World Patent Information 39 (2014) 50–57

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Knowledge dissemination patterns in the information retrieval industry: A case study for automatic classification techniques

Ricardo Eito-Brun*

Universidad Carlos III de Madrid. c/Madrid, 124, Getafe (Madrid), Spain

A R T I C L E I N F O

Article history:
Available online 16 July 2014

Keywords:
Citation analysis
Automatic document classification
Document clustering
Bibliometric indicators
Innovation capability assessment

A B S T R A C T

Patents provide valuable information to identify flows in the transfer of technical knowledge and assess the innovation capabilities of the actors involved in different industries. Patent citations are also recognized as a valid tool to measure the impact of innovations and to identify key influencers in diverse activity sectors.

This study analyzes a collection of U.S. patents granted in the period between 1990 and 2012 for the subject "automatic document clustering and classification", a key technology within the Information Retrieval and Text Mining disciplines. The purpose of this research is to identify – using citation analysis – the most productive and influential companies and journals, and the patterns followed in the transfer and sharing of technical knowledge. The paper identifies the most productive organizations (those that have been granted a higher number of patents) and those with a higher impact (organizations whose patents have received a major number of citations), and compares the generated rankings with those obtained using traditional bibliometric indicators. The conclusions provide an overview of the innovation landscape in the area of study, and suggest to which extent bibliometric indicators match the conclusions obtained after analyzing productivity and impact using patent citation.

1. Introduction

The production of patents is an indicator usually applied to identify the companies and research centers leading different areas of technical knowledge. The number of patents granted to different institutions is a signal of the institutions' productivity and their capability to innovate. Patent citations have also been recognized as a valuable source of information to assess the impact of research [1,2]. The patents receiving a higher number of citations are considered as being the most influential. This approach can also be used to assess the impact that organizations involved in innovation activities and new product development have on subsequent research. Organizations whose patent portfolio receives a higher number of citations are supposed to have a bigger influence on the global technical progress achieved by all the parties involved in particular areas of knowledge.

This article presents the conclusions of a bibliometric study completed on a collection of U.S. patents granted in the area of "automatic document classification and clustering". The purpose of this research was to identify the most influential companies and research institutions and give visibility to the knowledge transfer patterns that characterize this discipline. Although the study has been done on a sample set of patents granted on this specific area, the methodology proposed may be used to analyze any other subject domain. Document classification and clustering techniques fall within the Text mining discipline. This is an application of Information Retrieval and Computational Linguistics aimed to help users identify and extract new knowledge from large collections of documents and textual corpus. Text mining focuses on the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources [3]. Text mining focuses on the extraction of knowledge from textual repositories of unstructured information.

Text mining applies Information Retrieval techniques, automated classification and artificial intelligence. The study focuses on patents related to two of these areas: document clustering and automatic text classification. Both share a common objective: the creation of groups of documents with similar semantics, to give information systems' end-users the capability of interactively exploring large document sets and retrieve relevant items based on their similarity. This helps improve information retrieval effectiveness: most of the information retrieval models based on

* Tel.: +34 669 796498.
E-mail address: reito@bib.uc3m.es.

http://dx.doi.org/10.1016/j.wpi.2014.06.003
0172-2190/© 2014 Elsevier Ltd. All rights reserved.

keywords require a match between the terms used in the queries and the terms appearing in the documents. Documents are only retrieved if they contain the words entered by the users when searching the database. This makes sure that documents that are relevant to the user's interest but do not contain the query terms are excluded from the results. Document clustering and classification widen the possibility of retrieving relevant documents. Similar documents are put together within groups represented by a surrogate or centroid. End-user queries are compared with these surrogates, and the most relevant groups are retrieved. Then, users can explore the elements within the group and select individual documents. The advantage of automatic classification and clustering is that documents within the same group may not necessarily contain the terms used in the query. This is because their similarity is calculated using the documents' full text, while traditional search based on keywords just considers the few words typed in the query. Problems due to synonymy or homonymy may be avoided by applying classification and clustering techniques. Regarding the difference between document classification and clustering, in classification the text of the documents is compared with a set of existing classifiers to determine the membership of the document to a specific group. These classifiers usually consist of a set of terms that represent concepts or subjects, and are obtained after a training process based on previously classified documents. In document clustering, documents are automatically compared to each other to create the initial set of groups. When the similarity between two documents – or between one document and the documents in an existing group – is higher than a predefined threshold, the document is incorporated as a member of the group. The set of terms that characterize each group are in turn dynamically revised as new documents are processed and added to the groups. In document clustering there is not a predefined set of classifiers.

Results of the quantitative analysis reported in this paper do not need to be understood as a judgment of the productivity or innovation capabilities of the cited companies or their commitment to develop innovative products in the area of study. Conclusions are based on a sample set of patents that constitute a partial representation of the outcomes of the innovation effort developed by companies. Quantitative data referring to company information are just provided to illustrate the method of analysis that is being proposed.

