

Technological Indicators of Nanocellulose Advances Obtained from Data and Text Mining Applied to Patent Documents

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

^aCenter for Technological Information in Materials – NIT/Materiais, Materials Engineering Dept,
Federal University of São Carlos – UFSCar, São Carlos, SP, Brazil

^bInformation Science Dept, Federal University of São Carlos – UFSCar, São Carlos, SP, Brazil

Received: January 23, 2014; Revised: November 21, 2014

Abstract: Nanocellulose is remarkable cellulose-based nanomaterials that have a potential for innovation and sustainable appeal. Their advances can be assessed using patent indicators and text mining techniques. The aim of this study was at analyzing the advances in nanocellulose based on indicators compiled from patents filed at the United States Patent and Trademark Office (USPTO) from 2000 to 2012. Assignees, technological subjects, highly cited patents, applications and types of nanocellulose were obtained by mining structured and unstructured data. The results highlighted the different interests in the USA market, mainly after 2007. Mined terms from titles and abstracts could add further information to the analysis. However, although the method applied was useful, it was not sufficient to identify all applications and types of nanocellulose involved in the sample analyzed, therefore it is recommended that other document parts be included in future analyses.

Keywords: *sustainability, data mining, text mining, patent analysis, technological mapping*

1. Introduction

Nanocellulose is an emerging and sustainable nanomaterial, which has been the target of scientific and technological research due to its interesting functional properties and higher mechanical properties than regular cellulose fibers or other conventional reinforcement materials¹⁻⁵. Leaders in producing cellulose and pulp, such as the USA, China and Brazil, as well as others with large forest reserves, notably Canada, Finland and Sweden, have supported nanocellulose developments. The United Kingdom and France may also be accounted for this group due to their initiatives and contributions⁵. The global market of nanocellulose has a good perspective, especially in the USA whose market is forecasted to be US\$ 250 million by 2020⁶. In spite of many potential applications of nanocellulose^{1-4,7}, it is estimated that it will be mainly used in composite materials, paper, medical devices, filters and electronics⁶.

Nanocellulose is a generic term to refer to cellulose-based nanomaterials and many others nomenclatures can be observed in the literature^{1,2,4,5}. Basically, there are two types of cellulose-based nanomaterials, whose nomenclatures have recently been standardized as cellulose nanofibrils (CF) and cellulose nanocrystals (CC)⁸, and the main differences between them rely primarily on the crystalline degree of the cellulose¹⁻³. CF comprises crystalline and amorphous domains of cellulose chains while CC contains only the crystalline cellulose. Wood and pulp, natural fibers, plants in general, forest and agricultural residues are examples

of sources for CF and CC, which are prepared by using mechanical and acid hydrolysis processes (top-down approaches), respectively. Bacterial fermentation (bottom-up approach) is another process that produces CC, and it has been called bacterial cellulose (BC) in most studies to differ from CF and CC obtained from top-down approaches¹⁻⁴.

Despite renewed characteristics, potential properties and research efforts in nanocellulose, there are considerable challenges to be overcome. For example, even though an upscale in the production of nanocellulose is expected in coming years⁶, costs should be minimized^{1,2,5}. The systematic characterization of nanocellulose sources, properties and behaviors, such as mechanical, rheological, surface properties and interaction with living tissues are also issues not completely solved. Furthermore, due to the agglomeration trends, more developments are needed concerning surface modification and functionalization^{1-3,9}.

A significant increase in the number of scientific publications and patent documents in nanocellulose have recently been observed^{4,5,9}. To follow progress in nanocellulose, technological indicators can be compiled using quantitative techniques applied to patent documents using well-established procedures, criteria and recommendations in patent statistics¹⁰⁻¹⁴. Indeed, patent documents are an interesting source of information as it merges technical, legal and business data in a well-established format^{11,12,14-16}. Patent statistics have been increasingly used in policy reports and a new field of academic research has emerged aiming at developing new methodologies to analyze and use of patent data as an indicator of science and

*e-mail: douglas@nit.ufscar.br

technology research output¹⁰⁻¹⁸ as well as in technological monitoring and strategic analysis^{4,5,10,12,17,18}.

Patent indicators have also some limitations, though. Not all inventions are patented, thus, some companies prefer secrecy and other mechanisms to gain market dominance. There is also evidence of differing patent behavior across industries and countries over time. Furthermore, standards across patent offices affect patent numbers, although underlying inventive activities may remain unaffected, and the counting of patent documents gives the same weight to all patents regardless their economic and technical value¹⁰⁻¹². An approach to select economically and technically relevant patent documents is to analyze those patents filed in the USA market because of their great economic importance¹⁰⁻¹², despite home advantage issues¹⁹. Citation analysis of patents is also a usual method to enhance relevant documents and to track knowledge. It has been evidenced that the number of citations a single patent receives reflects its technological and commercial influence¹⁰⁻¹².

Due to the increasing volume of information available and challenges involved in collecting and analyzing this amount of data and texts, data-mining and text-mining techniques have been researched, so that scientist, managers and engineers are able to keep up with what has been done in a specific research field, especially in the context of engineering^{5,10-12,16,17}. Data-mining applied to patent documents deals with structured part of documents, which essentially includes the bibliographic data, such as assignees, inventors, priority country (country of the first application), dates and classification codes^{11,12,16}. However, data-mining results might not be sufficient to point out clearly technical aspects of inventions, for instance fabrication routes, applications and products involved. To fulfill this gap, text-mining techniques may help by analyzing unstructured texts (also called free texts) of patent documents, such as titles, abstracts, technical report, claims etc.¹⁴⁻¹⁶ Data- and text-mining complement each other in the purpose to build a whole framework of competitors, historical developments and trends, technological monitoring, hotspots, gaps and opportunities to support strategic technological planning^{11,14,15,17}.

