

Article

Exploring Promising Research Frontiers Based on Knowledge Maps in the Solar Cell Technology Field

Inchae Park, Keeun Lee and Byungun Yoon *

Department of Industrial & Systems Engineering, Dongguk University-Seoul, 30, Pildong-ro 1 gil, Jung-gu, Seoul 100-715, Korea; E-Mails: inchae@dongguk.edu (I.P.); kelee@dongguk.edu (K.L.)

* Author to whom correspondence should be addressed; E-Mail: postman3@dongguk.edu; Tel: +82-2-2260-8659; Fax: +82-2-2269-2212.

Academic Editor: Andrew Kusiak

Received: 24 July 2015 / Accepted: 1 October 2015 / Published: 9 October 2015

Abstract: Given the challenging environmental issues in the energy sector, the importance of strategic research and development (R&D) planning has been emphasized to manage a turbulent business situation. This study aims to propose a methodology for exploring promising research frontiers in the energy sector. To this end, first, core documents are collected from scientific documents such as patents and academic papers. Second, the research frontiers are extracted by clustering the core documents based on the bibliographic relations. Third, a knowledge map is generated by mapping the relations between research frontiers. Finally, the promising research frontiers (RFs) are identified by conducting dynamic analyses and the contents of promising RFs are suggested. As an illustration of the method, the field of solar cell technology is chosen and analyzed. As a result, the promising research frontiers from the patent knowledge map are related to development (D) themes and promising research frontiers from scientific paper knowledge map are related to the research (R) themes. The proposed method and results can be utilized by researchers, R&D policy makers, and administrations in practice.

Keywords: bibliographic analysis; patent map; research frontier; solar cell technology

1. Introduction

Recently, many challenging environmental issues such as the increase in energy demands, the exhaustion of fossil fuel and climate change have emerged, making many countries pay attentions on

various influential policies and R&D efforts. In particular, recently, shale gas has become a critical substitute source of natural gas, enabling the structure of power in the energy industry to radically be reorganized. Although the development of shale gas was originated from the United States, interests have spread to potential shale gas in the rest of the world. To promptly cope with such rapidly changing situations, strategic research and development (R&D) planning is significantly important because sophisticated technology can realize the needs of such energy demands. Basically, new sources of energy cannot be extracted and developed with traditional technologies that are unable to meet totally different technological needs of them. To this end, detection of promising and emerging research themes in energy technology is regarded as a priority task. However, most of the strategies for development of energy technology have relied on the consideration of national policies and qualitative analyses without a systematic process or methodology. Thus, such R&D processes are subject to yield rough research themes based on intuitions of policy-makers and technologists.

Though some researches utilize quantitative approaches such as bibliometric analyses of patents and publications in energy technology, most of the studies suggest simple statistical graphs and analyze the status quo of technologies [1–4]. The sector of energy technology inevitably needs technology forecasting because a lot of energy issues are global-wide problems that all countries should collaborate to solve for the future of human being. However, fruitful databases of patents and academic papers have been intensively utilized to anticipate promising energy technology. Thus, a systematic approach based on a knowledge map can be a solution to overcome the limitations of previous researches on bibliometric analysis. As one of the knowledge maps, a patent map is able to be utilized for various functional areas. As Yoon *et al.* [5] suggested for the usage of patent map, it can be used to: (1) identify industry or specific technology trends and understand competitors' technology assets and strategy for administration area; (2) select new product themes, understand portfolios of technology assets, and identify technology vacuums in the R&D area; (3) identify portfolios of R&D human resources in the personnel area; and (4) identify patent infringements and investigate life cycle and scope of patents in the patent management area.

The present research aims to propose a quantitative methodology identifying promising research frontiers of energy technology. A knowledge map of solar cell technology as an illustration of the proposed approach is developed by visualizing research frontiers that are derived from bibliographic information from patents and scientific articles. The hybrid approach of network analysis and principal component analysis is utilized as a visualization technique of a knowledge map based on bibliographic coupling. Promising research frontiers from both patent and scientific paper knowledge maps are suggested by conducting dynamic analysis. Furthermore, practical contents of promising research frontiers are suggested by conducting text-mining technique.

The study is structured as follows. Section 2 briefly explains the relevant studies and describes the research methodology. Based on the methodology, Section 3 illustrates the proposed approach and shows the results of analysis, using solar cell technology and Section 4 describes implications and discussions of the results. Finally, Section 5 draws conclusions and policy implications, discussing the contributions and limitations of this study.

2. Methods

2.1. Background

2.1.1. Research Frontier

The concept of a research front was originally introduced by de Solla Price [6]. According to Price, research fronts are research domains under development where papers cite each other densely. Analysis of research frontiers, so called research fronts, was originated by pioneers of bibliometric citation analysis [7]. The most highly cited scientific papers were defined as a “research frontier” since the number of citations can be interpreted as signs of their importance as sources of new knowledge [8,9]. As a similar concept, Garfield and Small [10] identified “hot field” clusters of highly cited papers within three years from publication.

Quantitative methods have been applied to identify and track the research frontiers. Small and Griffith [11] represented scientific specialty by computing co-citation strength between pairs of documents. Braam *et al.* [12] analyzed the frequency of indexing terms and classification codes co-occurring in the publications. Clustering methods were improved with co-word based map [13,14]. Persson [15] defined research frontiers by developing a research front map based on citation-based bibliographic coupling. Fujita *et al.* [16] detected research fronts and promising fields by division into clusters in combinational citation networks. Shibata *et al.* [17] developed a method of detecting emerging research fronts based on topological measures in citation networks of scientific publications. The studies on comparing the performance of cluster solutions such as co-citation, bibliographic coupling, direct citation, and hybrid approaches have been conducted in order to improve research fronts mapping performance [18–20]. Lucio-Arias and Leydesdorff [21] utilized title words, cited references, and sequence numbers as an indicator of research front activity. Saka *et al.* [22] analyzed the dynamic changes and trends of scientific research fronts using the database of articles with high citation rates among the published articles. Toivanen [7] analyzed the evolution of Brazilian research frontiers with bibliometric methods.

2.1.2. Bibliometric Analysis

A bibliometric analysis refers to the research methodology utilized in information sciences, employing quantitative analysis and statistics to describe the distribution patterns of articles within a given topic, field, institution, or country [23]. Bibliometric methods can be applied to assess the scientific outputs or research patterns of authors, institutes and countries. In addition, this analysis can contribute to the progress of science and technology in many ways by allowing analysts to analyze international collaborations, laying the academic foundation for the evaluation of new developments, identifying major scientific actors, performing technology forecasting and so on. Two main methods of bibliometric analysis are science/technology mapping and performance analysis. The purpose of science/technology mapping is to visualize the conceptual structure of research and technology, showing the evolution of them. However, the performance analysis aims at evaluating the citation impact of the scientific/technological production of different actors.

Thus, it is related to various fields such as bibliometrics, scientometrics, informetrics and so on. These three terms have so considerable fuzzy borderlines that these are used as almost similar concept. Before the terminologies were used, Cole and Eales [24] conducted first bibliometric study although they used terminology of “statistical bibliography”. Hulme [25] conducted further work using patents in order to measure social progress in Britain. Lotka [26] studied frequency distribution of scientific productivity using the decennial indices of *Chemical Abstract*. Gross and Gross [27] conducted their citation-based study in order to support the decision which chemistry periodicals should best purchased by small college libraries. Zipf [28] formulated a law on word frequency or occurrences.

Many following researchers [29–35] have claimed the term “bibliometrics” is coined by Pritchard [36]. Pritchard explained the term *bibliometrics* as “the application of mathematical and statistical methods to books and other media of communication”. The term *scientometrics*, which is Russian equivalent of the term “naukometriya”, were almost simultaneously introduced by Nalimov and Mul’chenko [37]. They defined scientometrics as “the application of those quantitative methods which are dealing with the analysis of science viewed as an information process”. The term *Informetrics*, which comes from German term “informetrie”, was proposed by Nacke [38] to cover the part of information science.

Hood and Wilson [39] explained the differences among the terminologies. Bibliometrics deal with general information process and focus on the literature itself of science and scholarship, while scientometrics is restricted to the measurement of science communication. However, it offers more information on science and technology after measuring and analyzing literature output; for example, socio-organizational structures, governmental policies toward science and technology and so on. Informetrics stands for a more general subset of information science dealing with mathematical statistical analysis of communication process in science and also deals with electronic media.

2.1.3. Knowledge Map

A knowledge map or mapping is generally a representative tool or technique concerning “knowledge about knowledge” rather than knowledge itself [40–42]. A knowledge map is identified in terms of various perspectives as: (1) a visualization tool that enables the identification of the competence of a firm by analyzing the knowledge portfolio for knowledge manager [40]; (2) a type of mental diagram that allows the complex idea to be arrayed in a logical sequence for the knowledge management system developer [43]; and (3) a spatial arrangement plan representing inclusion/dependence relationships among knowledge for the knowledge user [44]. Knowledge maps are classified as hierarchical/radial knowledge maps, networked knowledge maps, knowledge source maps, and knowledge flow maps, in terms of their structure [45].

Studies on knowledge maps utilizing science and technology information such as patents and scientific papers are widely conducted with various methods. Traditional science and technology knowledge maps suggested statistical graphs on papers, textbooks, and reports [36]. The range of mapping data on knowledge maps has been extended to citations, indices, classifications, authors, *etc.*, in terms of bibliographic information. Textual information of scientific documents such as patents and papers is also utilized to generate knowledge maps. Thus, various types of knowledge maps on science and technology are visualized by diverse visualization techniques such as matrix, network analysis,

multi-dimensional scaling (MDS), self-organizing mapping (SOM), and generative topological mapping (GTM).

In terms of studies on knowledge maps with scientific papers, Persson [15] suggested a research front map based on bibliographic coupling with articles published in *Journal of the American Society for Information Science (JASIS)*. White and McCain [33] utilized co-citation analysis and MDS to visualize academic research areas. Van Raan [46] generated inter-journal bibliometric maps in order to track how scientific knowledge has developed, centering on the interdisciplinary relationship. Boyack *et al.* [47] presented a knowledge map and examined differences in subject areas by comparing co-citation and the inter-citation analysis. Medina and Leeuwen [48] presented a network map that comprises the most deeply related journals of the seed journal. Su and Lee [49] made a survey by generating a keyword-based network with the articles published in *Research Policy*.