2. Patents and the transfer of technical knowledge

Patents do not only provide information about the achieved technical progress in a specific area, they also give details about patterns in knowledge transfer between basic research (usually disseminated and published in academic journals) and practical research done with commercial purposes (traced by means of patents). Three different kinds of knowledge transfer patterns may be considered:

- Transfer of knowledge as a whole, considering all the involved actors and its evolution in time. This can be represented through the measurement of the productivity and impact achieved by the different companies and institutions at different periods.
- Transfer of knowledge focused on individual organizations. This analysis is aimed to identify: a) those companies having a stronger influence on the research conducted by the analysis' target company and, b) the influence of this company on the research conducted by other parties. This analysis should not consider just patent citations, but also the citations to technical reports and papers present in patents.
- Transfer of the knowledge disseminated through academic journals. It is based on the citations that articles published in journals have received in the analyzed set of patents. It may be understood as an indicator of the impact of basic, academic research, on the practical innovation process embedded in patents.

This combined analysis of knowledge transfer patterns has been applied on a subset of patents for the area of study. The proposed methodology might be applied to any other subject area or with a wider scope within the Information Retrieval discipline. In any case, when presenting the conclusions it should be considered that the analysis of knowledge transfer patterns through patents has some limitations. First of all, patents are not fully representative of the innovation outcomes, as not all the innovation results translate into patents. In addition, the constraints and restrictions applied in some countries to the patentability of software-intensive or software-supported inventions clearly have a negative impact on our area of study. Another issue in patent citation analysis is the fact that citations may be added to patents by their authors or by the experts in charge of assessing the documents and deciding whether the innovation may be patented or not. In his classical work, Callon et al. [4] mentioned that the inclusion of references and bibliography in patents may not be as rigorous as in the articles published in academic journals.

Regardless of these constraints, patents are the only document type that is publicly available and may be used to objectively assess the status of the innovation achieved in different areas and how this knowledge is transferred from scientific and academic research and converted into practical, applicable inventions. The analysis of patent citations is today one of the main methods at our disposal to obtain indicators about:

- The relationship between science or basic research, and techniques or applied research and the dynamics of this interaction.
- The relationships between the research completed by different inventors and organizations.
- The identification of the most influential journals in the development of innovations.

3. Related work

Bibliometric indicators have been traditionally applied to assess the impact of journals and the productivity of personal researchers. This study applies three of them – Bradford's law, h-index and g-index -, to assess companies based on the citations received by their patents.

Bradford's law [5,6] is a classical bibliographical bibliometric technique used to identify the core journals in any knowledge area. It was proposed by Samuel Bradford in 1948, to study the distribution of the scientific literature. This technique is usually applied to study the distribution of citations to identify the most relevant journals in a specific area. Bradford's law states that there is an inverse relationship between the number of articles published in a subject area and the number of journals in which the articles appear. The conclusion of this analysis identifies the "core" set of journals in the discipline. This helps librarians decide the journals the libraries should subscribe to, making a better investment of the budget available for acquisitions. This technique has been used in other bibliometric studies focused on assignees' productivity [7] and on the analysis of non-patent references in patent citations [8].

The most popular indicators for assessing researchers are the Hirsch Index or h-index and the g-index. The first one, the h-index quantifies the cumulative impact and relevance of the scientific output of an individual [9,10]. This index has been widely used to compare the researchers in different knowledge areas [11–13], including patent analysis [14–16]. It takes into account the quantity of papers published and the citations that they have received. A particular researcher has an index h, if h of his N papers has received at least h citations each, and the other (N-h) papers have received less than h citations. The h-index may be used not only to rank producers of knowledge, but also to identify the core intellectual products (papers, articles, patents, etc.) produced by a person or organization.

The second indicator under discussion is the g-index [17]. It was proposed as an improvement to the h-index, to improve some of its limitations [18]. More concretely, the h-index is criticized because it is not sensitive to the set of non-cited or lowly cited items and to the highly cited papers. Although the first constraint is generally accepted as a positive characteristic of the index, the second one is considered to be negative. The author of the g-index, Leo Egghe, designed an index that also included the papers receiving a higher number of citations. This g-index is defined as the unique, largest number such that the top g papers together receive g^2 or more citations, and g-index will be always greater or equal to the h-index for a particular author. A comparison of the performance of g-index and h-index has been recently published by Abramo et al. [19].

These indexes are also valid tools to rank organizations based on their productivity and on the impact of their scientific production and patent portfolios. Additional bibliometric indexes could be applied when analyzing patents, for example the citation speed index [20] or immediacy index that measures how fast recently published items are incorporated as citations in new documents, the classical impact factor used to rank academic journals, the Publication Efficiency Index (PEI) used to know if the impact of the publication is aligned with its research effort, or the Eigenfactor to name a few [21].