Mining unstructured texts is a newer, empiric and much less developed field when compared to mining structured data of patent documents, thus it has currently been the subject of research^{14-16,20-26}. There are advances in methodological practices and tools to perform analysis and visualization, such as text summarization, noun-phrase extraction and natural language processes, clustering, vector representation and mapping^{16,20-26}. Titles and abstracts have been the main sources of unstructured text, although all parts of documents can be eventually explored¹⁶.

There are some few studies concerning technological indicators and patent analyses regarding nanocellulose. Durán, Lemes and Seabra⁷ aimed at updating the technological advances in cellulose nanocrystals by analyzing the content of patent documents obtained from important repositories without using any quantitative technique, which decreased the number of document evaluated. Despite the lack of some methodological criteria, for instance, the search expression and criteria to select the

documents analyzed, the authors provided an assessment regarding preparation, treatments, and applications of the nanomaterial. Charreau et al.⁴ attempted to visualize the technological development of cellulose nanocrystals, microfibrillated cellulose and bacterial cellulose using data mining technique. They obtained evolution trends, top patentee and most cited patents. Milanez et al.⁵ forecasted worldwide trends of the scientific publications and patent applications using growth curves, compared nanocellulose evolution with other nanomaterials and the relation between scientific and technological advances among the USA, Japan, China, Canada, Brazil and some European countries (France, Sweden, the UK, Germany and Finland). However, these studies explored only structured data of patent documents, leaving out information that is potentially relevant, such as materials, processes and other technical aspects that could be assessed by text mining. Furthermore, none of them focused to analyze the nanocellulose patenting activity at the USA market, which may also be significant one for nanocellulose⁶ as it already is for nanotechnology²⁷⁻²⁹.

Hence, the purpose of this study was to map the nanocellulose patenting activity in patent applications filed in the United States Patent and Trademark Office (USPTO¹) from 2000 to 2012. It provides the evolution of patenting and technological subjects, priority countries, main patent assignees, and most cited patents with focus on application and materials, specially the type of nanocellulose.

2. Material and Methods

2.1. Data-mining procedure to compile and analyze patent document indicators

Technological indicators were developed using data mining procedure on a dataset of 500 full text patent documents in nanocellulose filed at the USPTO - United States Patent and Trademark Office³⁰, from 2000 to 2012, in which the retrieval procedure, databases used and other criteria are described in topic 2.3. An additional analytical element was the identification of the top 15 most cited patent documents using the Derwent Innovations Index website³¹, whose values were obtained according to procedure also explained in detail in Topic 2.3¹¹.

The indicators were compiled and analyzed according to the guidelines and procedures recommended by the OECD Patent Statistic Manual¹² and by the field of quantitative studies^{10,14,16}. The indicators were developed with support of the VantagePoint® (version 5.0) software. To select the years of filing, priority countries, the assignees' names and International Patent Classification (IPC) codes from the sample data-mining technique were applied. The annual number of patent filed at the USA was analyzed comparatively to the worldwide perspective¹¹. The

¹ The USPTO is the federal agency that issues patents and trademarks in the USA.

¹¹ The citation analysis was conducted in July 16th 2014, therefore it counted the number of citation that documents received until this date.

¹¹ The worldwide value considered the patent records retrieved from the Derwent Innovations Index database.

technological subject was depicted using the subclass level of IPC codes.

Assignees' names were standardized, because there is no control of this information in the filing act, and the name considered for standardization was the one found in the assignee website, as recommended by the OECD Patent Statistics Manual¹². These websites also were consulted for further information about the assignee. Only assignees with five or more patents were ranked. A profile of activity for each assignee was developed by considering their main subject categories and mined terms that characterized the type of nanocellulose and application (which were obtained according to the procedure in Topic 2.2). To the top cited patents, a profile was developed taking into account the number of citation, priority year, assignee name, subject categories, mined terms. Co-assignee patents were also identified as a way to determine collaborative developments^{11,12}.

2.2. Text-mining procedure to compile and analyze patent document indicators

Noun-phrase terms related to types of nanocellulose - cellulose nanofibrils (CF), cellulose nanocrystals (CC), bacterial cellulose (BC) - and applications of the inventions were extracted from patent titles and abstracts by using the following procedure, which is analogous to practices found in other studies^{10,14,16}. Natural language process (NLP) was used to mine phrases from titles and abstracts texts³². Stop words were removed and routines to identify and group meaningful terms concerning application and types of nanocellulose were applied. Due to the uncontrolled language of the patent texts, only terms that appeared at least in five documents were considered¹⁶.

2.3. Procedure to collect the full text patent document sample

The procedure used to collect the full text patent document sample followed the recommendation provided by the WIPO^{IV} Guide to Patent Information³³, which suggest the use of keywords and IPC codes in a search expression. Nanocellulose-related terms were extracted from scientific articles and reviews^{1-3,5}, however no IPC code specific to nanocellulose was found. The nanotechnology IPC code (B82) did not fill the interest of this study, even when combined to term related to cellulose, thus it was discarded. The final expression considered for collecting the main patent document sample was the following:

("bacterial cellulose*" OR "cellulos* crystal*" OR "cellulos* nanocrystal*" OR "cellulos* whisker*" OR "cellulos* microcrystal*" OR "cellulos* nanowhiskey*" OR "nanocrystal* cellulose*" OR "cellulos* nano-whisker*" OR "cellulos* nano-crystal*" OR "nano-crystal cellulose*" OR "cellulos* micro-crystal*" OR "cellulos* microfibril*" OR "microfibril* cellulose*" OR "cellulos* nanofibril*" OR "nanofibril* cellulose*" OR "micro-fibril* cellulose*" OR "nano-fibril* cellulose*" OR "cellulos* micro-fibril*" OR "cellulos* nano-fibril*" OR "cellulos* nanofiber*" OR "nanocellulose*" OR "cellulos* nanoparticle*" OR

"nano-cellulose*" OR "nanopartic* cellulose*" OR "nanosiz* cellulose*" OR "cellulos* nanofill*" OR "nano-siz* cellulose*" OR "cellulos* nano-fiber*" OR "cellulos* nanoparticle*" OR "cellulos* nano-fill*" OR "nano-particl* cellulose*")

A core of bibliographical patent records about nanocellulose was previously collected from the worldwide Derwent Innovations Index (DII)³¹ database on Jun 26th, 2013. Sample consistence was checked exhaustively and 505 from the 1,480 records had at least one patent filed at the USPTO. The USA patent documents were retrieved with the support of GETIPDL software³⁴ using the US patent number found in the 505 DII bibliographic records.