In terms of studies of knowledge maps of patents, Shin and Park [50] generated a patent claim map using text-mining and network analysis. Yoon, Yoon and Park [5] suggested technology vacuum maps, claim point maps, and technology portfolio maps utilizing the SOM method. Segev and Kantola [51] identified patent trends by presenting knowledge maps of patents by SOM. Lee *et al.* [52] discovered new technology opportunities by conducting text-mining and principal component analysis. Son *et al.* [53] developed a GTM-based patent map to automatically identify patent vacuums.

In terms of studies of knowledge map of other sources, Yoon *et al.* [54] developed a core R&D map, R&D trend map, R&D concentration map, R&D relation map, and R&D cluster map using a research proposal database. Chen and Chen [55] and Chen [56] established a design patent map with patent examiners. Kim and Park [57] proposed user-centric service map to identify a new service opportunity from potential needs.

2.2. Research Methodology

2.2.1. Research Concept

This research aims to suggest promising research frontiers by generating knowledge maps of patent and scientific paper. Basically, promising research areas have potentials that they provide significant influences to subsequent patents and papers. Such impacts can be measured by the cited frequency in the citation relationship. In particular, since research frontiers need to cover academic as well as practical worlds, the proposed approach uses two different databases of papers and patents. The results of analysis on two data will be complementarily interpreted to investigate the possibility of promising research themes. The overall research concept is shown as Figure 1. Bibliographic information from the patent or scientific paper database is collected in the research area. Core documents which are highly cited by other documents are extracted from the collected documents. The core documents are clustered to identify research frontiers. The research frontiers are visualized as a knowledge map. In the visualization process, emerging or enhanced research frontiers are identified by conducting dynamic analysis. Those emerging or enhanced research frontiers are identified as promising research frontiers for the future. The related keywords of the promising research frontiers are finally suggested as promising contents.

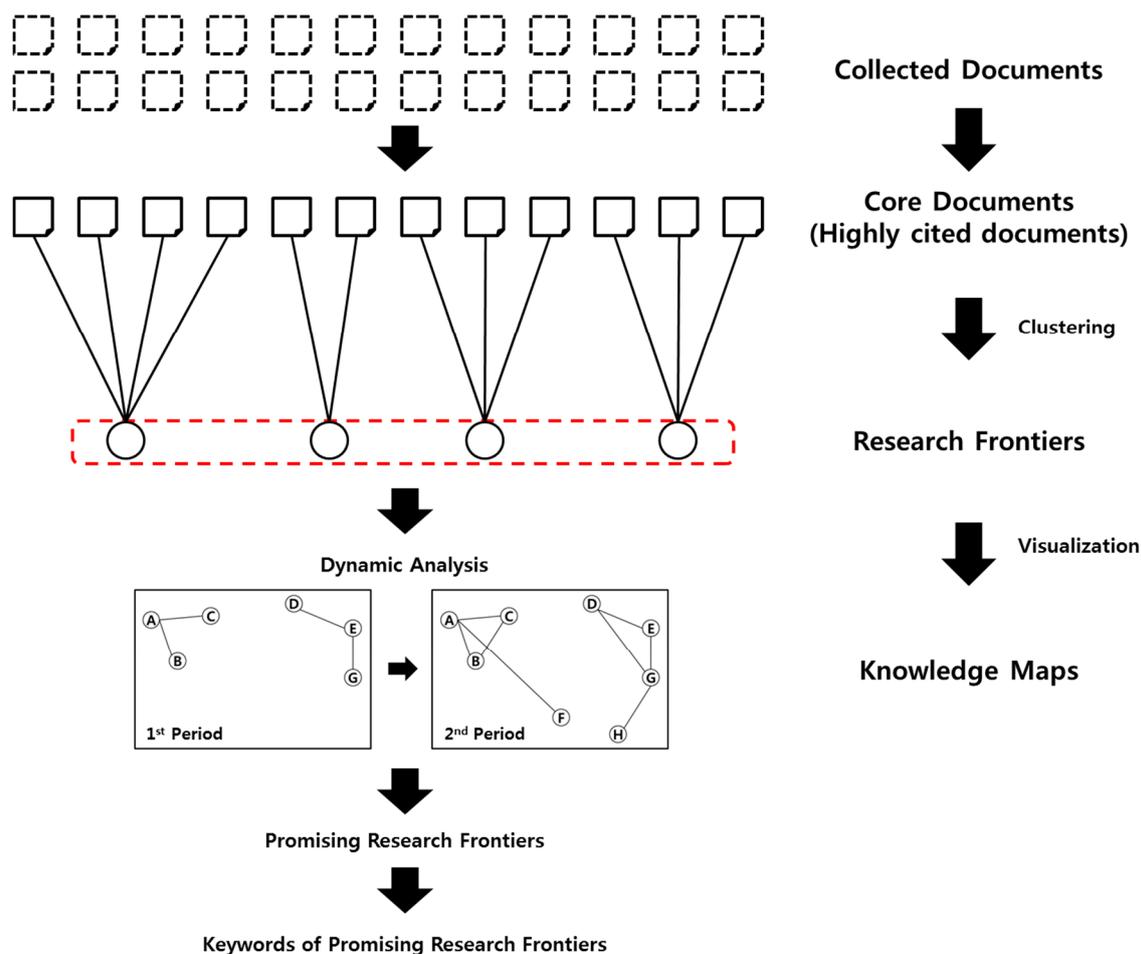


Figure 1. Research concept.

2.2.2. Research Framework

The overall research process to explore promising research areas by developing knowledge maps consists of several steps, as shown in Figure 2. The first step is the data collection from the patent database or scientific paper database because two databases are main data sources by covering both practical and academic research activities. Second, the core documents are extracted based on the number of citations since the highly cited documents can be considered as core data sources which contain high quality technological information. Third, research frontiers are identified by conducting k-means clustering with core documents. Similar documents that are correlated in terms of technological contents can be grouped to investigate trends of influential technologies. Fourth, the research frontiers are visualized as knowledge maps by conducting a hybrid approach of principal component analysis and network analysis. Fifth, promising research frontiers are determined by conducting a dynamic analysis which is a comparison process of knowledge maps between the first and second periods. If significant changes can be examined by the comparisons over time, promising research themes can be and drawn and highlighted. Finally, promising contents are suggested by conducting text-mining from the related documents in promising research frontiers because concrete research activities can be pursued on the basis of research keywords that are extracted by the text-mining.

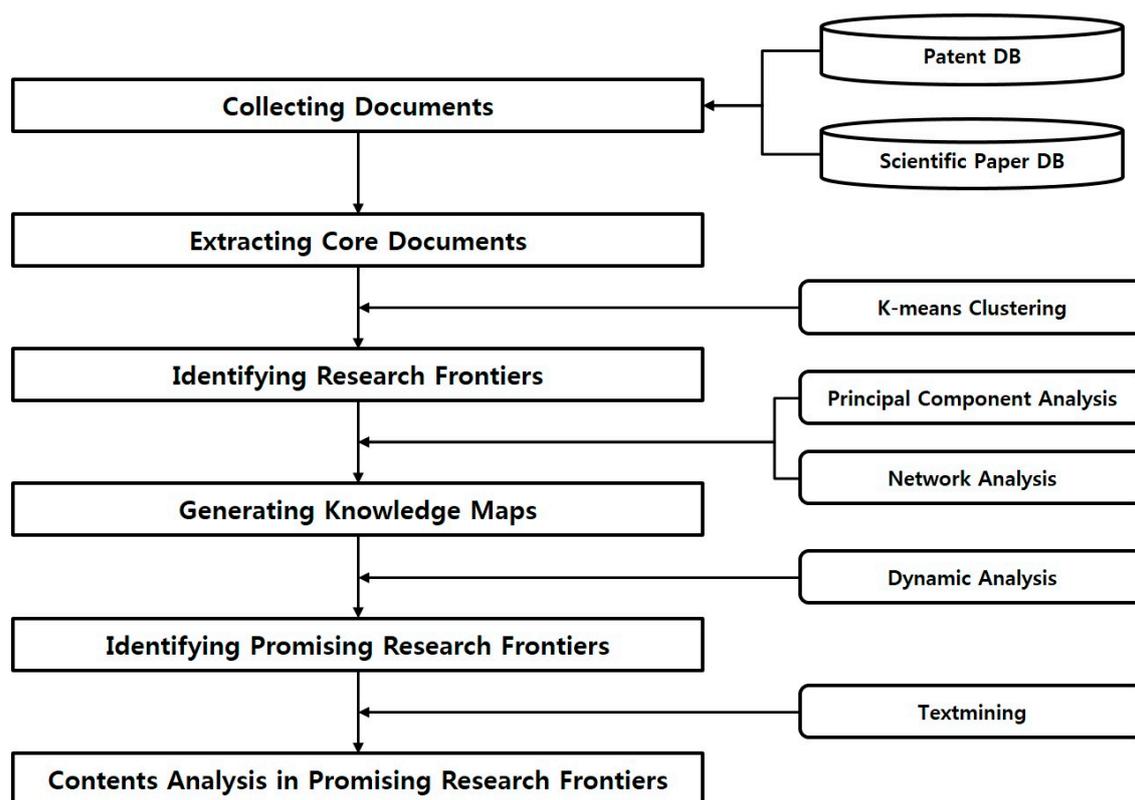


Figure 2. Research process.

2.2.3. Detailed Procedures

Data Collection

First of all, a technology field and scope of analysis such as publication year and types of documents for analyzing are identified. Afterward, a search formula is built to collect data. When building the search formula, domain experts' participation or utilization of technology classification table makes for better search results when collecting document data. After the search process, bibliographic information from a reliable patent or scientific paper database such as the United States Patent and Trademark Office (USPTO) database or the web of science (WoS) database is collected. The bibliographic information contains title, abstract, the number of citations, citing reference list. The collected raw data is refined by eliminating noise which is irrelevant in the documents data or have missing values.

Core Documents Extraction

Core patent and scientific papers exert so important an influence and impact in the technology field that core patent identification is highly necessary. In this research, core documents are identified based on the number of citations. Highly cited documents are extracted as core documents since the documents can be considered as documents that contain valuable technological information. In this process, suitable thresholds are selected to extract core patents and scientific papers since the characteristics of each database are different. The previous research on identifying hot research areas by bibliometric methods conducted by National Institute of Science and Technology Policy (NISTEP) utilized 1% of the threshold value based on the number of citations to extract core papers [22]. However, the threshold for a patent

database should be higher than that of the papers since the number of scientific paper publications and citation is much more than that of patent registration and citation in general. Furthermore, visualization of data should be considered since a map with too much data could not be comprehended intuitively when choosing the threshold.

