World Patent Information 50 (2017) 64–72

Contents lists available at ScienceDirect

World Patent Information

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

Claim-based patent indicators: A novel approach to analyze patent content and monitor technological advances

Douglas Henrique Milanez [a, b, *], Leandro Innocentini Lopes de Faria [a, b], Roniberto Morato do Amaral [a, b], José Angelo Rodrigues Gregolin [a]

[a] Federal University of São Carlos, Centre for Technological Information in Materials, Dept. of Materials Engineering, 13565-905, São Carlos, SP, Brazil
[b] Federal University of São Carlos, Centre for Technological and Organizational Intelligence Research, Dept. of Science Information, 13565-905, São Carlos, SP, Brazil

ARTICLE INFO

Article history:
Received 20 December 2016
Received in revised form
8 August 2017
Accepted 21 August 2017

Keywords:
Bibliometrics
Nanotechnology
Nanomaterials
USPTO
Innovation forecasting

ABSTRACT

This paper proposes a new method for developing patent indicators by text mining patent claims according to their drafting structure. We apply the method on nanocellulose as a case of study, although any subject could be the target of investigation. The results show that patent claims are a more reliable source of key terms to develop technical indicators than, for example, patent titles and abstracts. Indicators from patent claims in combination with other traditional indicators developed from bibliographic patent data may contribute significantly to the analytical process of technological forecasting, monitoring and competitive intelligence studies.

1. Introduction

1.1. Patents as a source of technological indicators

Patent statistics have been frequently adopted to follow technological trends, innovative activities, market analysis, and players involved in the innovation process, at least since the second half of the last century. In 1985, Pavitt [1] had already noticed the increasing use of patent statistics and he associated it to the growing recognition of the importance of technological change in firm competitiveness, improvements in technologies of information storage and retrieval, and the need for statistical evidence to support personal experience and expert opinion. The high volume of data currently available also justifies the need for quantitative analysis based on patent information [2].

Patent documents are rich sources of technological and business information. They codify part of the tacit knowledge generated by the technological development of new products, processes, methods, compositions, apparatus, etc. They also enable reasonably standardized quantitative data to perform accurate assessments using data and text mining approaches [2–4]. These assessments allow evaluating the effects of trade and industry production, development of industrial sectors, policy making related to scientific technological activities, and links between science and technology [1–3,5]. Much research and reports on using patent indicators can be found in the literature [for instance, see Refs. 3–13]. Moreover, Madani and Weber [4] performed an interesting review of patent mining evolution using bibliometrics and keyword network analysis.

In the context of technological forecasting, many studies have been conducted using patents as technological sources for quantitative evaluation. Considering patent analysis as a means of investigating phases of the technology life cycle, Chachetti et al. [14] overviewed the technological activity in hydrogen storage materials using patent indicators and network analysis for gathering insights on the future developments in this field. Caviggioli [11] investigated technology fusion by verifying convergence relying on the International Patent Classification (IPC) of patents filed at the European Patent Office between 1991 and 2007. His hypothesis was that the first occurrence of a patent incorporating a combination of IPC subclasses signals a new instance of fusion.

* Corresponding author. Federal University of São Carlos, Centre for Technological Information in Materials, Dept. of Materials Engineering, 13565-905, São Carlos, SP, Brazil.
E-mail address: douglasmilanez@yahoo.com.br (D.H. Milanez).

http://dx.doi.org/10.1016/j.wpi.2017.08.008
0172-2190/

Another example is the approach for forecasting promising technology proposed by Kim and Bae [12], where forward citations, triadic patent families and independent claims were used to assess promising technology clusters.

Most of the indicators are assembled by mining the bibliographic data of patent documents, which usually includes the IPC codes, priority numbers, countries of filing, inventors and patentees, etc. [15]. In more specialized studies, citations to prior patents and non-patent literature have also been used to improve the data mining exploitation of patent documents [16]. Van Raan [3] provided a detailed and insightful review of patent citation analysis and a new approach to map technology-relevant areas, focusing on patent-patent citation in the context of economic value of patents as well as citation to non-patent literature to landscape S&T linkage. One of the main conclusions is that only 3–4% of scientific publications covered by Web of Science or Scopus are cited by patents.

Non-controlled free text fields, such as titles and abstracts, have also been mined in order to extract key-terms that could depict the technical content of a set of patent documents. Courtial, Callon and Sigogneau [9] applied co-word analysis to normalized terms extracted from patent titles to get a panorama of a given field. The work of Brietzman [8] is another example of mining terms from patent titles and abstracts to assess industry R&D. Nevertheless, the use of titles and abstracts as a source for term extraction is limited and cannot provide all relevant aspects of an invention [17,18].