Although DII patent records are interesting to compile patent indicators due to their coverage, we sought to explore the potentiality of the full text documents to provide useful information, particularly from their original titles and abstracts. Titles and abstracts from the DII records were not used because the database modifies them in order to provide more information and this would affect the final analysis. The full text data were then imported into the data and text mining software and the final sample comprised 500 patent documents, after removing carefully duplicate records. Therefore, the studied accounted two sources of information for the same invention: bibliographic data from DII records and title and abstract from full texts retrieved from the USPTO database.

The most cited patent filed at the USA were retrieved redoing the search at the DII database, but now using the US patent number (the same used to retrieve the full text document at the USPTO database). After performing the search, the patents were listed according to the most cited to the less cited and the top 15 patents were selected to be assessed.

2.4. Statistical calculations performed

The annual growth rate (G_i) was obtained using Equation 1, where N_i is the number of patent documents in the year "i" and N^{i-1} is the number of patent documents in the year before ("i-1").

$$G_i = \frac{(N_i - N_{i-1}) * 100}{N_{i-1}} \quad (1)$$

The share (S) of patent documents was calculated using Equation 2, where S_i is the number of patent documents from an assignee, technological subject, application or type of nanocellulose (i) and S_t is the total number of patent documents in the context of the analysis (t).

$$S = \frac{S_i * 100}{S_t} \quad (2)$$

3. Results and Discussion

3.1. Evolution, priority countries and technological subjects

The number of patent documents increased rapidly from 2000 to 2010 for both patents of worldwide perspective and those filed at the USPTO, according to Figure 1, and

^{IV} World Intellectual Property Organization.

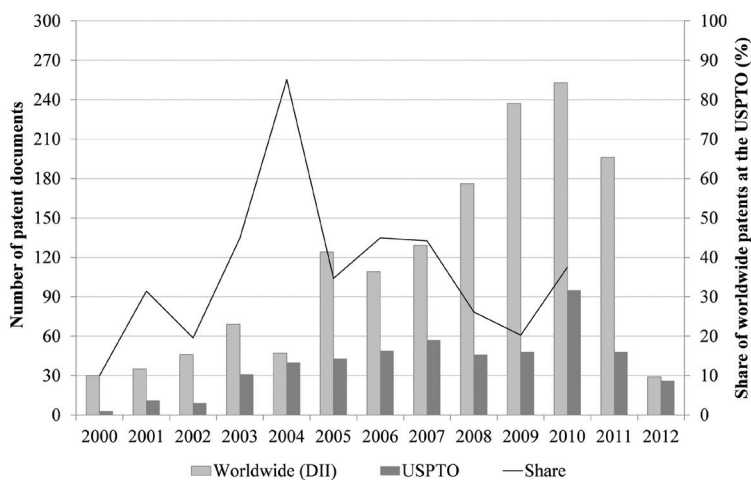


Figure 1. Annual number of patent documents in nanocellulose worldwide and at the USPTO and the relation between them from 2000 to 2010. Source: USPTO and DII.

this might be consequence of knowledge advances in nanocellulose and the promising forecasted market size for this nanomaterial¹⁻⁵. For the values of 2011 and 2012, the decrease might be due to the regular unpublished period of 18 months for patent applications and eventual database indexing delays³⁵. Interestingly, the patents filed at the USPTO accounted for 36%, on average, of the total number of patent applications in nanocellulose worldwide, which peaked in 2004 when the value accounted for 85%. This result highlights the importance of the USA market for nanocellulose.

The priority countries of the patents filed at the USPTO consisted mainly of major producers of cellulose and pulp (the USA, Japan, India, Finland, Sweden, Germany, Brazil)³⁶, as shown in Figure 2. Obviously, the USA had home advantage in the number of patents accumulated in the period of analysis¹⁹, but the presence of other countries/regions suggests that they were also concerned about the American market. The absence of Canada, an important country for the development of nanocellulose⁵, may be consequence of the recurrent practice of their assignees, which prefer to file their patents firstly at the USPTO and then extend the protection to Canada¹².

Medical, dental or toilet preparations (A61K) accounted for 51% of the patents filed at the USPTO, which was not considered as a relevant technological topic pointed for nanocellulose⁶. In part, it can be explained by the fact that the technological development of this subject decreased in 2008-2011^v period, as shown in Figure 3^{vi}, and the nanomaterial is not the focus of development. The number of patent documents regarding the use of nanocellulose in preparation of food and beverage (A23L) also declined in the same period. On the other hand, a considerably increase between 2004-2007 and 2008-2011 periods was observed

^v Despite the fact that not all patent documents from 2011 were indexed when the search was conducted, it seems the scenario of patenting will not change the trends observed.

^{vi} See the complete meaning of each technological subject at www.wipo.int/ipcpub/

to technological subject related to materials engineering and processes. The number of patent of the subject D21H (pulp composition or paper treatment) grew 533%, which might be consequence of pulp industry interests, while C08B (polysaccharides) raised 236%, suggesting that the nanomaterial become the focus of development. Layered products (B32B) and macromolecular compositions (C08L) rose, respectively, 164% and 90.9% between the same periods and these subjects can be associated to nanocellulose surface modifications and composite materials, respectively. Furthermore, fermentation and enzyme-using processes (C12P) increased 100% and, in part, it is associated to bacterial cellulose or nanocellulose as a secondary product of biofuels production.