Research Frontiers Identification

The extracted core documents are clustered by k-means clustering—which is popular for cluster analysis in data mining—to identify the research frontiers (RFs) that are the leading research themes in the technology area. RFs which are extracted from the patent data are technical RFs and those from the scientific paper data are academic RFs. The core documents are clustered based on the reference information; that is to say the documents that cite the same references are similar to each other. The data format for k-means clustering in this step is like Table 1. In this step, a suitable k should be selected by sensitive analysis. After cluster analysis, the extracted RFs are named by domain experts after reviewing the titles and abstracts of documents.

Table 1. Data format for K-means clustering.

	Ref. 1	Ref. 2	...	Ref. n – 1	Ref. n
Doc. 1	1	1	...	0	0
Doc. 2	0	0	...	1	1
...
Doc. n – 1	0	1	...	0	0
Doc. n	0	1	...	0	1

Knowledge Maps Generation

Relations among RFs are visualized as two types of network formatted knowledge maps: a patent knowledge map and a scientific paper knowledge map. The proposed visualizing method is a hybrid approach that is a combination of network analysis and principal component analysis. In the proposed network, the node is RF and link is the bibliographic coupling relation, which means the similarity between documents based on the number of common references. Furthermore, the proposed knowledge map provides position information additionally based on the first two principal component scores by conducting principal component analysis (PCA). PCA is conducted with the data formatted as in Table 2 to present RF's position. Thus, the X-axis is the first principal component score and Y-axis is the second principal component score. The node position indicates the similarity relation between one and other RFs in a whole map while the link indicates a one-on-one similarity relation between two RFs. For example, nodes are positioned based on the principal component scores in Figure 3. The node A or G is similar to B and C or to F and H, respectively, when it comes to a whole network. However, the node A is linked to node G though A is far from G since the two nodes are strongly bibliographically coupled.

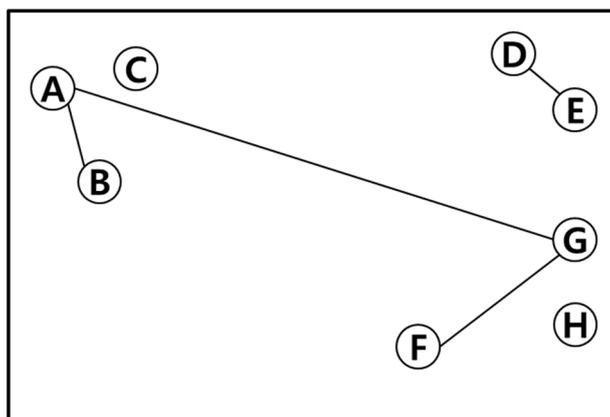


Figure 3. Concept of knowledge map.

Table 2. Data format for principal component analysis.

	Ref. 1	Ref. 2	...	Ref. n – 1	Ref. n
RF 1	2	0	...	10	0
RF 2	0	0	...	3	1
...
RF n – 1	0	1	...	0	3
RF n	0	12	...	9	4

Table 3. Original bibliographic coupling strength matrix.

	Doc. 1	Doc. 2	...	Doc. n – 1	Doc. n
Doc. 1	2	0	...	10	0
Doc. 2	0	5	...	1	12
...
Doc. n – 1	10	1	...	12	9
Doc. n	0	12	...	9	4

Bibliographic coupling strength should be calculated in order to visualize the one-on-one similarity relation between RFs. To this end, the original bibliographic coupling strength, which is defined as the number of common references, is calculated as a matrix as in Table 3. However, the original bibliographic coupling strength should be normalized since the documents have different numbers of references. The normalized coupling strength (NCS) is defined as [58]

$$NCS_{ij} = \frac{r_{ij}}{\sqrt{n_i n_j}} \tag{1}$$

where NCS_{ij} is the normalized coupling strength between document i and j , r_{ij} is the number of references common to both i and j , n_i is the number of references in the reference list of document i , and n_j is the number of references in the reference list of document j . The range of the NCS value is zero to one. The original coupling strength matrix between documents is transformed to the normalized coupling strength matrix between documents. The normalized coupling strength matrix between documents is finally transformed by calculating an average value of normalized coupling strength between RFs like Table 4. The RFs are visualized based on the average normalized coupling strength as nodes in the map. The

coupling relation between RFs is visualized as a link with a suitable threshold value. In addition, the core research frontier (RF) which has high degree centrality value in the knowledge network can be identified.

Table 4. Average normalized coupling strength matrix.

	RF 1	RF 2	...	RF n – 1	RF n
RF 1	1	0.56	...	0.76	0.32
RF 2	0.56	1	...	0.44	0.67
...
RF n – 1	0.76	0.44	...	1	0.11
RF n	0.32	0.67	...	0.11	1

Promising Research Frontiers Identification

A promising RF is identified on the basis of the results from dynamic analysis which is a comparison process of the number of core patents in two periods. To this end, the data are split into two time period. This paper defined two types of promising RFs: emerging and enhanced. First, an emerging RF is an RF that is newly emerged since the core documents do not exist in the first time period and emerge in the second period. Second, an enhanced RF is the RF for which the number of core documents in the RF increases in the second period. Thus, the promising RFs are the RFs that are newly emerged or enhanced as the core documents are newly emerged and increased in the dynamic analysis. In addition, the relation between core RFs and promising RFs can be observed to investigate and compare the current important RFs and future important RFs.

Contents Analysis in Promising Research Frontiers

Keywords can play a critical role in coming up with innovative research ideas because the identification of relevant keywords is the first step for idea generation. Thus, this research utilizes the text-mining approach to extract the contents of promising RFs. Since it is difficult to derive promising RFs keywords from a few core documents in the promising RF, promising RF's relevant documents should be collected again. To this end, first, a search formula of promising RFs is developed to collect relevant patent or paper data. Second, data is refined by eliminating noise because the search results may include irrelevant documents. Although this process can be automatically performed, the intervention of experts is necessary for better performance of keyword analysis. Finally, promising keywords are extracted based on the high frequent keywords from the abstract of collected documents by text-mining. The suggested promising keywords can be utilized as the contents of promising RFs.

3. Results

3.1. Data Collection and Core Documents Extraction

Solar cell power generation is one of the strongest technologies for applications of solar energy. The governments of many countries have promoted the solar cell industry to solve energy problems, allowing the industry to become one of the fastest growing new industries. A solar cell can be defined as an

electrical device that converts the energy of light directly into electricity by the photovoltaic effect. Energy harvesting powered devices have the potential for widespread use in buildings as sensors in building management systems. Thus, the solar cell technology is one of critical research areas in the energy sector, having industrial impacts in terms of market size as well as technology diffusion. Patent data on the solar cell technology that were granted from 2008 to 2012 is collected from the four patent offices: Unites States (US), Europe (EP), Korea (KR), and Japan (JP). To collect the patent data, patent search queries are built by research areas with domain experts' participation from Korea Institute of Energy Research (KIER) shown in Table A1. A technology classification related to the solar cell technology is utilized in order to collect the relevant patent data. The citation frequency of 1668 patents is distributed as shown in Table 5. The highest cited frequency is 91 and 956 patents have not been cited. The 175 patents among 1668 patents which are cited more than five times and occupy approximately 10% are selected as core patents since these are considered as highly cited patents of the total patent data set (We conducted *t*-test with self-citation rate of selected patents to verify whether the selection of core patents is biased. The references citing the selected patents are additionally collected and the patents are split into two groups which are top 5% cited of patent group (group 1) and the patents group (group 2) which is occupied from 5% to 10%. As a result, the difference of self-citation rate between group 1 and group 2 is not statistically significant (p -value = 0.152 > 0.05). The result implies that the self-citation may not lead to serious biased results on the selection of core patents. Results: Group 1 (N = 82, mean = 0.3571, S.D. = 0.3989); Group 2 (N = 93, mean = 0.2750, S.D. = 0.3493)). In this research, 174 patents are analyzed and one patent is excluded for having no citing reference.

Table 5. Patent citation frequency distribution.

Citation Frequency	Number of Patent	Cumulative Number of Patent	Rate	Cumulative Rate
91	1	1	0.06%	0.06%
72	1	2	0.06%	0.12%
62	1	3	0.06%	0.18%
.....
.....
5	35	175	2.10%	10.49%
4	52	227	3.12%	13.61%
3	70	297	4.20%	17.81%
2	146	443	8.75%	26.56%
1	269	712	16.13%	42.69%
0	956	1668	57.31%	100.00%
Total	1668		100.00%	

The solar cell technology relevant scientific paper data that was published from 2008 to 2012 and indexed in SCI-Expanded or SSCI is collected from the Web of Science database. Paper search query is composed of combination of keywords “solar”, “photovoltaic”, and “cell*” in the “title” category. The citation frequency of 13,162 papers is distributed as Table 6. The 131 patents which are cited more than 194 times and occupy about 1% are selected as core papers since these are considered as highly cited

papers of the total paper data set. In this research, 130 papers are analyzed and one paper includes insufficient information.

Table 6. Scientific paper citation frequency distribution.

Citation Frequency	Number of Paper	Cumulative Number of Paper	Rate	Cumulative Rate
1686	1	1	0.01%	0.01%
1675	1	2	0.01%	0.02%
.....
.....
194	1	131	0.01%	1.00%
.....
.....
2	1072	9836	8.14%	74.73%
1	1244	11,080	9.45%	84.18%
0	2082	13,162	15.82%	100.00%
Total	13,162		100.00%	

The threshold to select a technology relevant core document is different since the properties of patents and papers are different. The number of published papers is much greater than that of granted patents and the number of citations of patent is much less than that of scientific paper publication since patent citation is related to complicated business situations.

3.2. Research Frontiers

The solar cell technology relevant 174 core patents are clustered by conducting k-means clustering. The 45 clusters are extracted by sensitive analysis with energy technology domain experts. These 45 clusters are named by domain experts after reviewing the title and abstract of the patents. The 45 clusters are identified as research frontiers (RFs) since these consist of highly cited core patents and imply research themes. The 45 RFs are matched to the research areas (RAs) which are the highest class level in the solar cell technology classification which is utilized when collecting relevant documents (see Table 7). Though several RFs have only one core patent, the core patents can stand for research theme because the core patents are extracted as 10% highly cited patents among all collected patents. In addition, the results can be guaranteed because the RFs are identified by domain experts after clustering analysis.

Table 7. Research frontiers from core patents.