4. Methodology

The conclusions in this paper are based on a detailed selection of patents granted in the period between 1990 and 2012 in the automatic document classification and text clustering area. Patents in other sub-areas of text mining like text summarization and feature extraction have been excluded, as well as those that apply clustering algorithms and techniques to process images with purposes not directly related to the semantic categorization of document content. Although the application of this selection criterion could be considered to be a restriction, it is remarked that text summarization and feature extractions share a similar theoretical foundation and that, in some cases, it is not easy to assign a contribution to a specific sub-area.

Following these criteria, a total number of 533 patents containing 11,898 citations to other patents and 5.804 citations to other types of documents (journal articles, books, technical reports, etc.) have been selected from an initial set of patents retrieved from the Thomson Reuters Delphion database. This database was selected because its records incorporate the list of cited patents and additional bibliography. The initial set of patents was identified using full-text search and classification codes. Different terms were combined: “document clustering”, “document categorization”, “summarization”, “text mining” or “information retrieval”. The retrieved patents were later screened and reviewed to discard those not dealing with the target area of study. The final subset of patents cover an interesting period that includes the first steps in the development of the World Wide Web and the emergence of new, successful companies in the Information Access industry. In fact, the analysis shows a significant shift regarding the most productive and influential companies in this area before and after the generalization and wide adoption of the World Wide Web. The selected sample also represents the state of the art and the evolution of classification and clustering related technologies in the period under study and gives a better understanding on how the main actors of the Information Access industry positioned their IP portfolio and research efforts in the early years of the World Wide Web.

Once the final set of patents was identified, each document was processed to collect the following data:

- Organization owning the rights on the patent, and the year when the patent was granted. This date was used instead of the year of application. Patent renewals were not considered in the analysis.
- Number of citations to other patents. This data is used to compare the percentage of citations given to patents with the percentage of citations given to other documents, including academic journals and technical reports.
- Organizations owning the rights on the cited patents. These data are the basis to identify those organizations with a bigger impact, acting as knowledge spillovers.

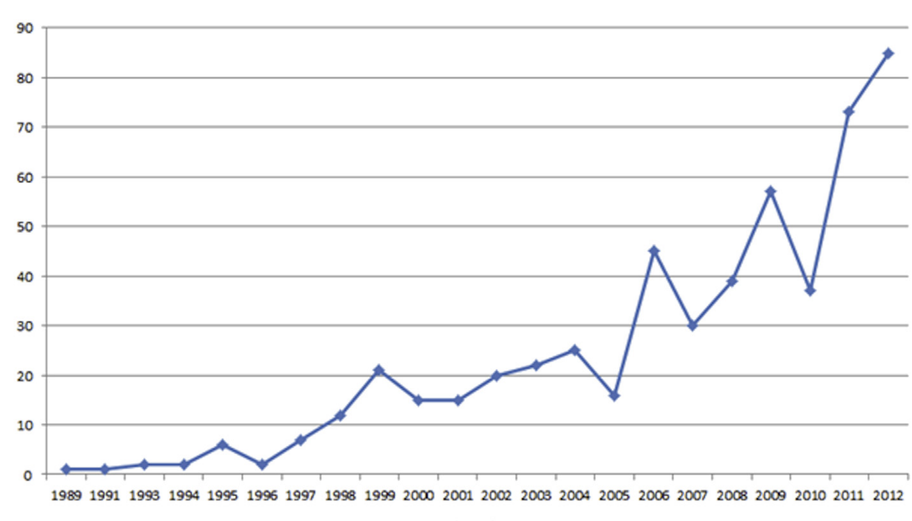

Fig. 1. Patents granted per year.

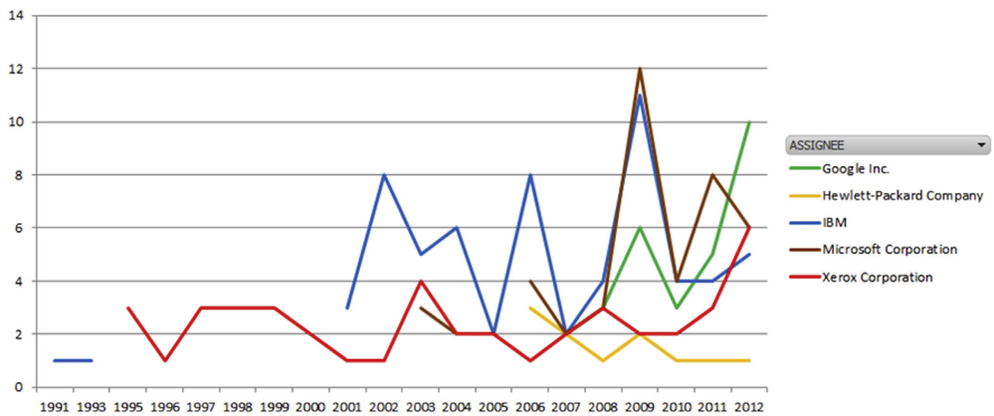

Fig. 2. Patents granted per year (5 major players).

