404	1.16	0.24	1.60	0.70
705/405	1.63	0.56	0.21	0.13
705/406	1.40	0.97	1.33	0.52
705/407	0.97	0.31	1.16	0.78
705/408	1.06	0.23	1.26	0.48
705/409	1.93	0.52	3.07	0.31
705/410	1.23	0.55	2.04	0.42
705/411	1.98	0.28	1.17	0.79

The keywords are extracted from the abstracts, which contain the essential information of the patents, using the text-mining package, "TextAnalyst 2.1." A total of 989 keywords are extracted; however, it should be noted that the software does not understand the context or meaning of words so that there may be some irrelevant or insignificant words included in the result. Those insignificant words are then manually screened out and a total of 490 keywords are selected as a final set.

After mining keywords, the resulting lists of descriptors were standardized to eliminate different spellings and variants of the same terms. In contrast to the conventional co-word analysis which generates a symmetrical matrix with an empty diagonal, matrix of patents vs. keywords are asymmetrical.

3.3. Patterns of knowledge flow

The technological classes are positioned on the two-dimensional map based on the amount of knowledge exchanged with the base BM patents, as shown in Fig. 5. Taking the mean value as a criterion, each class can be classified as either high (above mean) or low (below mean) in the knowledge utilization (KU) dimension or knowledge dissemination (KD) dimension.

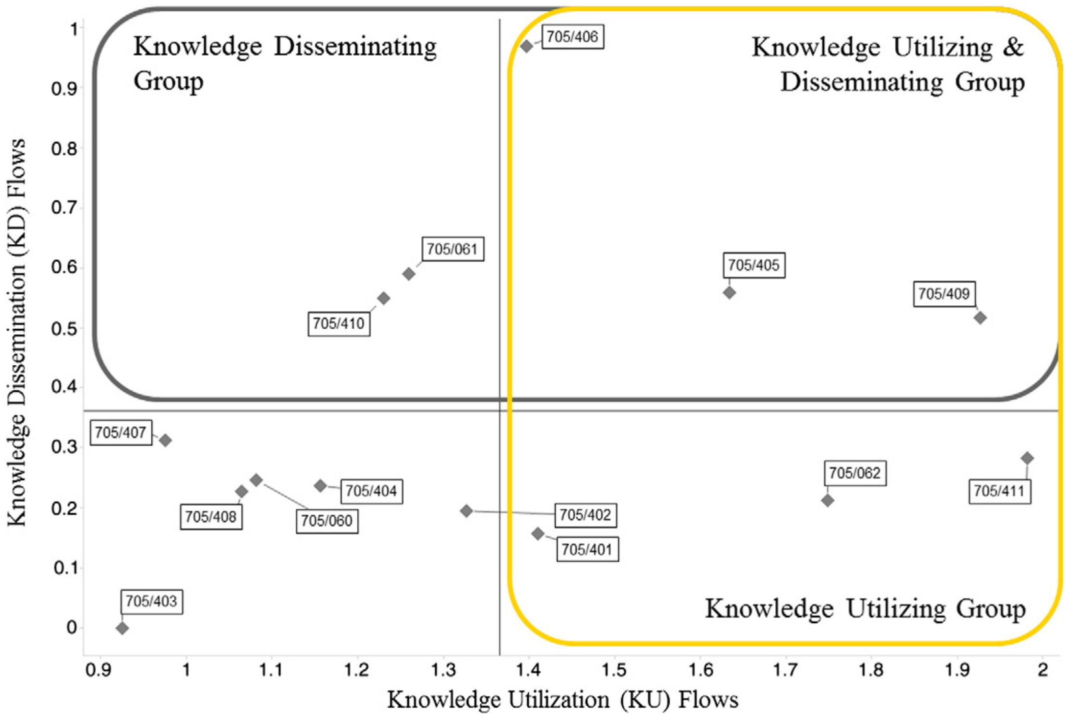


Fig. 6. Positioning of base patent classes driving knowledge flows in technology classes.