The mined terms concerning types of nanocellulose could be extracted to the top technological subject, as can be seen from Table 1, in spite of the low number of patent documents found. Bacterial cellulose was associated to all subjects, but as a preferred type to A61K, A61P, A61L, C12P and A23L. These can be consequence of its peculiar characteristics when compared to nanocelluloses obtained from top-down approaches: bacterial cellulose has higher purity and capacity of water absorption². Moreover, the subject C12P corroborates with the bottom-up process to obtain bacterial celluloses.

Pulp compositions or paper treatments (D21H) was linked mainly to cellulose nanofibrils developments, as consequence of mechanical defibrillation of pulp. The area of polymers (C08L) and process of compounding (C08J) appeared with all types, suggesting that the developments of composite materials did not claimed a specific type of cellulose-based nanomaterials as reinforcement, despite its apparent importance to cellulose nanocrystals. However, there was no preference of type to layered products (B32B).

3.2. Main assignees patenting profile

The main assignees shared together only 26% of the total patent documents analyzed, as shown in Table 2. Patenting behavior across these assignees may differ due to their own interests and strategies^{12,17}. At least one

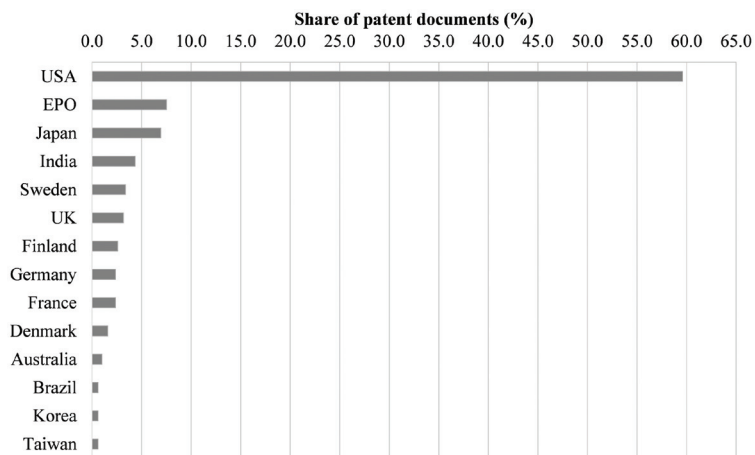


Figure 2. The main priority countries of patent documents filed at the USPTO. Source: DII.

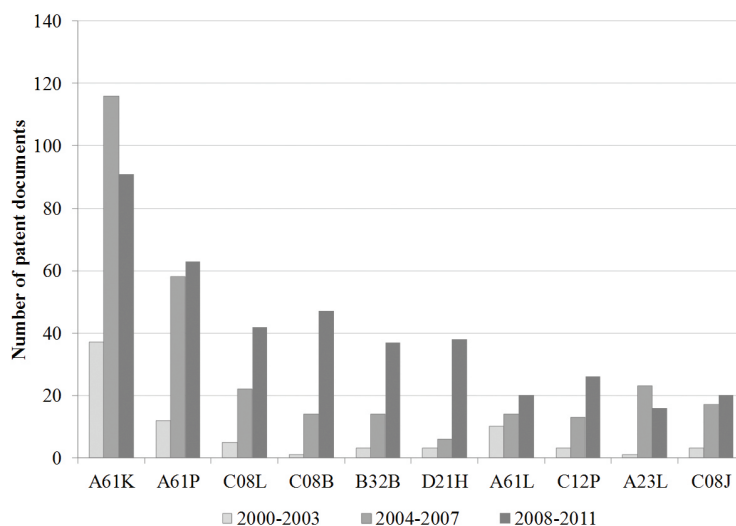


Figure 3. Evolution of the top 10 technological subjects. Source: USPTO.

Table 1. Types of nanocellulose to the top technological subject.

Technological subject	BC	CF	CC
A61K: medical, dental or toilet preparations	10	1	2
A61P: therapeutic activity of preparations	4		1
C08L: macromolecular compositions	5	7	6
C08B: polysaccharides	4	8	3
B32B: layered products	3	6	1
D21H: pulp composition, paper treatment (coatings, etc.)	1	12	
A61L: sterilizing, aspects of bandage, dressing or surgical article	5	1	1
C12P: fermentation or enzyme-using processes	14	2	
A23L: preparation of food or beverage	2		
C08J: processes of compounding, after-treatment	1	2	3

Source: USPTO.

application addressed by patents of each assignee could be identified from key terms extracted, as shown in Table 2, even though the number of patents associated to that application in most of the cases was low when compared to the total of patents of that assignee. In many cases, the type of nanocellulose involved could also be recognized,

except for the pharmaceutical-related companies (Pfizer, Randbaxy Lab., Aurobindo Pharma. and Teva Pharma.) and for FMC Co., Kimberly-Clark and Swetree Tech. The lack of information about the type of nanocellulose may be part of their patenting strategy¹² or even issues regarding the methodological procedures adopted in this study. However,

Table 2. Main assignee profile compiled from patent documents in nanocellulose.