To fill this gap, researchers have been exploiting the content from the full text patent documents. For instance, key-terms have been extracted with routines based on the morphological and syntactic structure of sentences [17–26]. Other more specific techniques rely on segmenting and combining terms [17,21,26,27], summarizing before extracting key-words [17], vector representation and grouping approaches [28–30], clustering analysis [7,17,30] and mapping [22,25,26,29]. Nonetheless, less attention has been paid to considering rules for the drafting of each part of the patent document, such as the specific rules dictating the writing of patent claim sentences.

1.2. Patent claims as a source of indicators

Patent claims can be understood as “the heart of a patent” because it specifies the invention's scope of protection. According to the United State Patent and Trademark Office (USPTO) [31], claims must point out “the subject matter that the inventor or inventors regard as the invention, […] (and what) defines the scope of the protection of the patent”. Thereby, claims are a valuable source of technical terms [32]. To extract useful key-words from the claims, several methodologies have been proposed with different aims, such as improvement of wording and translation during the filing period of a patent document [19,33], information retrieving [34], claim overlapping and legal analysis [22,35], conceptual mapping [22], and technological forecasting and competitive intelligence [23–26]. Nevertheless, none of these studies dealt properly with the particularities of the patent claim structure. For instance, claim sentences can be relatively long, reaching more than 200 words, which can cause failures in the natural language processing [36].

According to the WIPO Patent Drafting, a claim can be divided into three main parts [32]: 1) the preamble, which introduces the category of invention or an essential part of the invention – can be a process, a method, a composition, a product, etc.; 2) the transition phrase, which separates the preamble from the body of the claim; and 3) the body of the claim, which details the characteristics of the invention. Patent claims are also divided into independent claims, which contain the general aspect of the invention, and dependent claims, which details each part of the invention – and may always be referred directly or indirectly to an independent claim, i.e. all dependent claims should be grouped together with the independent claim(s) to which they refer to [31,32].

The claim's “internal structure” can be exemplified with two sentences present in a patent document, as shown in Table 1. These two sentences are the first and the second claims from the patent document numbered US20110086948 [37] and we have already separated them according to the structure of writing. The first sentence is an independent claim while the second one is a dependent claim. Both preambles refer to a product (composite material) and the body of these claims contains useful key words that details the referred invention: “nanocellulose”, “maleic anhydride graft poly(ethylene-octene) copolymer resin”, “nylon-4 resin”, etc. In the approach that we propose, we considered the claim structure in the text mining routine that will be described in this paper and it allowed us to extract more accurately detailed information regarding products, compositions and processes protected by patents.

This paper presents a new method for developing patent indicators using patent claims as a source of key-terms. The idea is to investigate technical details in order to provide technological indicators with high aggregate value thereby supporting technological forecasting studies and decision making. The method take into account the structure of drafting the claim, and it was designed to analyze patents filed in the United States Patents and Trademarks Office (USPTO), however the logical principle of depicting the claim structure can be applied to patent documents in other offices. We used nanocellulose as a case study to delineate a sample of full-text patent documents, but it can be used for any other subject or technological field.

1.3. Nanocellulose: a sustainable nanomaterial

The technological developments of nanotechnology and nanomaterials have grown at least since the beginning of this century. High budgets to support research and development in these topics have been executed due to their great potential in promoting innovations. Consequently, nanotechnology has been the target of monitoring and forecasting activities based on patent indicators [for instance, see Refs. 38–44]. Nanocellulose is an emerging and economically promising nanomaterial that can be obtained from renewable sources, such as plants, woods, natural fibers, etc. It has been estimated that the American market for nanocellulose in 2020 will be US$ 250 million and that its production can reach 780 tons in 2017 [45].

The mechanical properties of nanocellulose are higher than the ones from conventional cellulose fibers. Furthermore, the attractive properties of nanocellulose also include biocompatibility and biodegradability, gas barrier, water absorption and rheological and optical properties. Among the main applications, besides being a reinforcing agent in composite materials and paper, we can mention packaging, optically transparent paper for electronic devices, texturizing agents in cosmetics and food, dressings and bioartificial implants [46–51].

Nanocellulose is a generic term for a set of cellulose-based nanomaterials, which also include cellulose nanofibrils, cellulose nanocrystals, bacterial cellulose [18,47,49,51,52]. In the top-down manufacturing process, nanostructures are obtained by