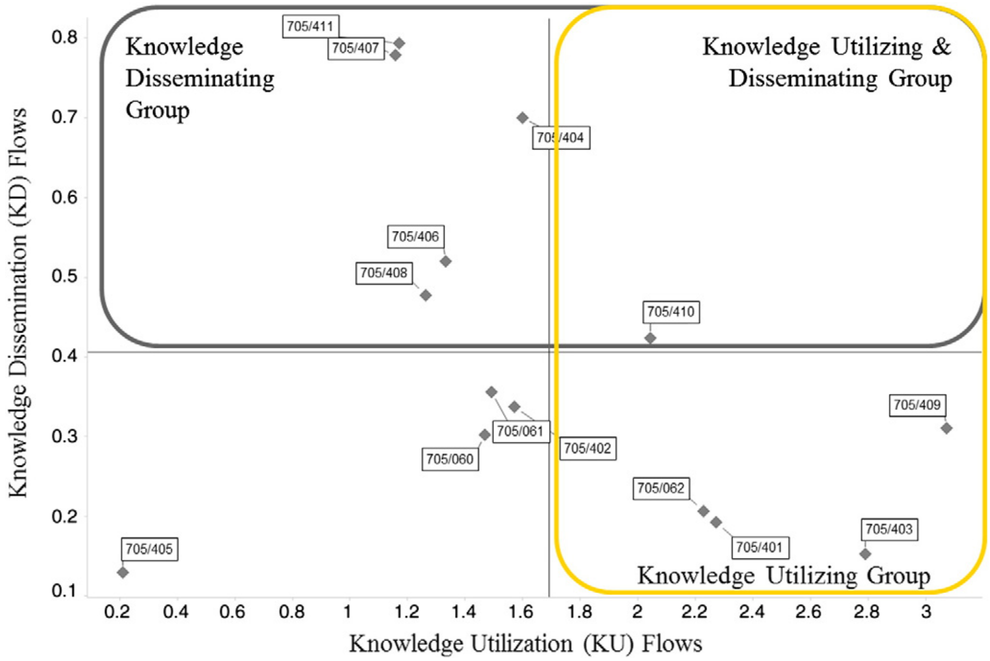


Fig. 7. Positioning of base patent classes driving knowledge flows in technology-based BM classes.

The classes are then categorized into three groups, depending on the amount of knowledge flow: high KU–high KD, high KU–low KD and low KU–high KD. The detailed characteristics of knowledge flow patterns are described in the following Table 3.

3.3.1. High KU–high KD

The technological classes in this group are characterized by a high degree of both knowledge inflow and outflow through the technology-based BMs. They are mainly ICTs and the postage metering system BMs related to subject (cryptography & cost/price) specific technologies: for example, register cryptography, postage meter system (cryptography & cost/price), etc.

3.3.2. High KU–low KD

The technological classes in this group are characterized by a high degree of knowledge inflow and a low degree of knowledge outflow through the technology-based BMs. The affiliated classes include general technologies related to printing and the postage metering system BMs with supporting technologies; for example, printing, typewriting machines, specialized function performed, recharging, record keeping, etc.

3.3.3. Low KU–high KD

The technological classes in this group are characterized by a low degree of knowledge inflow and a high degree of knowledge outflow through the technology-based BMs. The affiliated classes include ICTs related to data/image processing and the BMs with business system infra technologies; for example, electrical computers and digital processing systems, image analysis, special goods or handling procedure, automated electrical financial or business practice.

3.4. The degree of knowledge flow driven by technology-based BM patents

Table 4 presents the average degree of knowledge flows driven by technology-based BM patents. The average value is used to see the contribution of individual base BM patent class to knowledge flow.

Table 5
Group of knowledge flow drivers: technology classes.