- Bibliographic citations in the patent (excluding citations to other patents). This includes citations to technical reports, thesis, dissertations, proceedings and articles published in academic journals. With these data it is possible to know the number of citations obtained by the different academic journals and conferences. These data are the basis to identify those journals with a major impact. It is also useful to assess the impact of companies through the citations received by their technical reports.

These data have been analyzed applying bibliometric techniques. The conclusions of this analysis are summarized in the following sections.

5. Description of the data set

The data set includes 533 patents and 17,702 citations. The data set is the result of an exhaustive selection of patents related to the target domain - automatic clustering and classification – from the selected data source. The size of the data set and the number of analyzed citations is similar to the data sets used in other Scientometric studies [22,23] and it is considered representative to obtain reliable conclusions. Figure 1 displays a chart with the number of patents granted in the different years of the period under study. It shows a significant increment in the number of patents starting from 1997. This increment may be explained by two reasons: a) the major investment on Information Technology that characterized that period, and b) the widely adoption of web technologies and the need for designing more sophisticated solutions to help organizations manage the complexity of a growing amount of un-structured data.

One interesting aspect in the evolution of the number of patents is the presence of downward and upward peaks observed in years 2005, 2007 and 2010. This pattern, that requires further investigation that is out of the scope of this research, is also observed when analyzing the evolution of the number of patents granted to the key players in the industry. Figure 2 shows this behavior in the number of patents granted to Google Inc., IBM and Microsoft. This chart also identifies the key players that have made a continuous research effort on this area. For example, it is observed that Xerox Corp. has kept a regular production of patents since 1995, although their cumulative figures are lower than those of other competitors.

Regarding citations, the sample set includes 11,898 citations to other patents and 5804 citations to other types of documents (reports, monographs, web pages, conference proceedings and articles from academic journals). The analysis of the impact of organizations' innovations is based on the citations received by their patents, Although the inclusion of other, non-patents, documents should be a positive aspect, the scope of the study has focused on patent citations to assess the productivity and impact of organizations' innovations. This is due to two different factors: a) the difficulties in getting the authors' affiliation data for document types other than patents, and b) number of patents is more significant than academic papers to assess organizations' innovation capabilities and output of applied research efforts. Table 1 shows the average number of citations per patent and year. The general trend shows an increment in the average number of citations per patent, with some temporal oscillations in the values of the two variables.

Table 1
Distribution of average number of citations per year.

	N° citations to other patents (average)	N° bibliographic citations (average)
1995	12.16	4
1996	7.5	7
1997	6.28	13.42
1998	12.16	8
1999	9.8	5.71
2000	12.8	8.1
2001	16.93	6.3
2002	13.45	7.28
2003	10.95	5.87
2004	16.8	9.41
2005	14.87	7.5
2006	20.04	12.75
2007	19.4	17.45
2008	23.12	13.9
2009	35.17	21.54
2010	24.37	16.65
2011	30.68	18.89
2012	27.47	14.44

6. Analysis of productivity

The analysis of the total number of patents per organization during the whole period shows the most productive companies: those with a higher number of patents (see Fig. 3). The ranking is led by IBM, followed by Microsoft Corp., Xerox Corp., and Google Inc. in the fourth position. The patents granted to these companies represent 34.21% of the total number of patents.

7. Analysis of impact

Analysis of citations constitutes the basis for assessing the most influencing organizations. The companies with a higher number of