Research Area (RA)	RF No.	Research Frontier (RF)	Core Patent Frequency
Crystalline silicon solar cell	RF2	Back contact-type solar cell	1
	RF3	Back side contact solar cell	1
	RF7	Contact for a photovoltaic cell	1
	RF9	Crystal silicon solar cell	1
	RF14	Hetero-junction crystalline silicon solar cell	2
	RF20	Metal contact structure for solar cell	1

Table 7. Cont.

Research Area (RA)	RF No.	Research Frontier (RF)	Core Patent Frequency
Crystalline silicon solar cell	RF25	Poly silicon ingot	1
	RF27	Process and system for semiconductor wafer	1
	RF30	Single crystal silicon solar cell	5
	RF32	Solar cell contact formation	1
	RF39	Texturing silicon surface in solar cell	1
	RF45	High-efficiency crystalline silicon solar cell	2
Thin film solar cell	RF4	Bifacial thin film solar cell	1
	RF5	Bulkheterojunction thin film solar cell	1
	RF6	CIGS based thin film	10
	RF8	Contact layers for thin film solar cells	2
	RF13	Forming crystalline layer	1
	RF15	Integrated thin-film solar cell	2
	RF17	Large-scale production of thin-film solar cells	2
	RF19	Material for thin film photovoltaic cell fabrication	2
	RF21	Multijunction thin film solar cell	3
	RF24	Photovoltaic module architecture	7
	RF29	Silicon based thin film solar cell	2
	RF37	Tandem-type thin film solar cell	2
	RF41	Thin film solar cell manufacturing	21
	RF42	Thin film solar cell module	2
	RF43	Three-dimensional thin-film solar cells	6
	RF44	Treatment of thin film photovoltaic cell	8
Organic/ quantum dot/nano convergent solar cell	RF1	A process for producing metal oxide thin films	1
	RF10	Dye-sensitized solar cell	14
	RF22	Nano- solar cells	7
	RF23	Organic solar cell	15
	RF28	Quantum dot solar cell	9
	RF40	Thin film of organometallic compound	1
The other solar cell	RF11	Electrode for photovoltaic cells	9
	RF12	Forming a tiled substrate structure	1
	RF16	Laminated solar battery	1
	RF18	Manufacturing method for solar modules	11
	RF26	Power generation system	1
	RF31	solar arrays	3
	RF33	Substrate and collector grid structures	5
	RF34	Substrate for transparent electrodes	1
	RF35	Sunlight detecting system	1
	RF36	System for cleaning surface of solar cell	1
RF38	Terminal box	3	

The solar cell technology relevant 130 core scientific papers are assigned to 45 predefined research frontiers which are derived from core patents by domain experts. However, four research frontiers are newly added since some of the core scientific papers are not assigned to the existing 45 research frontiers. The RFs from core scientific papers consist of 11 existing RFs and four newly added RFs which are

polymer solar cell (RF46), plasmonic solar cell (RF47), mesoscopic solar cell (RF48), and solar water splitting (RF49), as shown in Table 8.

Table 8. Research frontiers from core scientific papers.

Research Area (RA)	RF No.	Research Frontier (RF)	Core Paper Frequency
Crystalline silicon solar cell	RF30	Single crystal silicon solar cell	1
	RF45	High-efficiency crystalline silicon solar cell	8
Thin film solar cell	RF5	Bulkheterojunction thin film solar cell	17
	RF19	Material for thin film photovoltaic cell fabrication	1
	RF37	Tandem-type thin film solar cell	2
	RF41	Thin film solar cell manufacturing	3
Organic/quantum dot/nano convergent solar cell	RF10	Dye-sensitized solar cell	24
	RF22	Nano- solar cells	15
	RF23	Organic solar cell	13
	RF28	Quantum dot solar cell	5
The other solar cell	RF11	Electrode for photovoltaic cells	1
	RF46	Polymer solar cell	34
	RF47	Plasmonic solar cell	4
	RF48	Mesoscopic solar cell	1
	RF49	Solar water splitting	1

RAs which are thin film solar cells and organic/quantum dot/nano convergent solar cells account for a high portion of core patent data. There are many core patents in CIGS based thin film (RF6) and thin film solar cell manufacturing (RF41) of thin film solar cell RA, dye-sensitized solar cell (RF10) and organic solar cell (RF23) of organic/quantum dot/nano convergent solar cell RA, and manufacturing methods for solar modules (RF18) of the other solar cell RA. However, RAs which are organic/quantum dot/nano convergent solar cell and other solar cell account for a high portion of core scientific paper data. There are many core scientific papers in bulk heterojunction thin film solar cells (RF5) of thin film solar cell area and dye-sensitized solar cell (RF10) of organic/quantum dot/nano convergent solar cell RA, and polymer solar cell (RF46) of the other solar cell RA.

3.3. Knowledge Maps

Two types of knowledge maps from patent and scientific papers are generated by conducting principal component analysis (PCA) and network analysis. The results imply the position on the knowledge map and Tables A2 and A3 show the results of PCA from the patent and scientific paper research frontiers. The position among the respective RFs depicts the relationship in the whole knowledge map.

Furthermore, the RFs of patents and scientific papers are visualized based on the bibliographic coupling relation by network analysis (see Figures 4 and 5). The links of the knowledge maps visualize the one-on-one relationship by linking normalized bibliographic coupling strength between nodes that are greater than the threshold value even if the nodes are far from each other with respect to the position. The threshold value is 0.02 which is selected by conducting sensitive analysis to visualize the relation among the research frontiers. The core research frontiers which have a high degree of centrality values in

the maps can be derived from maps, as summarized in Tables 9 and 10. The core research frontier nodes are colored red in the maps. Most of the core RFs which are derived from patent knowledge maps are related to structure of the solar cell such as RF33, 3, 8, 14, 20, and 30 while RF6 is related to material and RF44 is related to methods. However, core RFs derived from the scientific paper knowledge map consist of two attribute-related RFs such as RF23 and 46, two structure-related RFs such as RF5 and 37, a method-related RF such as RF45, and a material-related RF such as RF19.

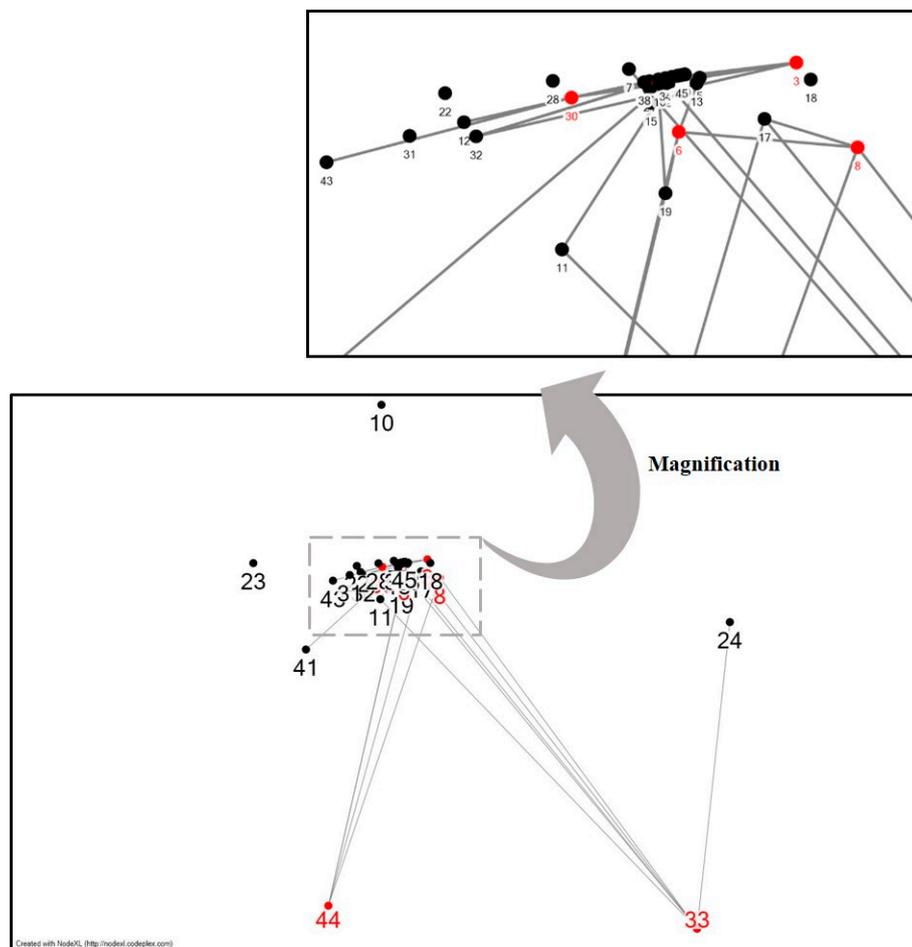


Figure 4. Patent knowledge map (2008–2012, threshold = 0.02).

Table 9. Core research frontiers from patent knowledge map.

RF No.	Core Research Frontier	Degree Centrality	Types of Technology
RF33	Substrate and collector grid structures	0.1364	Structure
RF3	Back side contact solar cell	0.0909	Structure
RF6	CIGS based thin film	0.0909	Material
RF8	Contact layers for thin film solar cells	0.0909	Structure
RF14	Hetero-junction crystalline silicon solar cell	0.0909	Structure
RF20	Metal contact structure for solar cell	0.0909	Structure
RF30	Single crystal silicon solar cell	0.0909	Structure
RF44	Treatment of thin film photovoltaic cell	0.0909	Method

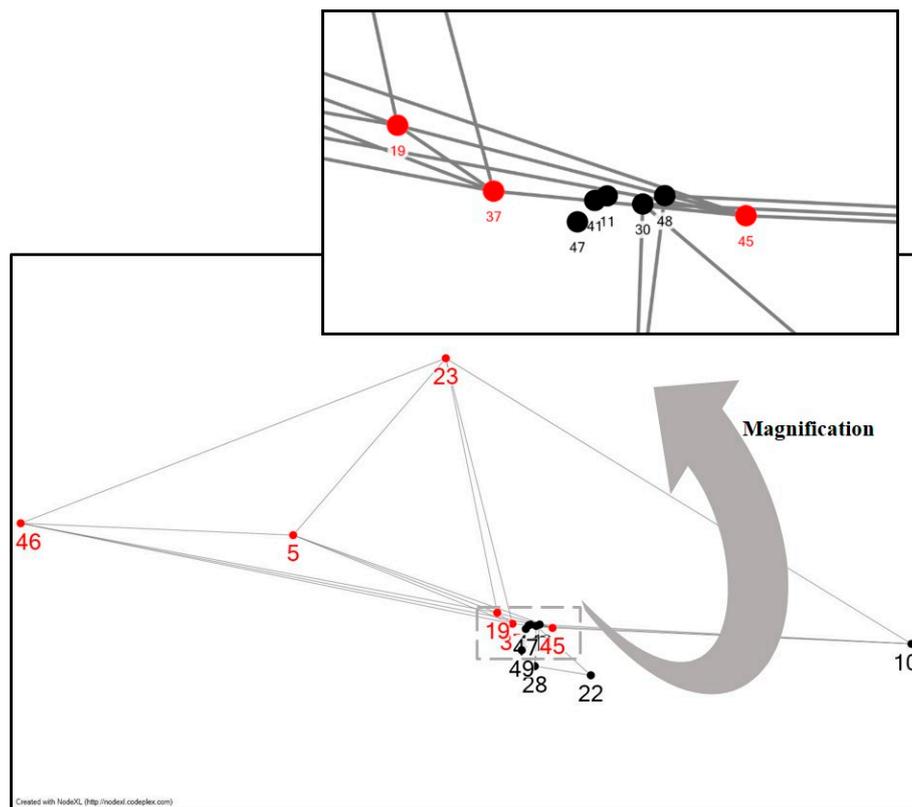


Figure 5. Scientific paper knowledge map (2008–2012, threshold = 0.02).