Assignees (country)	#P*	Main technol. Subject (Simplified IPC code)	# P	Applications	# P	(#P) Type of nanocellulose**		
						BC	CC	CF
Procter & Gamble (USA)	24	C11D: detergent, soaps, soap making, glycerol.	19	Detergent composition	6	3	-	-
FPInnovations (Canada)	14	C08B: polysaccharides	8	Adhesives	1	-	6	1
		C08L: macromolecular compositions.	8	Nanocomposites	1	-	-	-
				Hydrogels	1	-	-	-
Pfizer (USA)	14	A61K: medical, dental or toilet preparations.	14	Pharmaceutical composition	6	-	-	-
Stora Enso (Finland)	11	D21H: pulp composition, coating of papers, final treatment of paper	11	Coatings	1	-	-	7
FMC Co (USA)	10	A61K: medical, dental or toilet preparations	10	Coatings	1	-	-	-
				Cosmetics	1	-	-	-
				Food and beverage	1	-	-	-
Upm-Kymmene (Finland)	10	D21H: pulp composition, coating of papers, final treatment of paper	7	Drilling composition	1	-	-	1
Kimberly-Clark (USA)	9	A61F: medical devices (filters, prostheses, etc.)	7	Absorbent article	4	-	-	-
		H01M: conversion of chemical energy into electrical energy	3					
UT-Battelle (USA)	8	A61F: medical devices (filters, prostheses, etc.)	3	Nanocomposites	3	6	-	-
		A61K: medical, dental or toilet preparations	3					
Randbaxy Lab (India)	6	A61K: medical, dental or toilet preparations	6	Pharmaceutical comp.	3	-	-	-
Univ. Texas (USA)	6	C12N: micro-organisms or enzymes, culture media	4					
		C12P: fermentation or enzyme-using processes	4	Medical devices	1	3	-	-
Aurobindo Pharm. (India)	5	A61K: medical, dental or toilet preparations	5	Pharmaceutical composition	2	-	-	-
Sofradim (France)	5	A61L:sterilizing, aspects of bandage, dressing or surgical article	4					
		C12P: fermentation or enzyme-using processes	4	Medical devices	5	3	-	-
Swetree Tech (Sweden)	5	D21H: pulp composition, coating of papers, final treatment of paper	4					
		C08B: polysaccharides	4	Nanocomposites	1	-	-	-
Teva Pharma (USA)	5	A61K: medical, dental or toilet preparations	5	Pharmaceutical composition	2	-	-	-

*#P means the total number of patent documents; **BC = Bacterial Cellulose; CC = Cellulose Nanocrystals; CF = Cellulose Nanofibrils; Source: USPTO and WIPO.

an improvement in the analysis was possible because the mined terms may suggest trends of type of nanocellulose preferences for some assignees.

The nanocellulose patents from Procter & Gamble, whose interest lies in consumer goods³⁷, were associated to detergent compositions, a result that can be seen by both the technological subject and application terms. Bacterial cellulose was found in only 12.5% of its patent documents. In turn, the patent documents from FMC Co., a chemical manufacturing company³⁸, were related to a broad portfolio of applications, including medical, dental or toiletry preparations (A61K), cosmetics, food and beverage. Additionally, FMC Co. produces Avicel[®], a product based on microcrystalline cellulose commonly used in pharmaceutical compositions, which also emphasizes the company's nanocellulose-related interests³⁹. Kimberly-Clark is a

company that produces paper-based consumer products for personal care use⁴⁰ and 77.8% of its patent documents were related to medical devices (A61F) while 44.4% indicated applications as absorbent articles.

The patents from FPInnovations, a research center working with sustainable forestry, pulp, paper and the packaging industry⁴¹, were mainly addressed to polysaccharides and derivatives (C08B), as well as macromolecular compositions (C08L). The type of nanocellulose involved in its technologies included both cellulose nanocrystals and cellulose nanofibrils, while the main mapped applications were adhesives, hydrogels and nanocomposites. The University of Texas assigned patent documents mainly to medical device applications, using bacterial cellulose, and their technologies used fermentation processes, microorganisms and enzyme technologies.

UT-Battelle is a private non-profit company created by a partnership between the University of Tennessee and the Battelle Memorial Institute⁴². It showed a broad technological profile, including preparations for medical, dental or toilet purposes (A61K), as well as implantable filters, prostheses and other medical devices (A61F). It also had patents concerning processes for conversion of energy (H01M), which is aligned with its function of managing and operating the Oak Ridge National Laboratory for the U.S. Department of Energy⁴². The mined titles and abstracts also showed that 75% of their technologies included bacterial cellulose as a type of nanocellulose and interest in nanocomposites applications (37.5%).

The Finnish Stora Enso and UPM-Kymmene and the Sweden Swetree Technologies are companies linked to forest products, pulp and paper⁴³⁻⁴⁵, and their patent interests were associated to pulp compositions. However, the nanocellulose applications mapped were different: Stora Enso focused in coatings, UPM-Kymmene in drilling compositions and Swetree Tech. in nanocomposites. Cellulose microfibrils were present in technologies from the Finnish companies and they also co-assigned two patent applications concerning manufacturing nanostructured paper or board for printing application in the period of analysis. Finally, The French Sofradim Products⁴⁶ demonstrated

interests in fermentation processes (C12P) and methods or apparatus for sterilizing materials/objects (A61L) while the application claimed was clearly related to medical device (100%) using bacterial cellulose (66.7%).

3.3. Analysis of the highly cited patents from the sample

From the total of 500 US patent documents which comprises the sample analyzed, 329 (65.8%) received at least one citation until July 2014. Obviously, older documents trend to have more citation because there is a higher time span to cite them. The top 15 most cited patents had at least 19 citations and they included a variety of technological developments in nanocellulose, as can be seen from Table 3. Companies owning highly cited patents (Table 3) also were among the assignees with the greatest number of patents (Table 2). Most of the assignees from these highly cited patents are from the USA, except two from Japan (University of Tokyo and Sharp) and another from Switzerland (Roche). Moreover, universities seemed to have an important role in the development of new technologies in nanocellulose since they signed in one third of the patents analyzed. This is in agreement with the other authors' statement about the high scientific dependence of the technological advances in nanocellulose^{5,10,12}.

Table 3. Top 15 most cited patents profile from the sample analyzed.