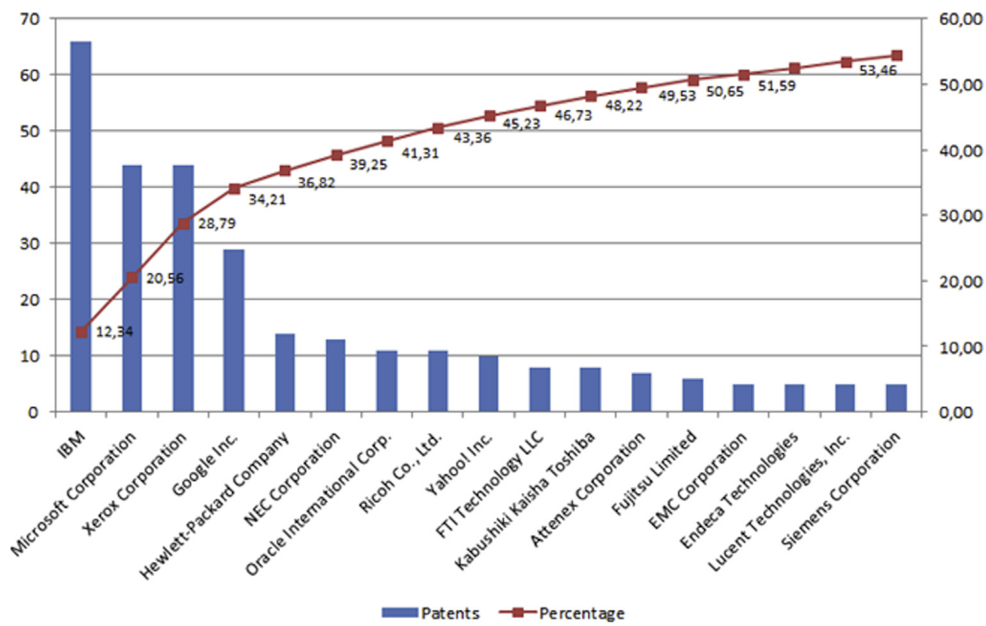

Fig. 3. Patents granted per organization in the period 1990–2012.

citations, whose patents had a major impact on subsequent research, are shown in Table 2:

Figure 4 shows the distribution of citations per company (including self-citation) for the whole period:

The previous data were obtained including self-citation. In the context of this analysis, self-citation occurs when one patent granted to a company includes references to other patents granted to the same company. As part of the analysis, self-citations have been removed with no impact on the ranking of the most influential companies with the exception of Google, that moves from the 11th position to the 21st position (see Table 3). The exclusion of self-citation seems to be a reasonable step, as self-citation does not represent any transfer or flow of knowledge between different parties. It is remarked anyway that self-citation exclusion has not led to different conclusions when analyzing the impact of the actors involved in this market.

Table 2 and Fig. 4 provide a visual display of the companies whose patents have had a higher impact. It is possible to establish a parallelism with one of the traditional bibliometric methods described in a previous section: Bradford's Law. By applying this classical analysis to patent citations, it would be possible to identify the "core institutions" generating inventions or innovations in a specific area. The analysis divides the distribution under study into three or more zones, having in each zone the same number of citations. The number of members in each zone increments following the pattern 1:n:n^2:n^3, etc. With the sample data used in this study, it would be possible to set up a distribution with four zones, each one containing around 2300 citations. The first one, known as the "core", includes four companies: IBM, Xerox, Microsoft and Oracle. The second zone includes 23 companies, the third one 132 and the last one 1258.

This basic analysis does not reconcile the data about productivity (number of published items) with the data about impact (citations received by the published items). Companies might have a large number of patents that have received a small number of citations, or a small number of patents that have received a large number of citations. This is one of the subjects of study in bibliometrics, and researchers have proposed indicators to assess both factors. One of the questions leading this research was to know whether the proposed bibliometric indicators represent with accuracy the intuitive idea that innovative companies should be characterized both by their productivity and by the impact of their innovations. The preliminary, combined analysis of productivity and impact for the sample data set has led to the creation of charts plotting the raw data about the number of granted patents and the number of citations they have received. A bubble chart is used, in which the *x*-axis represents the productivity (number of granted patents), and the size of the bubbles and their position on the *y*-axis represents the impact of the patent portfolio of the company

Table 2
List of organizations sorted according to the number of citations received.

Organization	Received citations (including self-citation)	Received citations (excluding self-citation)	% In total number of citations (including self-citation)
IBM	1087	981	9.15%
Xerox Corp.	625	505	5.26%
Microsoft Corp.	593	457	4.99%
Oracle Corp.	239	205	2.01%
Hitachi Ltd.	199	195	1.67%
Digital Equipment Corp.	151	151	1.27%
AT&T Corp.	144	144	1.21%
Fujitsu Ltd.	129	129	1.09%
HNC, Inc.	129	126	1.09%
NEC Corp.	126	111	1.06%
Google Inc.	115	55	0.97%
Sun Microsystems, Inc.	99	98	0.83%
Hewlett Packard Company	92	83	0.77%
Ricoh Company Ltd.	89	85	0.75%
Canon Inc.	86	82	0.72%
Matsushita Electric Industrial Co. Ltd.	84	84	0.71%
Apple Computer Inc.	83	83	0.70%
Lucent Technologies Inc.	78	77	0.66%
Amazon.com	77	77	0.65%
Fuji Xerox Co., Ltd.	72	71	0.61%