Group	Subclass	Title
Knowledge utilizing & disseminating group	705/409	Rate updating
	705/405	Data protection
	705/406	With specific mail handling means
Knowledge utilizing group	705/411	Display controlling
	705/062	Having printing detail (e.g., verification of mark)
	705/401	Postage metering system (cost/price)
Knowledge disseminating group	705/061	Reloading/recharging
	705/410	Specialized function performed

Table 6
Group of knowledge flow drivers: technology classes & technology-based BM classes.

Group	Subclass	Title
Knowledge utilizing & disseminating group	705/410	Specialized function performed
Knowledge utilizing group	705/409	Rate updating
	705/403	Recharging
	705/401	Postage metering system (cost/price)
	705/062	Having printing detail (e.g., verification of mark)
Knowledge disseminating group	705/411	Display controlling
	705/407	Including mailed item weight
	705/404	Record keeping
	705/406	With specific mail handling means
	705/408	Specific printing

The technology-based BM patents in 705/409 (Rate updating) and 705/411 (Display updating) class are characterized by a high degree of knowledge inflow from technology patent classes. While 705/409 BM patents utilized a high degree of knowledge from technology-based BM patent classes as well, 705/411 BM patents utilized a relatively low degree of knowledge from technology-based BM patent classes. In the case of 705/403 and 705/401, it utilized the knowledge much more from technology-based BM.

In the aspect of knowledge dissemination, 705/411 BM patents have relatively less influence on technology and a very high degree of knowledge dissemination on business methods. 705/407 and 705/404 BM patent classes are also the ones which have disseminated a high degree of knowledge to BM patent classes. On the other hand, 705/409, 705/406, 705/405 and 705/062 BM patent classes disseminated knowledge more to technology classes than BM patent classes.

3.5. Classification of knowledge flow drivers

The base BM patents as knowledge flow drivers are classified based on the amount of knowledge exchanged between the base BM patents and cited/citing patents, as shown in Figs. 6 and 7. Taking the mean value as a criterion, the base BM patents at patent class level are then classified into three groups, depending on the amount of knowledge flow in and out: knowledge utilizing & disseminating group, knowledge utilizing group and knowledge disseminating group. Tables 5 and 6 present the groups of knowledge flow drivers in technology classes and in technology classes & technology-based BM classes, respectively.

3.5.1. Knowledge utilizing group

Knowledge utilizing group increases utility of the existing patents by acquiring a significant amount of knowledge from existing patents. Simply put, this group acts as knowledge application. The group includes printing detail (e.g., verification of mark) and postage metering system (Cost/Price) as technology classes and technology-based BM classes.

3.5.2. Knowledge disseminating group

Knowledge disseminating group spreads its valuable knowledge that will be widely used by the following patents and acts as knowledge provision. This group includes reloading/recharging and specialized function performed as technology classes, and record keeping, with specific mail handling means, including mailed item weight as technology-based BM classes.

3.5.3. Knowledge utilizing/disseminating group

Knowledge utilizing/disseminating group facilitates knowledge flow by both acquiring and disseminating a significant amount of knowledge. The affiliated classes include data protection, with specific mail handling means, rate updating as technology classes and specialized function performed as technology-based BM class.

4. Implication and conclusion

The study proposed an elaborated approach that explores knowledge flows through technology-based BMs. The proposed approach is integrating patent citation analysis and text mining technique, so that the range and degree of knowledge flows are measured together. In addition, possible problems that may arise when only text mining or citation analysis is used alone are reduced by combining citation analysis and text mining when verifying the degree of knowledge flows between technology-driven BMs and their cited/citing patents. The degree of knowledge measured by the proposed approach can induce the patterns of knowledge flow in detail. The technological classes are categorized into three different groups based on the amount of knowledge they provided to or received from the technology-based BM patents: high knowledge utilization–high knowledge dissemination, high knowledge utilization–low knowledge dissemination, and low knowledge utilization–high knowledge dissemination.