Table 10. Core research frontiers from scientific paper knowledge map.

RF No.	Core Research Frontier	Degree Centrality	Types of Technology
RF5	Bulkheterojunction thin film solar cell	0.3333	Structure
RF19	Material for thin film photovoltaic cell fabrication	0.3333	Material
RF23	Organic solar cell	0.3333	Attribute
RF37	Tandem-type thin film solar cell	0.3333	Structure
RF45	High-efficiency crystalline silicon solar cell	0.3333	Method
RF46	Polymer solar cell	0.3333	Attribute

3.4. Promising Research Frontiers

Promising research frontiers are identified by conducting dynamic analysis. The whole data set is split into two periods. In the present research, 2008 to 2010 is the first period and the more recent two years, 2011 to 2012, is the second period. Then, knowledge maps, utilizing the patent and scientific paper date which is registered and published from 2008 to 2010, are generated. The dynamic analysis is the process of comparing the knowledge maps of the first period (Figures 6 and 7) and of the second period (Figures 4 and 5) which represent the entire period of knowledge mapping with the cumulative data. The results of the dynamic analysis are shown in Tables 11 and 12.

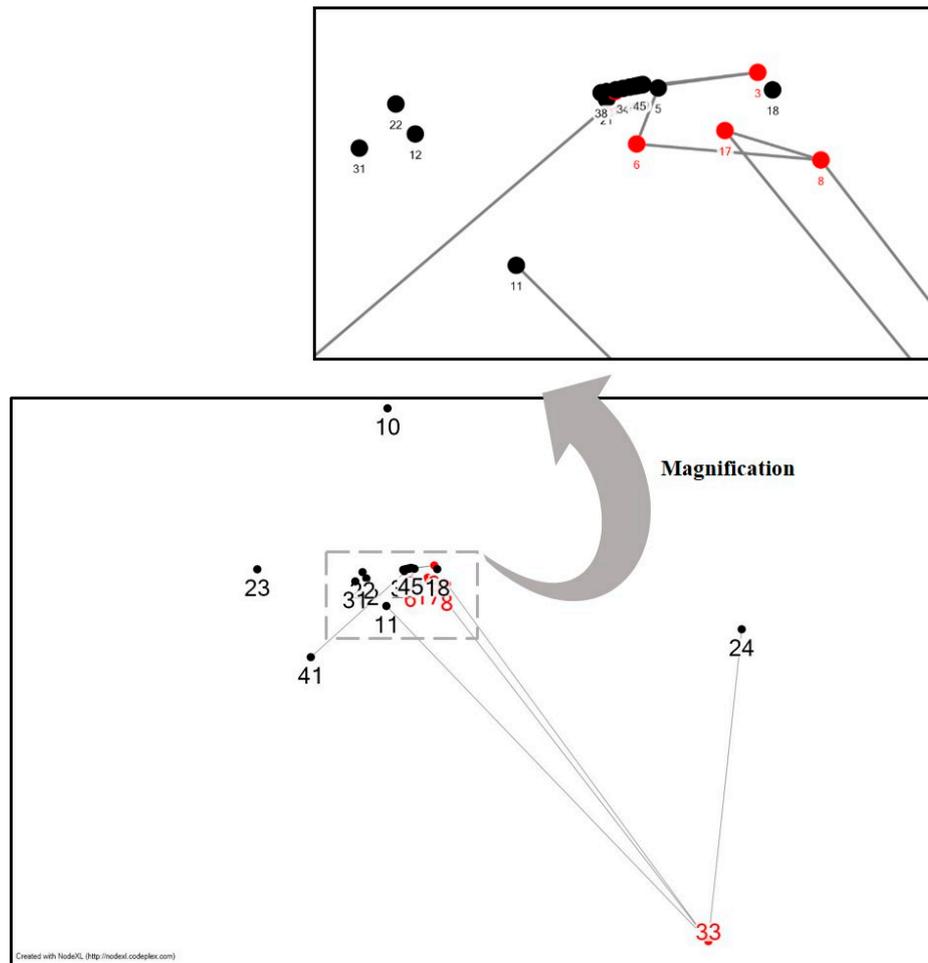


Figure 6. Patent knowledge map (2008–2010, threshold = 0.02).

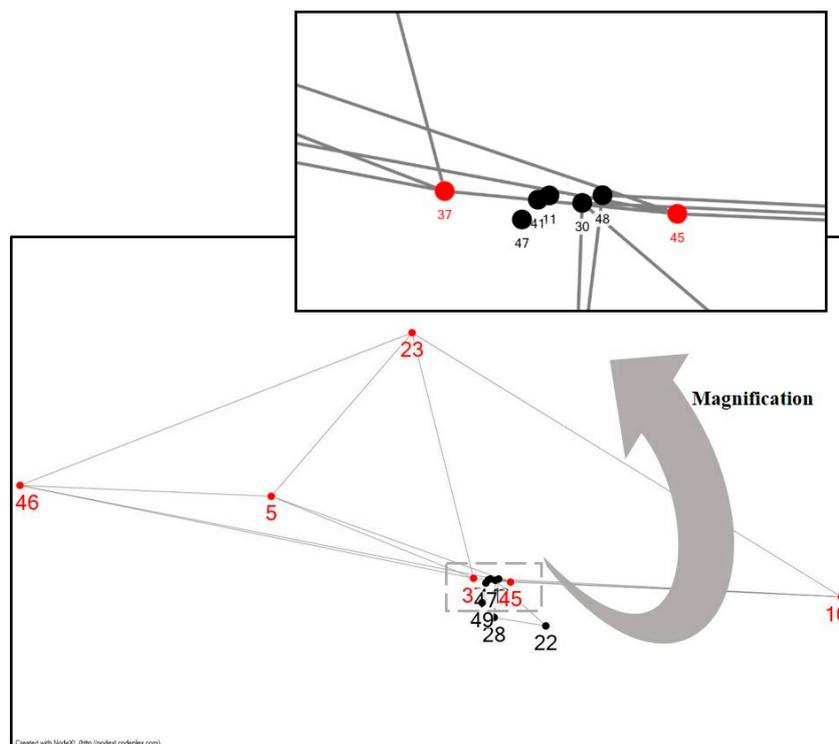


Figure 7. Scientific paper knowledge map (2008–2010, threshold = 0.02).

Table 11. Promising research frontiers from patent knowledge map.

Types of Promising RF	RF No.	Promising Research Frontiers	Types of Technology
Emerging Research Frontier	RF2	Back contact-type solar cell	Structure
	RF7	Contact for a photovoltaic cell	Structure
	RF9	Crystal silicon solar cell	Attribute
	RF13	Forming crystalline layer	Method
	RF15	Integrated thin-film solar cell	Structure
	RF19	Material for thin film photovoltaic cell fabrication	Material
	RF27	Process and system for semiconductor wafer	Method
	RF28	Quantum dot solar cell	Attribute
	RF30	Single crystal silicon solar cell	Structure
	RF43	Three-dimensional thin-film solar cells	Structure
	RF44	Treatment of thin film photovoltaic cell	Method
Enhanced Research Frontier	RF6	CIGS based thin film	Material
	RF10	Dye-sensitized solar cell	Method
	RF11	Electrode for photovoltaic cells	Attribute
	RF18	Manufacturing method for solar modules	Method
	RF22	Nano- solar cells	Attribute
	RF23	Organic solar cell	Attribute
	RF31	solar arrays	Structure
	RF34	Substrate and collector grid structures	Structure
	RF37	Tandem-type thin film solar cell	Structure
	RF38	Terminal box	Structure
	RF41	Thin film solar cell manufacturing	Method
RF42	Thin film solar cell module	Structure	

Table 12. Promising research frontiers from scientific paper knowledge map.

Types of Promising RF	RF No.	Promising Research Frontiers	Types of Technology
Emerging Research Frontier	RF19	Material for thin film photovoltaic cell fabrication	Material
Enhanced Research Frontier	RF5	Bulkheterojunction thin film solar cell	Structure
	RF23	Organic solar cell	Attribute
	RF28	Quantum dot solar cell	Attribute
	RF37	Tandem-type thin film solar cell	Structure
	RF41	Thin film solar cell manufacturing	Method
	RF45	High-efficiency crystalline silicon solar cell	Method
	RF46	Polymer solar cell	Attribute

Twenty three promising RFs among 45 RFs are derived from the patent knowledge map. Five RFs, nine RFs, four RFs, and five RFs are derived from crystalline silicon solar cell, thin film solar cell, organic/quantum dot/nano convergent solar cell, and other solar cell research areas, respectively. Ten promising RFs are related to structure of solar cells, six promising RFs are related to methods, five promising RFs are related to attributes, and two promising RF are related to materials, as shown in Table 11. Eight promising RFs among 15 RFs are derived from the scientific paper knowledge map. One RF, four RFs, two RFs, and one RF are derived from crystalline silicon solar cell, thin film solar cell, organic/quantum dot/nano convergent solar cell, and other solar cell research areas, respectively. Three

promising RFs are related to attributes. Two promising RFs are related to structure and methods, respectively, and a promising RF is related to materials, as shown in Table 12.