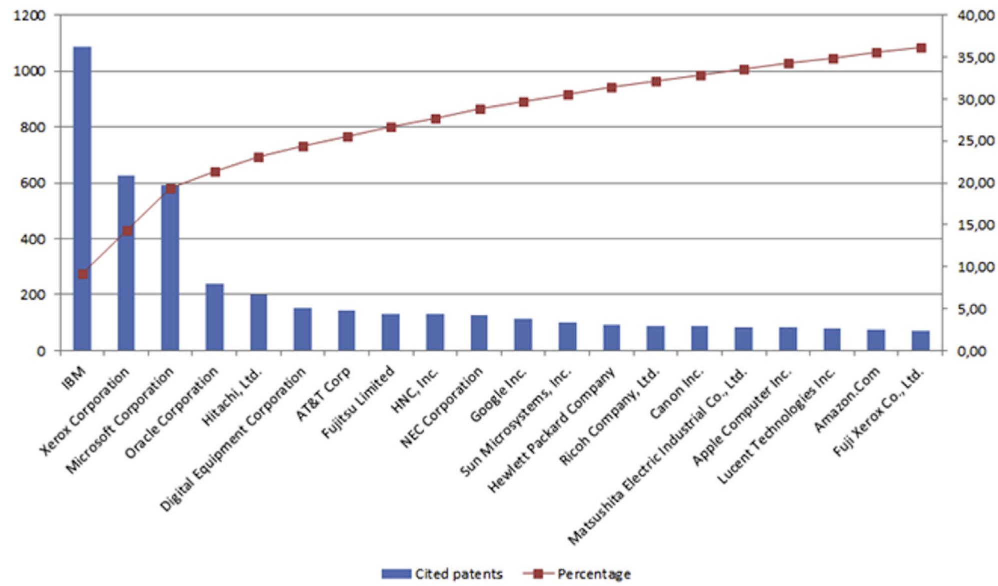

Fig. 4. Identification of the most influential companies based on number of citations.

(number of citations received). Figure 5 shows the results obtained when plotting the data for all the available years.

These charts may be generated for different time periods to analyze the evolution of companies and the changes in the productivity-impact scenario. The chart corresponding to the years 1995–2000 shows a different ranking, with Xerox Corp. having a greater productivity than IBM (see Fig. 6).

In the years 2001–2006, IBM consolidated its position as a leader regarding productivity and impact, and Microsoft started standing-out (see Fig. 7).

Finally, in the last five-year period, 2007–2012, the chart shows a relevant shift, with a significant shift in Microsoft's position and the appearance of Google with an outstanding productivity but still with a lower impact (see Fig. 8).

Using this model, it is possible to track the evolution of companies regarding productivity and impact, and represent the dynamics of their innovation outcomes using vectors in a 2-dimensional space. The changes in the relative positioning of companies in the chart also show a relevant aspect of the knowledge transfer pattern for the area of study as a whole: the companies that become active producers of knowledge consumed by other actors.

8. Productivity and impact. Bibliometric indicators

This section reports the results obtained when applying the h-index and the g-index on the data set. The values for the h-index, excluding self-citations, are shown below:

The calculation of the g-index for the sample set of data shows the results in the next table (see Table 4). It is remarkable the shift in the position of companies like Microsoft or DEC.

Table 3
List of organizations sorted according to h-index.

h-index	Organizations with this h-index
5	IBM, Xerox Corp.
4	DEC, Infoseek
3	Microsoft, Fujitsu, Amazon.com, Canon, HNC, AT&T.
2	Intel, HP, Apple, Oracle, Hitachi, Yahoo!, Toshiba, Google, Accenture, Lexis-Nexis, Lucent, Lycos, MIT, SAP, Syracuse University, University of California, etc.

In addition to the measurement of the companies' productivity and impact, and the assessment of the feasibility of using traditional bibliometric indicators with patent collections, the sample data set has led to the identification of journals and academic conferences having a higher impact on applied research. Bradford analysis has been used to identify the core set of journals receiving most of the citations. This list includes the following journals on the core area of the distribution:

Journal full name	Abbreviated name	% Citations
ACM SIGIR Int. Conf. of Research and Development on Information Retrieval	ACM-SIGIR Conf. Res. Dev. Info. Retrieval	15%
Information Processing & Management	INFORM PROCESS MANAG	5%
Journal of the American Society for Information	J AM SOC INFORM SCI	5%
Text Retrieval Conference	TREC	5%
Int. Conf on Machine Learning	ICML	4%

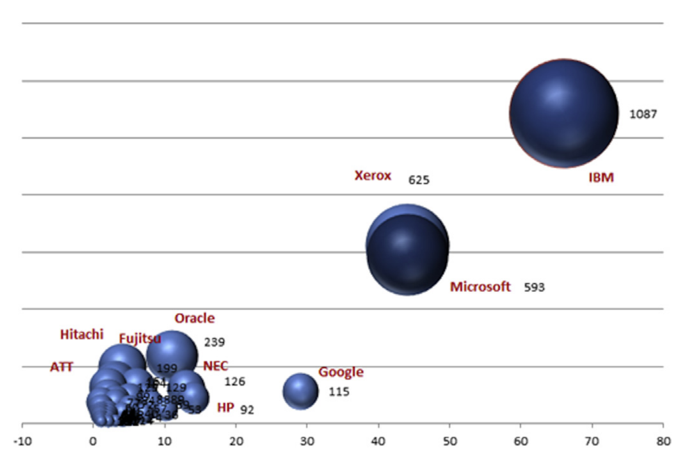

Fig. 5. Productivity and Impact Chart for the period 1995–2012.