As technology-based BMs have become a patentable subject matter, they played a critical role in knowledge flow. Since ICTs are the integral part of the technology-based BMs, it is often overlooked that most of technological knowledge flow occurs between the ICT classes and the technology-based BMs. It is, however, found that not only the ICT classes but also other general technologies exchange a substantial amount of knowledge with the technology-based BMs. Some technology-based BMs utilize the knowledge from general technologies more than from technology-based BMs as well as disseminate its knowledge to technologies more than to technology-based BMs. In contrast, some technology-based BMs utilize the knowledge from technologies, but its main area of knowledge dissemination is with technology-based BMs. It is also found that there are active knowledge flows between technology-based BMs of different classes.

The proposed approach showed its strengths for handling the unstructured documents in exploring knowledge flows. Moreover, this paper contributed to an improved understanding of the value and function of technology-based BMs and BM patents from the knowledge flow perspective. It also identified the significant sectors in which knowledge is actively exchanged through the technology-based BMs. It is suggested to focus on the development of these sectors to stimulate co-evolution of the technology-based BMs. Since many small and medium companies struggle with lack of necessary skills and knowledge [44], the proposed approach and the classification of knowledge flow drivers can help decision makers to get a comprehensive view of technology-based BMs. Pathways of knowledge dissemination will give an advantage to obtain the benefits of R&D without having to pay its full cost.

However, the proposed approach should be carefully applied to practice. Setting the appropriate level of analysis is one of the important issues. Also, it has to be accounted that the co-word analysis used as text mining is performed on the keywords, and the text is not analyzed directly. Although the bias introduced from keyword analysis is unknown, use of keywords continued to affect the credibility of the technique for years [25,45].

Our case study has some limitations and requires further studies. The scope of BM patent data used in the case study is limited to postage metering system. Data from a wider range should be used to understand the overall system of technology-based BMs. Also, the positioning map only shows a snapshot of the phenomenon. Creating maps for different time periods will present the evolving stages of knowledge flows brought about by the technology-based BMs.

References

[1] S.B. Chang, Using patent analysis to establish technological position: two different strategic approaches, Technol. Forecast. Soc. Chang. 79 (1) (2012) 3–15.

[2] S. Mounteney, European Patent Office update, *IP Europe*, January (2002), 32–8.

[3] NSF, Science and Engineering Indicators, National Science Board, Arlington, VA, 2002. (Retrieved from www.nsf.gov/sbe/srs/seind02/c6/c6s5.htm#top).

[4] A. Osterwalder, The business model ontology: a proposition in a design science approach, Institut d'Informatique et Organisation, 173, University of Lausanne, Ecole des Hautes Etudes Commerciales HEC, Lausanne, Switzerland, 2004.

[5] A.G. Pateli, G.M. Giaglis, Technology innovation-induced business model change: a contingency approach, J. Organ. Chang. Manag. 18 (2) (2005) 167–183.

[6] Y.C.J. Wu, Unlocking the value of business model patents in e-commerce, J. Enterp. Inf. Manag. 18 (1) (2005) 113–130.

[7] L.M. Rausch, International Patenting of Internet-related Business Methods, National Science Foundation, Arlington, VA, March 2003. (Retrieved from http://www.nsf.gov/statistics/infbrief/nsf03314/).

[8] C. Kim, S. Kim, M.S. Kim, Identifying relationships between technology-based services and ICTs: a patent analysis approach. World Academy of Science, Eng. Technol. 5 (2011) 12–22.

[9] C. Lee, H. Park, Y. Park, Keeping abreast of technology-driven business model evolution: a dynamic patent analysis approach, Tech. Anal. Strat. Manag. 25 (5) (2013) 487–505.

[10] R.M. Hunt, You can patent that? Are patents on computer programs and business methods good for the new economy? Bus. Rev. Q1 (2001) 5–15.

[11] S. Wagner, Business method patents in Europe and their strategic use-evidence from franking device manufacturers, Econ. Innovation New Tech. 17 (3) (2008) 173–194.

[12] R. Lukach, J.E. Plasmans, Measuring knowledge spillovers using patent citations: evidence from the Belgian firm's data, CESifo, 2002.