3.5. Contents of Promising Research Frontiers

Textual data of patents and scientific papers from 2008 to 2013 is collected using a search formula since the number of core documents is small. The core keywords of promising RFs, shown in Tables 13 and 14, are suggested by extracting highly frequent keywords excluding noises such as article, pronoun, conjunction, *etc.* Table 13 shows the results of only emerging research frontiers since there are many emerging and enhanced research frontiers. However, Table 14 consists of emerging and enhanced RFs. The derived keywords are considered as promising contents of promising RFs since those are from the emerging or enhanced RFs. Thus, the users should pay attention to those contents when they plan technological strategy on promising RFs.

Table 13. Keywords of promising research frontiers from patent knowledge map.

RF No.	Promising Research Frontier	Keywords
RF2	Back contact-type solar cell	quantum wells, silver particle, back contact-type, contact-type solar
RF7	Contact for a photovoltaic cell	oxide, dielectric, coating, dopant, wafer, p-type, n-type, diode, making, polymer
RF9	Crystal silicon solar cell	transistors, nanowire, ceramic, lamina, polymer, nanowhisker, stent, wafer, single-crystal, amorphous silicon, transparent insulator
RF13	Forming crystalline layer	electrode, insulating, polarizing, transistor, retardation, transmissive, capacitor, dielectric, spacer, sealant, photonic crystal
RF15	Integrated thin-film solar cell	oxide, indium, polymer, n-type, p-type, dopant, epitaxial, nanocrystals, diode, sputtering, quantum dot
RF19	Material for thin film photovoltaic cell	oxide, epitaxial, diodes, polymeric, optoelectronic, photoactive, selenide, solvent, electrochromic, liquid polysiloxane, microcrystalline silicon
RF27	Process and system for semiconductor wafer	wafer, semiconductor, chamber, laser, die, silicon, dielectric, annealing, plasma, epitaxial, pressure-sensitive, monocrystalline
RF28	Quantum dot solar cell	nanocrystal, anchor, epitaxial, nanowhisker, nanowire, optoelectronic, ligand, heterostructure, quantum dot, fused nanocrystals, epitaxial layer
RF30	Single crystal silicon solar cell	diodes, implanting, monocrystalline, etching, epitaxial, n-type single-crystal, photodiode, performed, resin, antifuse-configurable interconnect, nanowire photodiode
RF43	Three-dimensional thin-film solar cells	etching, extrusion, bicontinuous heterostructure, solvent, crystallographic, ligand-exchanged, elastomeric, co-extrusion, transparent cnt, titanium dioxide
RF44	Treatment of thin film photovoltaic cell	dopant, n-type, ligands, p-type, sputtering, ion, transistor, optoelectronic, copper-to-indium, etching, molybdenum, electromagnetic radiation

Table 14. Keywords of promising research frontiers from scientific paper knowledge map.

RF No.	Promising Research Frontier	Keywords
RF19	Material for thin film photovoltaic cell fabrication	x-ray diffraction, energy dispersive, band gap, dispersive x-ray, amorphous silicon, microscopy, elsevier ltd, chalcopyrite, degradation
RF5	Bulkheterojunction thin film solar cell	bulk-heterojunction, conversion, copolymers, elsevier, annealing, spectroscopy, photoactive, dissociation, nanoparticles, crystallinity, short-circuit, open-circuit
RF23	Organic solar cell	p3ht, pcbm, phthalocyanine, thiophene, spin-coating, plasmon, heterojunction organic, thermal annealing, fullerene derivative, conjugated polymers, electrochemical impedance, planar heterojunction, organic semiconductors, organic-inorganic hybrid, organic photovoltaics, schottky junction
RF28	Quantum dot solar cell	mesoporous, heterojunction, photoelectrochemical, quantum-dot-sensitized, photoluminescence, passivation, hydrothermal, superlattice, polysulfide electrolyte, quantum dot-sensitized, cds/cdse quantum, inas/gaas quantum
RF37	Tandem-type thin film solar cell	a-si:h, c-si, i-layer, co-evaporation, silicon-germanium, microcrystalline silicon, tandem organic, hydrogenated microcrystalline, uv-vis spectroscopy
RF41	Thin film solar cell manufacturing	cigs, quantum, azo, annealing, polycrystalline, heterojunction, microscopy, amorphous silicon, glass substrates, x-ray diffraction, tin oxide, circuit voltage
RF45	High-efficiency crystalline silicon solar cell	c-si wafers, crystalline silicon, surface texturing, edge isolation, inspection system, qsc silicon, conversion efficiency, force microscopy, mc-si, bv, impurity
RF46	Polymer solar cell	copolymers, heterojunction, donor-acceptor, pedot:pss, mdmo-ppv, photophysical, thermogravimetric, cyclic voltammetry, permeation chromatography, zno nanostructure

4. Discussion

The present research deals with research frontiers that are derived from patents and scientific papers based on the frequency of citation. The data are split into two periods for extract promising research frontiers. Trends of core documents in research areas and research frontiers can be shown in Tables 15 and 16. The number of core documents in the second period is less than that in the first period in terms of both patent and scientific paper RFs since the recently registered patents or published papers have less of a chance to be cited by others. Furthermore, the total number of core scientific papers in the second period is much less than that of patents since the threshold value of patents is much higher than that of scientific papers. Thus, rate is utilized to analyze the trend instead of the frequency. The rate of core patents in thin film solar cell RA rapidly increased and core patents in RF6, 43, 44 are increased while the core patents in RF30, 24, 41, 28, and 18 are decreased. The rate of core scientific papers in crystalline silicon solar cell and thin film solar cell RAs is drastically increased during the second period and organic/quantum dot/nano convergent solar cell RA is decreased. The rate of core scientific papers in RF45 is increased and that of core scientific papers in RF10 is remarkably decreased.

Core research frontiers are central nodes in a knowledge network of whole periods while promising research frontiers are newly emerging or enhanced nodes in the second period. Thus, the relationship between core research frontiers and promising research frontiers is presented in Tables 17 and 18. It can offer some meaningful insights for energy policy makers to understand the relation between core RFs and promising RFs.

Table 15. Change of core patents in research frontier.

Research Area	RF No.	1st Period (2008-2010)		2nd Period (2011-2012)		Total
		Frequency	Rate (%)	Frequency	Rate (%)	
Crystalline silicon solar cell	RF2	0	0.00%	1	1.59%	1
	RF3	1	0.90%	0	0.00%	1
	RF7	0	0.00%	1	1.59%	1
	RF9	0	0.00%	1	1.59%	1
	RF14	2	1.80%	0	0.00%	2
	RF20	1	0.90%	0	0.00%	1
	RF25	0	0.00%	1	1.59%	1
	RF27	0	0.00%	1	1.59%	1
	RF30	5	4.50%	0	0.00%	5
	RF32	0	0.00%	1	1.59%	1
	RF39	1	0.90%	0	0.00%	1
	RF45	2	1.80%	0	0.00%	2
	Subtotal		12	10.81%	6	9.52%
Thin film solar cell	RF4	1	0.90%	0	0.00%	1
	RF5	1	0.90%	0	0.00%	1
	RF6	1	0.90%	9	14.29%	10
	RF8	2	1.80%	0	0.00%	2
	RF13	0	0.00%	1	1.59%	1
	RF15	0	0.00%	2	3.17%	2
	RF17	2	1.80%	0	0.00%	2
	RF19	0	0.00%	2	3.17%	2
	RF21	3	2.70%	0	0.00%	3
	RF24	7	6.31%	0	0.00%	7
	RF29	2	1.80%	0	0.00%	2
	RF37	1	0.90%	1	1.59%	2
	RF41	19	17.12%	2	3.17%	21
	RF42	1	0.90%	1	1.59%	2
	RF43	0	0.00%	6	9.52%	6
RF44	0	0.00%	8	12.70%	8	
Subtotal		40	36.04%	32	50.79%	72
Organic/quantum dot/nano convergent solar cell	RF1	1	0.90%	0	0.00%	1
	RF10	10	9.01%	4	6.35%	14
	RF22	3	2.70%	4	6.35%	7
	RF23	9	8.11%	6	9.52%	15
	RF28	9	8.11%	0	0.00%	9
	RF40	1	0.90%	0	0.00%	1
Subtotal		33	29.73%	14	22.22%	47
The other solar cell	RF11	6	5.41%	3	4.76%	9
	RF12	1	0.90%	0	0.00%	1
	RF16	0	0.00%	1	1.59%	1
	RF18	9	8.11%	2	3.17%	11
	RF26	1	0.90%	0	0.00%	1

Table 15. Cont.

Research Area	RF No.	1st Period (2008–2010)		2nd Period (2011–2012)		Total
		Frequency	Rate (%)	Frequency	Rate (%)	
The other solar cell	RF31	1	0.90%	2	3.17%	3
	RF33	3	2.70%	2	3.17%	5
	RF34	1	0.90%	0	0.00%	1
	RF35	1	0.90%	0	0.00%	1
	RF36	1	0.90%	0	0.00%	1
	RF38	2	1.80%	1	1.59%	3
Subtotal		26	23.42%	11	17.46%	37
Total		111	100.00%	63	100.00%	174

Table 16. Change of core scientific papers in research frontier.

Research Area	RF No.	1st Period (2008–2010)		2nd Period (2011–2012)		Total
		Frequency	Rate (%)	Frequency	Rate (%)	
Crystalline silicon solar cell	RF30	1	0.89%	0	0.00%	1
	RF45	4	3.57%	4	22.22%	8
Subtotal		5	4.46%	4	22.22%	9
Thin film solar cell	RF5	14	12.50%	3	16.67%	17
	RF19	0	0.00%	1	5.56%	1
	RF37	1	0.89%	1	5.56%	2
	RF41	2	1.79%	1	5.56%	3
Subtotal		17	15.18%	6	33.33%	23
Organic/quantum dot/nano convergent solar cell	RF10	24	21.43%	0	0.00%	24
	RF22	15	13.39%	0	0.00%	15
	RF23	12	10.71%	1	5.56%	13
	RF28	4	3.57%	1	5.56%	5
Subtotal		55	49.11%	2	11.11%	57
The other solar cell	RF11	1	0.89%	0	0.00%	1
	RF46	28	25.00%	6	33.33%	34
	RF47	4	3.57%	0	0.00%	4
	RF48	1	0.89%	0	0.00%	1
	RF49	1	0.89%	0	0.00%	1
Subtotal		35	31.25%	6	33.33%	41
Total		112	100.00%	18	100.00%	130

Table 17. Promising research frontier related to core research frontier in patent knowledge map.