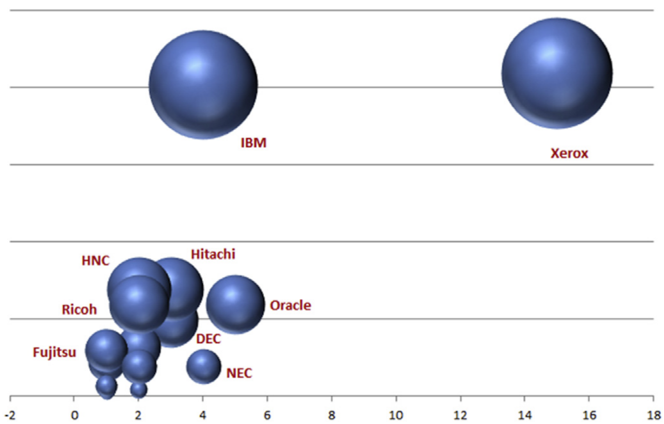

Fig. 6. Productivity and Impact Chart for the period 1995–2000.

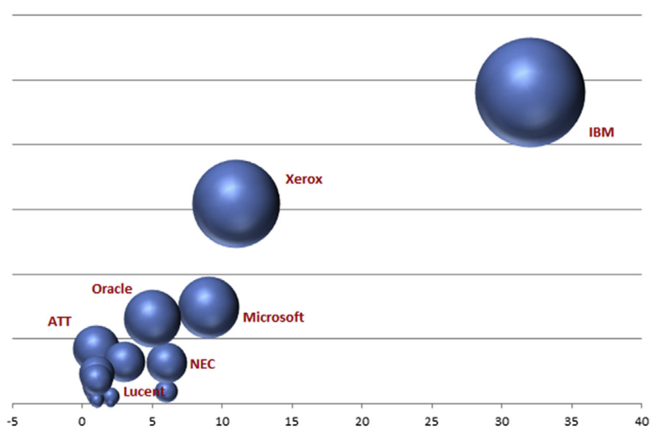

Fig. 7. Productivity and Impact Chart for the period 2001–2006.

9. Conclusions

Analysis of patents is a necessary activity to know the status of practical research and innovation in technical areas. Traditional bibliometric techniques based on citation analysis are used to identify those organizations leading the innovation processes. Applied methods need to combine the assessments of two variables: productivity, based on the number of granted patents, and impact based on the number of citations received by the companies' patent portfolio. The analysis of citations in different timeframes also provides relevant knowledge about shifts in the position of companies in the innovation landscape. Graphical representations and dynamic charts give an intuitive understanding of these changes and the market evolution. As part of this study, focused on a sample set of patents for the Information Retrieval discipline, two bibliometric indicators that combine impact and productivity have been calculated to assess the feasibility of using these methods in patent analysis. The g-index seems to provide a more accurate characterization of the ranking of companies involved in the studied area of knowledge: document clustering and classification, as it is more sensitive to the patents with a higher number of citations that may be excluded when calculating the h-index. The combination of these indicators with the proposed graphical representations provides an accurate, combined picture of the productivity and impact variables. In today's scenarios, characterized by the need for measuring the effectiveness of innovation strategies and investments at the corporate, institutional and national level, and by the availability of bigger data sets coming from different sources, innovation researchers will soon have the opportunity to compare and analyze complementary sets of related data and apply traditional bibliometric indicators in innovative ways to get more accurate pictures of the organizations' productivity and impact.

As shown in this study, citation analysis may be used to identify relevant conclusions regarding the transfer of technical knowledge between the agents involved in innovation activities: a) to which extent the actors provide knowledge that is consumed or reused by other actors, b) the flows of knowledge between organizations, based on the impact that their respective patent portfolios have on each other, and c) how the knowledge disseminated through academic journals is used in the creation of new inventions.

The method followed in this study could be applied to analyze any other area of knowledge or disciplines, and the combined analysis of patent citations and citations to other bibliographic materials can help give answers to questions related to the actual value of academic research and the transformation of basic research into practical innovations.

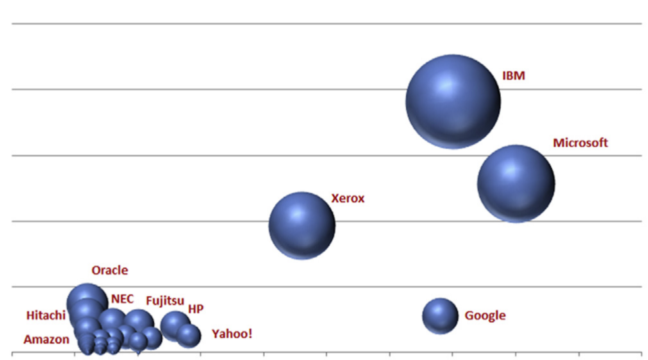

Fig. 8. Productivity and Impact Chart for the period 2007–2012.