[13] A.G. Hu, A.B. Jaffe, Patent citations and international knowledge flow: the cases of Korea and Taiwan, Int. J. Ind. Organ. 21 (6) (2003) 849–880.

[14] E.L. Glaeser, H.D. Kallal, J.A. Scheinkman, A. Shleifer, *Growth in cities* (No. w3787), National Bureau of Economic Research, 1991.

[15] R. Huggins, A. Johnston, Knowledge flow and inter-firm networks: the influence of network resources, spatial proximity and firm size, Entrepreneurship Reg. Devel. 22 (5) (2010) 457–484.

[16] K. Meagher, M. Rogers, Network density and R&D spillovers, J. Econ. Behav. Organ. 53 (2) (2004) 237–260.

[17] A. Sammarra, L. Biggiero, Heterogeneity and specificity of inter-firm knowledge flows in innovation networks, J. Manag. Stud. 45 (4) (2008) 800–829.

[18] G. Crespi, C. Criscuolo, J.E. Haskel, M. Slaughter, *Productivity growth, knowledge flows, and spillovers* (No. w13959), National Bureau of Economic Research, 2008.

[19] A.J. Nelson, Measuring knowledge spillovers: what patents, licenses and publications reveal about innovation diffusion, Res. Policy 38 (6) (2009) 994–1005.

[20] J.R. Allison, E.H. Tiller, Business method patent myth, Berkeley Tech. LJ. 18 (4) (2003) 987.

[21] E. Duguet, M. MacGarvie, How well do patent citations measure flows of technology? Evidence from French innovation surveys. Econ. Innovation New Tech. 14 (5) (2005) 375–393.

[22] J. Plasmans, R. Lukach, The patterns of inter-firm and inter-industry knowledge flows in the Netherland, http://www.cesifo-group.de/portal/page/portal/DocBase_Content/WP/WP-CESifo_Working_Papers/wp-cesifo-2010/wp-cesifo-2010-05/cesifo1_wp3057.pdf (From).

[23] R. Lukach, J. Plasmans, International knowledge flows from and into a small open economy: patent citation analysis, in: A. Spithoven, P. Teirlinck (Eds.), Beyond Borders: Internationalisation of R&D and Policy Implications for Small Open Economies, Elsevier, Amsterdam, The Netherlands, 2005, pp. 331–357.

[24] A.F.J. van Raan, R.J.W. Tijssen, The neural net of neural network research, Scientometrics 26 (1) (1993) 169–192.

[25] L. Leydesdorff, Words and co-words as indicators of intellectual organization, Res. Policy 18 (4) (1989) 209–223.

[26] C. Choi, Y. Park, Monitoring the organic structure of technology based on the patent development paths, Technol. Forecast. Soc. Chang. 76 (6) (2009) 754–768.

[27] H.J. No, Y. Park, Trajectory patterns of technology fusion: trend analysis and taxonomical grouping in nanobiotechnology, Technol. Forecast. Soc. Chang. 77 (1) (2010) 63–75.

[28] I.V. Wartburg, T. Teichert, K. Rost, Inventive progress measured by multi-stage patent citation analysis, Res. Policy 34 (10) (2005) 1591–1607.

[29] X. Wang, Y. Zhao, R. Liu, J. Zhang, Knowledge-transfer analysis based on co-citation clustering, Scientometrics 97 (3) (2013) 859–869.

[30] Information Retrieval: Data Structures & Algorithms, in: W.B. Frakes, R. A. Baeza-Yates (Eds.), Prentice-Hall, 1992.

[31] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.

[32] G. Salton, M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1986.

[33] F. Janssens, Clustering of scientific fields by integrating text mining and bibliometircs, (Ph.D. thesis,) Faculty of Engineering, Katholieke Universiteit Leuven, Belgium, 2007. (Retrived from http://hdl. Handle.net/1979/847).

[34] P. Calado, M. Cristo, M.A. Goncalves, E.S. De Moura, B. Riberiro-Neto, N. Ziviani, Link-based similarity measures for the classification of web documents, JASIST 57 (2006) 208–221.