RF No.	Core Research Frontier	Promising Type	RF No.	Related Promising Research Frontier
RF33	Substrate and collector grid structures	Emerging	RF2	Back contact-type solar cell
			RF7	Contact for a photovoltaic cell
		Enhanced	RF11	Electrode for photovoltaic cells
RF3	Back side contact solar cell	Emerging	RF30	Single crystal silicon solar cell
		Enhanced	-	-

Table 17. *Cont.*

RF No.	Core Research Frontier	Promising Type	RF No.	Related Promising Research Frontier
RF6	CIGS based thin film	Emerging	RF19	Material for thin film photovoltaic cell fabrication
			RF44	Treatment of thin film photovoltaic cell
RF8	Contact layers for thin film solar cells	Enhanced	-	-
		Emerging	RF44	Treatment of thin film photovoltaic cell
RF14	Hetero-junction crystalline silicon solar cell	Enhanced	RF6	CIGS based thin film
		Emerging	RF19	Material for thin film photovoltaic cell fabrication
RF20	Metal contact structure for solar cell	Enhanced	-	-
		Emerging	RF30	Single crystal silicon solar cell
RF30	Single crystal silicon solar cell	Enhanced	-	-
		Emerging	RF43	Three-dimensional thin-film solar cells
RF44	Treatment of thin film photovoltaic cell	Enhanced	-	-
		Emerging	RF19	Material for thin film photovoltaic cell fabrication
		Enhanced	RF6	CIGS based thin film

Table 18. Promising research frontier related to core research frontier in scientific paper knowledge map.

RF No.	Core Research Frontier	Promising Type	RF No.	Related Promising Research Frontier
RF5	Bulkheterojunction thin film solar cell	Emerging	RF19	Material for thin film photovoltaic cell fabrication
		Enhanced	RF23	Organic solar cell
			RF37	Tandem-type thin film solar cell
			RF45	High-efficiency crystalline silicon solar cell
			RF46	Polymer solar cell
RF19	Material for thin film photovoltaic cell fabrication	Emerging	-	-
		Enhanced	RF5	Bulkheterojunction thin film solar cell
			RF23	Organic solar cell
			RF37	Tandem-type thin film solar cell
			RF45	High-efficiency crystalline silicon solar cell
RF46	Polymer solar cell			
RF23	Organic solar cell	Emerging	RF19	Material for thin film photovoltaic cell fabrication
		Enhanced	RF5	Bulkheterojunction thin film solar cell
			RF37	Tandem-type thin film solar cell
RF37	Tandem-type thin film solar cell	Enhanced	RF46	Polymer solar cell
		Emerging	RF19	Material for thin film photovoltaic cell fabrication
			RF5	Bulkheterojunction thin film solar cell
			RF23	Organic solar cell
			RF37	Tandem-type thin film solar cell
RF45	High-efficiency crystalline silicon solar cell	Enhanced	RF46	Polymer solar cell
			RF19	Material for thin film photovoltaic cell fabrication
			RF5	Bulkheterojunction thin film solar cell
			RF37	Tandem-type thin film solar cell
			RF46	Polymer solar cell

Table 18. *Cont.*

RF No.	Core Research Frontier	Promising Type	RF No.	Related Promising Research Frontier
RF46	Polymer solar cell	Emerging	RF19	Material for thin film photovoltaic cell fabrication
		Enhanced	RF5	Bulkheterojunction thin film solar cell
			RF23	Organic solar cell
			RF37	Tandem-type thin film solar cell
			RF46	Polymer solar cell

Core and promising RFs derived from the patent knowledge map tend to be distributed to structure- and method-related research themes while core and promising RFs derived from the scientific paper knowledge map are equally distributed, as shown in Table 19. The industrial technologies such as structure- and method-related core and promising RFs tend to be derived from the patent knowledge map while the basic technologies such as attribute- and material-related core and promising RFs are more included in the scientific paper knowledge map than in the patent knowledge map. That is to say, RFs from the patent knowledge map are related to development and RFs from scientific paper knowledge map are related to research in the R&D research theme. Types of technology are based on the TEMPEST which is one of methods of classifying patents [59,60]. TEMPEST is an acronym of a viewpoint for analyzing patents. E is abbreviation of energy, representing principle of work devices or technology or power driving devices. M is for material, and P is for personality which depicts the function, property, or attribute of the technology. S (space) point of view approaches the present structure of product or technology. Finally, T (time) demonstrates manufacturing or controlling methods and processes. The present study distinguishes the type of technology as four types—structure, methods, attributes, materials—in Tables 9–12 and 19. The type regarding energy is excluded since it does not exist in the present case.

Table 19. Summary of research frontiers.

Types of Technology	Patent Knowledge Map		Scientific Paper Knowledge Map	
	Core RFs	Promising RFs	Core RFs	Promising RFs
Structure	6	10	2	2
Method	1	6	1	2
Attribute	0	5	2	3
Material	1	2	1	1
Total	8	23	6	8

5. Conclusions and Policy Implications

The present research proposes a methodology for identifying promising research frontiers and their keywords by generating knowledge maps based on scientific documents such as patents and academic articles. To this end, the core documents are derived from the collected scientific documents data based on high citation counts. Research frontiers, which core documents are clustered into, are visualized as a knowledge map to represent the relationship between research frontiers using network analysis and

principal component analysis. The promising research frontiers are identified by conducting dynamic analysis and the contents of the derived promising research frontiers are suggested.

The proposed method and results can be utilized in various ways since the method is based on the patents and scientific papers reflecting both industrial and academic perspectives. First, researchers can establish research direction by confirming core and promising research themes and contents. Second, R&D policy makers can utilize the results and methods as supporting data and a tool for decision making on upbringing policy of promising research. In energy policy, due to long development time and a large scale of R&D budgets, the proposed approach can importantly contribute to devising future energy policy. Third, firms can obtain insights on future technology, products, and market opportunities to get a good competitive position.

This study makes the following contributions. First, a quantitative approach is proposed to suggest promising research frontiers with both industrial and academic perspective by using scientific papers and patents data. Second, on visualization method, a hybrid approach of network analysis and principal component analysis are suggested. It can be used to visualize not only the relation among all but also the one-on-one relations with positions and links in a network. Finally, the characteristic difference between derived research frontiers from both patent and scientific knowledge maps is presented.

Despite these contributions, this study has some limitations. First, the highly cited documents which have emerged recently might be not included since recently emerging documents have a low chance of being cited. To overcome this limitation, estimation treatment on citation frequency of recently emerging documents or utilization of citation time lag distribution or data extraction by published years is needed in further study. Second, although it suggests a systematic approach, experts' participation is necessary in some processes. However, the appropriate experts' participation needs to be required for a high quality of knowledge maps. Thus, a future study can investigate which processes in this approach should be implemented automatically in a system or manually by experts.

Acknowledgments

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014R1A1A2054892).

Author Contributions

Inchae Park designed the study, outlined the methodology, analyzed the data, interpreted the results and wrote the manuscript. Keeun Lee analyzed the data and wrote the manuscript. Byungun Yoon implemented the research, designed the study, outlined the methodology, and helped complete the draft of this research. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix

Table A1. Search queries of the patent data.

Research Area (RA)	Search Query
Crystalline silicon solar cell	(((((crystalline* or crystal* or polycrystalline* or porous*) adj (silicon*)) near (solar adj (cell* or batter*))).ti,ab,cl. and (@RD >= 20080101 <= 20121231))
Thin film solar cell	(((((thin adj film*) or thin-film*) adj (solar or photovoltaic* or photo-voltaic*) adj (cell* or batter*))).ti,ab,cl. and (@RD >= 20080101 <= 20121231))
Organic/quantum dot/nano convergent solar cell	((((organic* or organic-based* or OPV or (quantum adj dot*) or nano* or (carbon adj nano*)) near ((solar* or photovoltaic* or photo-voltaic*) adj (cell* or battery*))).ti,ab,cl. and (@RD >= 20080101 <= 20121231))
The other solar cell	(((((solar* or photovoltaic* or photo-voltaic*) adj (cell* or batter*)) near (dye-sensitize* or (dye and sensitize*) or sensitize* or electrode* or simulat* or structure* or hybrid* or system) not ((thin adj film) or thin-film or organic* or organic-based* or OPV or (quantum adj dot*) or nano* or (carbon adj nano*) or crystalline* or crystal* or polycrystalline* or porous* or silicon*))).ti,ab,cl. and (@RD >= 20080101 <= 20121231))

Table A2. PCA results from patent RF.

Research Frontier	PC1	PC2
RF1	-36.2294	-9.81264
RF2	-35.3673	-10.4946
RF3	-24.0892	-7.23406
RF4	-35.0267	-9.41091
RF5	-32.5856	-9.54674
RF6	-34.4334	-17.887
RF7	-38.858	-8.22966
RF8	-18.7503	-20.2834
RF9	-35.4926	-9.62587
RF10	-44.3567	1367.402
RF11	-44.846	-36.2605
RF12	-53.6982	-16.3904
RF13	-32.858	-10.4107
RF14	-36.2743	-10.0376
RF15	-36.9143	-13.4135
RF16	-36.142	-10.4386
RF17	-26.8796	-15.8782
RF18	-22.8332	-9.82456
RF19	-35.6151	-27.4232
RF20	-35.6103	-9.62142
RF21	-37.0245	-11.2424
RF22	-55.4129	-11.919
RF23	-103.927	-9.92834
RF24	1677.531	-53.4246
RF25	-33.9014	-9.04046
RF26	-33.9014	-9.04046
RF27	-37.5177	-10.2497

Table A2. *Cont.*

Research Frontier	PC1	PC2
RF28	-45.6675	-10.0362
RF29	-37.0278	-10.0857
RF30	-43.9931	-12.6042
RF31	-58.6487	-18.5077
RF32	-52.5925	-18.5794
RF33	86.06667	-408.477
RF34	-35.6179	-9.60763
RF35	-34.038	-9.08518
RF36	-34.3148	-9.17595
RF37	-34.596	-9.26853
RF38	-37.4884	-10.2492
RF39	-33.9014	-9.04046
RF40	-34.4548	-9.22201
RF41	-78.8119	-74.5196
RF42	-35.0267	-9.41091
RF43	-66.286	-22.5906
RF44	-68.4117	-330.743
RF45	-34.1758	-9.13034

Table A3. PCA results from scientific paper RF.