Table 4
List of organizations sorted according to g-index.

g-index	Organizations with this g-index
6	IBM, Xerox Corporation
5	Microsoft, HNC, AT&T
4	DEC, Infoseek, Fujitsu, Amazon, SAP
3	Oracle, Canon, Hitachi, Yahoo!, Intel, HP, Syracuse Univ
2	Google, Lexis-Nexis, Toshiba, Accenture, Apple, MIT, etc.

Acknowledgment

This article has been developed from a presentation by the author at the ICIC meeting (International Conference on Trends for Scientific Information Professionals) held in Vienna 13 – 16 October 2013.

References

[1] Karki MMS. Patent citation analysis: a policy analysis tool. World Pat Inf 1997;19(4):269–72.
[2] Jaffe Adam B, Trajtenberg Manuel, editors. Patents, citations and innovations: a window on the knowledge economy. Cambridge, Ma: London: The MIT Press, cop; 2002.
[3] Hearst M. What is text mining?. Unpublished essay; Oct 2003. Available from:, http://people.ischool.berkeley.edu/~hearst/text-mining.html [accessed 03.01.14].
[4] Callon M, Courtial J-P, Penan H. Introduction à la scientométrie: de la bibliométrie à la veille technologique. París: Anthropos; 1990.
[5] Chung Y-K. Core international journals of classification systems: an application of Bradford's law. Knowl Organ 1994;21(2):75–83.
[6] Garg KC, Karki MMS, Krishnan Marg KS. Bibliometric study of world literature on patents. World Pat Inf 1988;10(4):237–42.

[7] Lo S. Patent coupling analysis of primary organizations in genetic engineering research. Scientometrics 2008;74(1):143–51.
[8] Lo Szu-chia S. Scientific linkage of science research and technology development: a case of genetic engineering research. J Sci 2010;82(1):109–20.
[9] Norris M, Oppenheim C. The h-index: a broad review of a new bibliometric indicator. J Doc 2010;66(5):681–705.
[10] Bornmann L, Werner M. The h index as a research performance indicator. Eur Sci Ed 2011;37:77–80.
[11] Ciriminna R, Pagliaro M. On the use of the h-index in evaluating chemical research. Chem Central J 2013;7(1):132.
[12] Kulasegarah J, Fenton JE. Comparison of the h index with standard bibliometric indicators to rank influential otolaryngologists in Europe and North America. Eur Arch Oto-Rhino-Laryngol 2010;267(3):455–8.
[13] van Raan A. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics 2006;67(3):491–502.
[14] Kuan Chung-Huei, Huang Mu-Hsuan, Chen Dar-Zen. Ranking patent assignee performance by h-index and shape descriptors. J Informetrics 2011;5(2):303–12.
[15] Guan JCh. Exploring the h-index at patent level. J Am Soc Inform Sci Technol 2009;60(1):35–40.
[16] Zhang S. Exploring the nonlinear effects of patent H index, patent citations, and essential technological strength on corporate performance by using artificial neural network. J Informetrics 2012;6(4):485–95.
[17] Egghe L. Theory and practise of the g-index. Scientometrics 2006;69(1):131–52.
[18] Costas R, Bordons M. The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level. J Informetrics 2007;1(3):193–203.
[19] Abramo G, D'Angelo C, Viel F. Assessing the accuracy of the h- and g-indexes for measuring researchers' productivity. J Am Soc Inf Sci Tec 2013;64(6):1532–2890.
[20] Bornmann L, Daniel H-D. The citation speed index: a useful bibliometric indicator to add to the h index. J Informetrics 2010;4(3):444–6.
[21] Andrés A. Measuring academic research. Cambridge: Chandos Publishing; 2009.
[22] Lo S. Patent analysis of genetic engineering research in Japan, Korea and Taiwan. J Sci 2007;70(1):183–200.
[23] Wang G, Guan J. Measuring science-technology interactions using patent citations and author-inventor links: an exploration analysis from Chinese nanotechnology. J Nanopart Res 2011;13(12):6245–62.

Dr. Ricardo Eito-Brun is an associate professor at Universidad Carlos III de Madrid, Spain, where he teaches different subjects related to knowledge organization, knowledge representation and information management. Ricardo holds a master degree in Informatics from Universidad Carlos III de Madrid and in Documentation and Information Science from University of Granada (Spain). His research interest is in information management practices, quality and process improvement, and he has been responsible for several large scale content management projects. He is the author of four books on mark-up languages and XML and numerous articles and conference papers in the field of information management.