[35] F. Janssens, V. Tran Quoc, W. Glanzel, B. De Moor, Integration of textual content and link information for accurate clustering of science fields, in: V.P. Guerrero-Bote (Ed.), Proc. Of the Intl. Conf. on Multidisciplinary Information Saicence and Technolgies (InSciT2006), Merida, Spain., 2006, pp. 615–619.

[36] P. Glenisson, W. Glanzel, F. Janssens, B. De Moor, Combining full text and bibliometric information in mapping scientific disciplines, Inf. Process. Manag. 41 (2005) 1548–1572.

[37] F. Janssens, W. Glanzel, B. de Moor, A hybrid mapping of information science, Scientometrics 75 (3) (2003) 607–631.

[38] J. Zhang, E.M. Rasmussen, Developing a new similarity measure from two different perspectives, Inf. Process. Manag. 37 (2) (2001) 279–294.

[39] M. McGill, M. Koll, T. Noreault, An evaluation of factors affecting document ranking by information retrieval systems, School of Information Studies, Syracuse University, Syracuse, NY, 1978.

[40] J. Zhang, R. Korfhage, A distance and angle similarity measure, J. Am. Soc. Inf. Sci. 50 (9) (1990) 772–778.

[41] M. Trajtenberg, R. Henderson, A. Jaffe, University versus corporate patents: a window on the basicness of invention, Patents, Citations and Innovations—A Window on Knowledge Economy, MIT Press, Cambridge, MA, 2002.

[42] J.L. Priego, A vector space model as a methodological approach to the triple helix dimensionality: a comparative study of biology and biomedicine

centres of two European national councils from a webometric view, Scientometrics 58 (2) (2003) 429–443.
[43] W.P. Jones, G.W. Furnas, Pictures of relevance: a geometric analysis of similarity measures, J. Am. Soc. Inf. Sci. 38 (1987) 420–442.
[44] E.E. Grandon, J.M. Pearson, Electronic commerce adoption: an empirical study of small and medium US businesses, Inf. Manag. 42 (1) (2004) 197–216.
[45] P.H. Healey, P. Hoch Rothman, An experiment in science mapping for research planning, Res. Policy 15 (5) (1986) 233–251.

Hyun Joung No Mailing address: 38-205, Department of Mechanical and Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-744, Republic of KoreaE-mail address: anesno1@snu.ac.kr
Full international contact details: tel) +82-2-880-1670, fax) +82-2-883-1546
Brief professional biography: Hyun Joung No is currently an endowed assistant professor at the Seoul National University (SNU) in the Department of Mechanical and Aerospace Engineering. She holds a Ph.D. and MS in technology management from SNU and a BS in materials engineering from University of Toronto, Canada. Her primary research interests are technology evolution and technology management, and include developing forecasting methodologies for technological convergence and emerging technologies.

Yoonjung An Department of Industrial Engineering, School of Engineering, Seoul National University, Seoul, Republic of Korea
Brief professional biography: Yoonjung An is a Ph.D. student in the Department of Industrial Engineering at Seoul National University. She received a master degree in Technology Management Economics and Policy Program from Seoul National University and a bachelor degree in Civil Engineering from University of Toronto. Her main research interests lie in the fields of strategic management of knowledge and innovation.

Yongtae Park Mailing address: 39-336, Department of Industrial Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-744, Republic of Korea
E-mail address: parkyt1@snu.ac.kr
Full international contact details: tel) +82-2-878-8358, fax) +82-2-878-8560
Brief professional biography: Yongtae Park is a professor of industrial engineering at the Seoul National University (SNU). He holds BS in industrial engineering from SNU, MS and Ph.D. in operations management, both from the University of Wisconsin-Madison, USA. He has authored many articles on technology management and service engineering published in numerous journals and, in 2009, he received the Award for Top 50 Researchers in Technology Innovation Management by the International Association of Management of Technology (IAMOT) and Elsevier.