Research Frontier	PC1	PC2
RF5	-691.262	142.8112
RF10	3144.054	-328.72
RF11	-57.4973	-243.063
RF19	-143.136	-190.857
RF22	95.75207	-470.292
RF23	-277.47	3181.938
RF28	-46.5078	-429.089
RF30	-43.0097	-249.118
RF37	-103.766	-239.56
RF41	-62.5159	-246.493
RF45	-1.22539	-257.838
RF46	-1629.9	192.6807
RF47	-69.5766	-262.38
RF48	-34.0751	-242.862
RF49	-79.8689	-357.157

References

1. Lee, W.J.; Sohn, S.Y. Patent analysis to identify shale gas development in China and the United States. *Energy Policy* **2014**, *74*, 111–115.
2. Altwies, J.E.; Nemet, G.F. Innovation in the us building sector: An assessment of patent citations in building energy control technology. *Energy Policy* **2013**, *52*, 819–831.

3. Wong, C.-Y.; Fatimah Mohamad, Z.; Keng, Z.-X.; Ariff Azizan, S. Examining the patterns of innovation in low carbon energy science and technology: Publications and patents of Asian emerging economies. *Energy Policy* **2014**, *73*, 789–802.
4. Lee, K.; Lee, S. Patterns of technological innovation and evolution in the energy sector: A patent-based approach. *Energy Policy* **2013**, *59*, 415–432.
5. Yoon, B.U.; Yoon, C.B.; Park, Y.T. On the development and application of a self-organizing feature map-based patent map. *R&D Manag.* **2002**, *32*, 291–300.
6. De Sollar Price, D.J. Networks of scientific papers. *Science* **1965**, *149*, 510–515.
7. Toivanen, H. The shift from theory to innovation: The evolution of Brazilian research frontiers 2005–2011. *Technol. Anal. Strateg. Manag.* **2014**, *26*, 105–119.
8. Moed, H.F. *Citation Analysis in Research Evaluation*; Springer: Berlin, Germany, 2005; Volume 9.
9. Garfield, E.; Merton, R.K. *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*; Wiley: New York, NY, USA, 1979; Volume 8.
10. Garfield, E.; Small, H. Identifying the changing frontiers of science. *Evolution* **1989**, *1*, 51–65.
11. Small, H.; Griffith, B.C. The structure of scientific literatures i: Identifying and graphing specialties. *Sci. Stud.* **1974**, *4*, 17–40.
12. Braam, R.; Moed, H.; Vanraan, A. Mapping of science by combined cocitation and word analysis. 1. Structural aspects. *J. Am. Soc. Inf. Sci.* **1991**, *42*, 233–251.
13. Peters, H.; van Raan, A.F. Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Res. Policy* **1993**, *22*, 23–45.
14. Peters, H.; van Raan, A.F. Co-word-based science maps of chemical engineering. Part II: Representations by combined clustering and multidimensional scaling. *Res. Policy* **1993**, *22*, 47–71.
15. Persson, O. The intellectual base and research fronts of jasis 1986–1990. *J. Am. Soc. Inf. Sci.* **1994**, *45*, 31–38.
16. Fujita, K.; Kajikawa, Y.; Mori, J.; Sakata, I. Detecting Research Fronts Using Different Types of Combinational Citation. Available online: http://2012.sticonference.org/Proceedings/vol11/Fujita_Detecting_273.pdf (accessed on 2 October 2015).
17. Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* **2008**, *28*, 758–775.
18. Jarneving, B. Bibliographic coupling and its application to research-front and other core documents. *J. Informetr.* **2007**, *1*, 287–307.
19. Jarneving, B. A comparison of two bibliometric methods for mapping of the research front. *Scientometrics* **2005**, *65*, 245–263.
20. Boyack, K.W.; Klavans, R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 2389–2404.
21. Lucio-Arias, D.; Leydesdorff, L. An indicator of research front activity: Measuring intellectual organization as uncertainty reduction in document sets. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 2488–2498.

22. Saka, A.; Igami, M.; Kuwahara, T. Science Map 2008. Study on Hot Research Areas (2003–2008) by Bibliometric Method. Available online: http://www.nistep.go.jp/achiev/ftx/eng/rep139e/pdf/NR139_ScienceMap2008.pdf (accessed on 6 October 2015).
23. Li, W.; Zhao, Y. Bibliometric analysis of global environmental assessment research in a 20-year period. *Environ. Impact Assess. Rev.* **2015**, *50*, 158–166.
24. Cole, F.J.; Eales, N.B. The history of comparative anatomy. Part I: A statistical analysis of the literature. *Sci. Prog.* **1917**, *11*, 578–596.
25. Hulme, E.W. *Statistical Bibliography in Relation to the Growth of Modern Civilization*; Butler and Tanner/Grafton & co.: London, UK, 1923.
26. Lotka, A.J. The frequency distribution of scientific productivity. *J. Wash. Acad. Sci.* **1926**, *16*, 317–323.
27. Gross, P.; Gross, E. College libraries and chemical education. *Science* **1927**, *66*, 385–389.
28. Zipf, G.K. *Human Behavior and the Principle of Least Effort*; Addison-Wesley: Menlo Park, CA, USA, 1949.
29. Fairthorne, R.A. Empirical hyperbolic distributions (bradford-zipf-mandelbrot) for bibliometric description and prediction. *J. Doc.* **1969**, *25*, 319–343.
30. Lawani, S.M. *Quality, Collaboration and Citations in Cancer Research: A Bibliometric Study*; Florida State University: Tallahassee, FL, USA, 1980.
31. Hertzfel, D.H. Bibliometrics, history of the development of ideas in. *Encycl. Libr. Inf. Sci.* **1987**, *42*, 144–211.
32. Brookes, B.C. Comments on the Scope of Bibliometrics. Available online: <https://uhdspace.uhasselt.be/dspace/handle/1942/827> (accessed on 6 October 2015).
33. White, H.D.; McCain, K.W. Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *J. Am. Soc. Inf. Sci.* **1998**, *49*, 327–355.
34. Soper, M.E. *The Librarian's Thesaurus*; American Library Association: Chicago, IL, USA, 1990.
35. Khurshid, A.; Sahai, H. Bibliometric distributions and laws: Some comments and a selected bibliography. *J. Educ. Media Libr. Sci.* **1991**, *28*, 433–459.
36. Pritchard, A. Statistical bibliography or bibliometrics? *J. Doc.* **1969**, *25*, 348–349.
37. Nalimov, V.V.; Mul'chenko, Z. *Naukometriya. Izuchenie Nauki Kak Informatsionnogo Protsessa*; Nauka: Moscow, Russia, 1969.
38. Nacke, O. Informetrie: Ein neuer name für eine neue disziplin. *Nachrichten Dok.* **1979**, *30*, 219–226.
39. Hood, W.W.; Wilson, C.S. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics* **2001**, *52*, 291–314.
40. Davenport, T.H.; Prusak, L. *Working Knowledge: How Organizations Manage What They Know*; Harvard Business Press: Boston, MA, USA, 1998.
41. Eppler, M.J. Making knowledge visible through intranet knowledge maps: Concepts, elements, cases. In Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 3–6 January 2001.
42. Wexler, M.N. The who, what and why of knowledge mapping. *J. Knowl. Manag.* **2001**, *5*, 249–264.
43. Gómez, A.; Moreno, A.; Pazos, J.; Sierra-Alonso, A. Knowledge maps: An essential technique for conceptualisation. *Data Knowl. Eng.* **2000**, *33*, 169–190.

44. Gordon, J.L. Creating knowledge maps by exploiting dependent relationships. *Knowl.-Based Syst.* **2000**, *13*, 71–79.
45. Le-Khac, N.-A.; Aouad, L.M.; Kechadi, M. Knowledge map: Toward a new approach supporting the knowledge management in distributed data mining. In Proceedings of the Third International Conference on on Autonomic and Autonomous Systems, Athens, Greece, 19–25 June 2007.
46. Van Raan, A.F. The Use of Bibliometric Analysis in Research Performance Assessment and Monitoring of Interdisciplinary Scientific Developments. Available online: <http://www.cwts.nl/TvR/documents/AvR-TFA2003.pdf> (accessed on 2 October 2015).
47. Boyack, K.W.; Klavans, R.; Börner, K. Mapping the backbone of science. *Scientometrics* **2005**, *64*, 351–374.
48. Medina, C.; Leeuwen, T. Seed journal citation network maps: A method based on network theory. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 1226–1234.
49. Su, H.-N.; Lee, P.-C. Knowledge map of publications in research policy. In Proceedings of the Portland International Conference on Management of Engineering & Technology, Portland, OR, USA, 2–6 August 2009.
50. Shin, J.; Park, Y. Generation and application of patent claim map: Text mining and network analysis. *J. Intellec. Property Rights* **2005**, *10*, 198–205.
51. Segev, A.; Kantola, J. Identification of trends from patents using self-organizing maps. *Expert Syst. Appl.* **2012**, *39*, 13235–13242.
52. Lee, S.; Yoon, B.; Park, Y. An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation* **2009**, *29*, 481–497.
53. Son, C.; Suh, Y.; Jeon, J.; Park, Y. Development of a gtm-based patent map for identifying patent vacuums. *Expert Syst. Appl.* **2012**, *39*, 2489–2500.
54. Yoon, B.; Lee, S.; Lee, G. Development and application of a keyword-based knowledge map for effective R&D planning. *Scientometrics* **2010**, *85*, 803–820.
55. Chen, A.; Chen, R. Design patent map: An innovative measure for corporative design strategie. *Eng. Manag. J.* **2007**, *19*, 14–29.
56. Chen, R. Design patent map visualization display. *Expert Syst. Appl.* **2009**, *36*, 12362–12374.
57. Kim, J.; Park, Y. Identifying a new service opportunity from potential needs: User-centric service map. In Proceedings of International Conference on Industrial Engineering and Engineering Management (IEEM), Macau, China, 7–10 December, 2010.
58. Glänzel, W.; Czerwon, H.-J. A new methodological approach to bibliographic coupling and its application to research-front and other core documents. In Proceedings of International conference on International Society for Scientometrics and Informetrics, River Forest, IL, USA, 7–10 June 1995.
59. Coh, B. *Technology Analysis and Patent Information Analysis*; Korean Intellectual Property Office: Seoul, Korea, 2012.
60. Jeong, Y.; Yoon, B. Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. *Technovation* **2014**, *39–40*, 37–52.