Patent Number/ Assignee name/ Priority year	Times cited	Technological Subjects	Application/ Type of nanocellulose
US2003216492 New Ice Ltd. (USA) 2002	52	B05D: surface treatments with liquids B29C: shaping of plastics, after-treatment B29L: particular articles B32B: layered products B65D: containers for storage C08J: processes of compounding, after-treatment C08K: inorganic or organic substances as compounding C08L: macromolecular compositions D21J: manufacture of articles from cellulosic suspension	Compostable containers -
US2005079386 Univ. of Texas (USA) 2003	40	B32B: layered products G02F: devices controlled by optical characteristics G09G: circuits for control of indicating devices	Electronic paper -
US2007178569 Univ. of Massachusetts (USA) 2006	39	B09B: disposal of solid waste C10L: fuels C12M: apparatus for enzymology C12N: micro-organisms or enzymes, culture media C12P: fermentation or enzyme-using processes	Biofuels -
US2006094320 Kimberly-Clark (USA) 2004	39	A61F: medical devices (filters, prostheses, etc.) B32B: layered products D04H: making textile, apparatus therefore	-
US2004241436 Univ. of California (USA) 2002	29	A61F: medical devices (filters, prostheses, etc.) B32B: layered products C12N: micro-organisms or enzymes, culture media D01D: methods or apparatus to manufacture filaments	-
US2002136755 Kimberly-Clark (USA) 2000	26	A01N: biocides, disinfectants, repellants, preservation A61F: medical devices (filters, prostheses, etc.) A61K: medical, dental or toilet preparations A61L: sterilizing, aspects of bandage, dressing or surgical article	Absorbent article Bacterial cellulose

Source: DII and USPTO.

Patent Number/ Assignee name/ Priority year	Times cited	Technological Subjects	Application/ Type of nanocellulose
US2007027108 CP Kelco Inc. (USA) 2005	24	A01N: biocides, disinfectants, repellants, preservation A23L: preparation of food or beverage A61K: medical, dental or toilet preparations C08B: polysaccharides C08L: macromolecular compositions C12P: fermentation or enzyme-using processes	- Bacterial cellulose
US2003130636 Kimberly-Clark (USA) 2001	22	A41B: textile clothes A61F: medical devices (filters, prostheses, etc.) A61K: medical, dental or toilet preparations A61L: sterilizing, aspects of bandage, dressing or surgical article	Absorbent article -
US2004123402 Procter & Gamble (USA) 2001	22	A61K: medical, dental or toilet preparations A61Q: use of cosmetics or toilet preparations	-
US2005013863 Depomed (USA) 2003	21	A61K: medical, dental or toilet preparations A61P: therapeutic activity of preparations	Pharmaceutical composition -
US2008190013 North Carolina State Univ. (USA) 2007	20	C08J: processes of compounding, after-treatment C10G: production of hydrocarbon mixture C10L: fuels D21C: production of cellulose, pulping liquors	Biofuels Cellulose nanofiber
US2010210501 Procter & Gamble (USA) 2008	19	C11D: detergent, soaps, soap making, glycerol	Detergent Bacterial cellulose
US2010233481 Univ. of Tokyo Sharp (Japan) 2007	19	B32B: layered products C08B: polysaccharides C12P: fermentation or enzyme-using processes D01F: chemical features of filaments D21H: pulp composition, paper treatment (coatings, etc.)	- Cellulose nanofiber
US2011223619 Genentech Inc (USA) Roche (Switzerland) 2007	19	A61K: medical, dental or toilet preparations A61P: therapeutic activity of preparations B23B: turning, boring C07D: heterocyclic compounds C12Q: measuring process (enzymes) G01N: determining chemical or physical properties of materials	Pharmaceutical composition Bacterial cellulose
US2004055095 Procter & Gamble (USA) 2001	19	A61K: medical, dental or toilet preparations A61Q: use of cosmetics or toilet preparations	-

Source: DII and USPTO.

The content of the document could be understood by the technological subject and, eventually, by the terms of application and type of nanocellulose mined from title and abstract. For instance, the highest cited patent refers to compostable containers (application and B65D code) made by polymers (C08L, C08K) as layered products (B32B) and many processes are involved (see subjects B05D, B29C, C08J and D21J). Indeed, it belongs to an American company that also owns a trademark to biodegradable containers⁴⁷, however, it was not possible to identify the type of nanocellulose used in the product. Another example would be the second most cited document which requested an electronic paper comprising layered product (B32B), a device controlled by optical characteristics (G02F) and circuits (G09G).

Seven patent documents were related to medical, dental or toilet preparations (A61K). Two of them claimed pharmaceutical composition and another two absorbent article as application of the invention. Bacterial cellulose emerged as the preferred type of nanocellulose in three documents. Five documents assigned layered products (B32B) and cellulose nanofiber was claimed as nanomaterial in one of those documents. In the case of medical devices (A61F), which appeared as subject in four documents, bacterial cellulose was identified in one document and, again, absorbent article was the application found in two documents. Definitely, bacterial cellulose was the unique type that could be identified in two patent documents (US2007027108 and US2010210501) and fermentation process (C12P) was present in three documents, even

though one claimed biofuels as application. On the other hand, cellulose nanofiber was claimed in only one document (US2010233481), but it is a generic name and do not foresee the manufacture approach of the nanocellulose used. Concerning biofuels, two technologies linked nanocellulose to this application (US2007178569 and US2008190013), which might be read as it would be a secondary product of the invention.

4. Conclusion

This paper has given an account of technological monitoring of nanocellulose, focusing on the patenting activity in patent applications filed at the United States Patent and Trademark Office (USPTO) from 2000 to 2012. The evolution and priority countries were assessed while a profile containing the assignee name, priority year, technological subject and terms mined from title and abstract were used to evaluate interests of main assignees and top cited patents.

A rapid annual growth of patenting activity was observed in the period, accompanying the worldwide trend. The share of patent applications in the USPTO resembled 36% in the annual average of world applications, and at least partially, it is due to the forecasted USA nanocellulose market size and the emergence of developments of the nanomaterial. The patents filed at the USPTO were not only from USA assignees, but also from other countries and regions, mainly from Europe, Japan and India. The highlighted assignees were from USA, Japan, Finland, India, Canada, France, and Sweden. The outcomes also showed the scientific dependence of technologies in nanocellulose, because universities and research institutes were present among the main assignees and the most cited documents.

Medical, dental or toilet preparations accounted for 51% of the patents filed at the USPTO. However, this subject stated decreasing from 2008. Technological subjects regarding materials and processes highly increased after 2008, though. Both assignee and top cited document profiles provided evidences of using nanocellulose in a variety of

applications, including cosmetic and drugs formulations, biodegradable composition, absorbent articles, methods to produce nanocellulose and products containing it, and technologies regarding biofuels and nanocellulose as a secondary product. Composite materials and surface treatments trend to increase as subjects to nanocellulose technological developments in coming future. In spite of technologies involving bacterial cellulose appear to be more developed, technologies containing cellulose nanofibrils or nanocrystals has emerged consistently in recent years probably as consequence of establishing plant pilots worldwide and at the USA.

The data and text-mining tools and complementary procedures used in this study to compile and support the analysis of the indicators have brought advances in highlighting type of nanocellulose and their applications. Moreover, text-mining indicators could complement the technological subjects and enhance the monitoring assessment. However, limitations to be overcome in future research could also be observed. For example, it was not possible to point out the type of nanocellulose for all patent documents, and when it was successful, the number of patent documents involved was low. The absence of information may also be part of assignees' strategy or that the nanomaterial is not the focus of development. Even so, applying the text mining approach for other parts of the patent documents, such as the detailed description of the invention and/or the patent claims, could improve the outcomes in future research.

Acknowledgements

The authors are grateful to the Brazilian National Council for Technological and Scientific Development (process number 160087/2011-2), to the São Paulo Research Foundation (process number 2012/16573-7) and to the Graduate Program in Materials Science and Engineering at the Federal University of São Carlos for supporting this work. We also would like to acknowledge the anonymous reviewers for their rich recommendations and contributions.

References

1. Siqueira G, Bras J and Dufresne A. Cellulosic bionanocomposites: a review of preparation, properties and applications. *Polymers*. 2010; 2(4):728-765. <http://dx.doi.org/10.3390/polym2040728>.
2. Klemm D, Kramer F, Moritz S, Lindström T, Ankerfors M, Gray D, et al. Nanocelluloses: a new family of nature-based materials. *Angewandte Chemie International Edition*. 2011; 50(24):5438-5466. <http://dx.doi.org/10.1002/anie.201001273>. PMID:21598362
3. Moon RJ, Martini A, Nairn J, Simonsen J and Youngblood J. Cellulose nanomaterials review: structure, properties and nanocomposites. *Chemical Society Reviews*. 2011; 40(7):3941-3994. <http://dx.doi.org/10.1039/c0cs00108b>. PMID:21566801
4. Charreau H, Foresti ML and Vazquez A. Nanocellulose patents trends: a comprehensive review on patents on cellulose nanocrystals, microfibrillated and bacterial cellulose. *Recent Patents on Nanotechnology*. 2013; 7(1):56-80. <http://dx.doi.org/10.2174/187221013804484854>. PMID:22747719
5. Milanez DH, Amaral RM, Faria LIL and Gregolin JAR. Assessing nanocellulose developments using science and technology indicators. *Materials Research*. 2013; 16(3):635-641. <http://dx.doi.org/10.1590/S1516-14392013005000033>.
6. Research and Markets. *The global market for nanocellulose to 2017*. United Kingdom; 2014. Available from: <www.researchandmarkets.com/reports/2271936/the_global_market_for_nanocellulose_to_2017>. Access in: 30/05/2014.
7. Durán N, Lemes AP and Seabra AB. Review of cellulose nanocrystals patents: preparation, composites and general applications. *Recent Patents on Nanotechnology*. 2012; 6(1):16-28. <http://dx.doi.org/10.2174/187221012798109255>. PMID:21875405
8. TAPPI. *Roadmap for the development of international standards for nanocellulose*. 2011. 36 p. Available from: <<http://www.tappinano.org/pdf/RoadmapforNanocelluloseStandards.pdf>>. Access in: 30/05/2014.
9. Siró I and Plackett D. Microfibrillated cellulose and new nanocomposite materials: a review. *Cellulose*. 2010; 17(3):459-494. <http://dx.doi.org/10.1007/s10570-010-9405-y>.

10. Moed HF, Glänzel W and Schmoch U. *Handbook of quantitative science and technology research*. New York: Kluwer Academic Publishers; 2004. 800 p.
11. Mogee ME. Patents and technology intelligence. In Ashton WB and Richard AK. *Keeping abreast of science and technology: technical intelligence for business*. Columbus: Battelle Press; 1997.
12. Organisation for Economic Co-operation and Development - OECD. *OECD Patent Statistics Manual. Patent Manual*. 2009. <http://dx.doi.org/10.1787/9789264056442-en>.
13. Salerno M, Landoni P and Verganti R. Designing foresight studies for Nanoscience and Nanotechnology (NST) future developments. *Technological Forecasting and Social Change*. 2008; 75(8):1202-1223. <http://dx.doi.org/10.1016/j.techfore.2007.11.011>.
14. Abbas A, Zhang L and Khan SU. A literature review on the state-of-the-art in patent analysis. *World Patent Information*. 2014; 37:3-13. <http://dx.doi.org/10.1016/j.wpi.2013.12.006>.
15. Fujii A, Iwayama M and Kando N. Introduction to the special issue on patent processing. *Information Processing & Management*. 2007; 43(5):1149-1153. <http://dx.doi.org/10.1016/j.ipm.2006.11.004>.
16. Tseng Y-H, Lin C-J and Lin Y-I. Text mining techniques for patent analysis. *Information Processing & Management*. 2007; 43(5):1216-1247. <http://dx.doi.org/10.1016/j.ipm.2006.11.011>.
17. Fleisher CS and Bensoussan BE. *Strategic and competitive analysis: methods and techniques for analyzing business competition*. New Jersey: Prentice Hall; 2002. 457 p.
18. Albuquerque EM. (Coord.). Patent activity in Brazil and abroad. In: São Paulo Research Foundation - FAPESP. *Science, technology and innovation indicators in São Paulo*. São Paulo; 2010. v. 1. Available from <<http://www.fapesp.br/indicadores/2010/volume1/cap5.pdf>>. Access in 24/07/2014.
19. Campus SK and Criscuolo P. The “home advantage” effect and patent families: a comparison of OECD triadic patents, the USPTO and the EPO. *Scientometrics*. 2006; 66(1):23-41. <http://dx.doi.org/10.1007/s11192-006-0003-6>.
20. Wanner L, Baeza-Yates R, Brüggemann S, Codina J, Diallo B, Escorsa E, et al. Towards content-oriented patent document processing. *World Patent Information*. 2008; 30(1):21-33. <http://dx.doi.org/10.1016/j.wpi.2007.03.008>.
21. Yang Y, Akers L, Klose T and Barcelon Yang C. Text mining and visualization tools: impressions of emerging capabilities. *World Patent Information*. 2008; 30(4):280-293. <http://dx.doi.org/10.1016/j.wpi.2008.01.007>.
22. Yoon B and Park Y. A text-mining-based patent network: analytical tool for high-technology trend. *The Journal of High Technology Management Research*. 2004; 15(1):37-50. <http://dx.doi.org/10.1016/j.hitech.2003.09.003>.
23. Yoon B and Park Y. A systematic approach for identifying technology opportunities: keyword-based morphology analysis. *Technological Forecasting and Social Change*. 2005; 72(2):145-160. <http://dx.doi.org/10.1016/j.techfore.2004.08.011>.
24. Yoon J and Kim K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics*. 2012; 90(2):445-461. <http://dx.doi.org/10.1007/s11192-011-0543-2>.
25. Yoon J and Kim K. An analysis of property–function based patent networks for strategic R&D planning in fast-moving industries: the case of silicon-based thin film solar cells. *Expert Systems with Applications*. 2012; 39(9):7709-7717. <http://dx.doi.org/10.1016/j.eswa.2012.01.035>.
26. Yoon J, Park H and Kim K. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics*. 2013; 94(1):313-331. <http://dx.doi.org/10.1007/s11192-012-0830-6>.
27. Chen H, Roco MC, Li X and Lin Y. Trends in nanotechnology patents. *Nature Nanotechnology*. 2008; 3(3):123-125. <http://dx.doi.org/10.1038/nnano.2008.51>. PMID:18654475
28. Wang G and Guan J. Value chain of nanotechnology: a comparative study of some major players. *Journal of Nanoparticle Research*. 2012; 14(2):702. <http://dx.doi.org/10.1007/s11051-011-0702-7>.
29. Milanez DH, Faria LIL, Amaral RM, Leiva DR and Gregolin JAR. *Patents in nanotechnology: an analysis using macro-indicators and forecasting curves*. *Scientometrics*. 2014; 101(2):1097-1112. Available from: <<http://link.springer.com/article/10.1007%2Fs11192-014-1244-4>>. Access in: 30/05/2014.
30. United State Patent and Trademark Office. Available from: <www.uspto.gov>. Access in: 30/05/2014.
31. Thomson Reuters. *Derwent Innovations Index*. 2014. Available from: <http://apps.webofknowledge.com/DIIDW_GeneralSearch_input.do?product=DIIDW&SID=4ApADyNtOVIQub2tsXM&search_mode=GeneralSearch>. Access in 30/05/2014.
32. Search Technology. *Vantage Point Manual*. 2006.
33. WIPO. *Guide to using patent information*. 2013. Available from: <http://www.wipo.int/export/sites/www/freepublications/en/patents/434/wipo_pub_1434_03.pdf>. Access in: 24/05/2014.
34. Ujihara K. *GetIPDL: Patent Downloader*. 2010. Available from: <www.getipdl.net>. Access in: 30/05/2014.
35. Milanez DH, Macedo TD, Amaral RM, Faria LIL and Gregolin JAR. Assessing an interval of confidence to compile time-dependent patent indicators in nanotechnology. In: Proceedings of the 14th International Society of Scientometrics and Informetrics Conference; 2013; Vienna, Austria. Vienna: Austrian Institute of Technology; 2013. v. 2. p. 1877-1880.
36. Brazilian Association for Cellulose and Pulp - BRACELPA. *Data from sector*. Available from: <<http://www.bracelpa.org.br/bra2/sites/default/files/estatisticas/booklet.pdf>>. Access in: 30/05/2014.
37. Procter & Gambler. 2014. Available from: <www.pg.com>. Access in: 30/05/2014.
38. FMC. 2014. Available from: <www.fmctechnologies.com>. Access in: 30/05/2014.
39. FMC. *Material safety data sheet: Avicel® PH Microcrystalline Cellulose*. 2014. Available from: <<http://www.fmcbiopolymer.com/Portals/Pharm/Content/Docs/avicelphmsds.pdf>>. Access in: 30/05/2014.
40. Kimberly-Clark. 2014. Available from: <<http://www.kimberly-clark.com>>. Access in: 30/05/2014.
41. FPIInnovations. 2014. Available from: <www.fpinnovations.ca>. Access in: 30/05/2014.
42. UT-Battelle. 2014. Available from: <<http://www.ut-battelle.org>>. Access in: 30/05/2014.
43. Enso S. 2014. Available from: <www.storaenso.com>. Access in: 30/05/2014.
44. UPM-Kymmene. 2014. Available from: <www.upm.com>. Access in: 30/05/2014.
45. Technologies S. 2014. Available from: <www.swetree.com>. Access in: 30/05/2014.
46. Lyonbiopole. *Sofradim Products*. 2013. Available from: <www.lyonbiopole.com/exchange/entreprise-53-sofradim-production.html>. Access in: 30/05/2014.
47. Trademarkia. *Verdepac*. 2014. Available from: <<http://www.trademarkia.com/verdepac-77313089.html>>. Access in: 24/06/2014.