



Scientific publications on innovation: an analysis using text mining

Cicero Aparecido Bezerra^[a], André José Ribeiro Guimarães^[b]

^[a] Doctorate in Production Engineering at the Universidade Federal de Santa Catarina (UFSC). Professor at the Universidade Federal do Paraná (UFPR). E-mail: cicero.bezerra@ufpr.br

^[b] Master of Science, Management and Information Technology at the Universidade Federal do Paraná (UFPR). Co-owner of the Abyssal Design. E-mail: andrejrg@gmail.com

Abstract

The scientific literature on innovation, associated with the business area, has shown not only, strong growth in recent decades, but also diversity in associated topics. In this sense, the objective of this study is to present the terms most used in scientific papers abstracts, where the word 'innovation' is part of the title, over the decades formed between the years 1960 and 2014, as well as the dynamic existing between them. To do this, we used the text mining technique in 40,383 abstracts extracted from some databases. Among the findings, it is emphasized that over the decades the articles has focused, increasingly, on pragmatic aspects, over the theoretical. Similarly, it was noted that the focus of research in innovation has moved from large companies to small organizations. Furthermore, we highlight the case studies as the main research method used. But in recent years, has increased the studies using quantitative methods.

Keywords: Innovation. Text mining.

Introduction

Formal academic papers on innovation can be traced back to 1888 (LAZZAROTTI; DALFOVO; HOFFMANN, 2011, p. 122), but even based on the assumption that innovation is a phenomenon that has always existed, research on the topic gained real importance in the late twentieth century – to Durisin, Calabretta, Parmeggiani (2010, p. 427) the 1990s marked the maturity of research in this area. Later, studies of innova-

tion processes and implementation of ideas continued to rise in response to changes in the socioeconomic environment.

The increase in the number of researches on innovation (specifically related to organizational studies) has also been accompanied by a plurality of subjects involved. For example, according to the classification proposed by Orlickas (2011, p. 75-101), for administrative techniques, it is possible to find several articles that address innovation related to reengineering, such as the Pellicelli, Meo, Cioffi (2012, p. 291), in which they argue that the reengineering process is an innovative approach to the organization, for it improves productivity, efficiency and costs; benchmarking, where Dembowski (2013, p. 10) notes that the understanding and evaluation of the organization regarding best practices helps improve their own performance; empowerment, where Çakar and Er-turk (2010, p. 348) realize the importance of encouraging employees to propose new and innovative ideas and how this impacts positively on the company's performance indicators; learning organization, in which the acceptance of the organization as a complex dynamic system, translated into active leadership, promotes innovation, learning and adaptation (GEER FRAZIER, 2014, p. 113). The same scenario is found when the term 'innovation' is faced with some of the management models proposed by Have, Have, Stevens, Elst, Pol Coyne (2003): Balanced Scorecard (BSC), where Frezatti, Bido, Cross, Camargo (2014, p. 390) conclude that the relationship between the BSC and the innovation process is marked by significant efforts to treat innovation as activities turned into results over time; ABC costing model, where Alcouffe and Guedes (2008, p. 55-56) propose a model in which the relationship between some innovative features (specifically, relative advantage, compatibility and complexity) and the adoption of the ABC model is expressed; the Hofstede's Cultural Dimensions that are employed by Halkos and Themes (2013, p. 674) to explain that the cultural values of countries have direct influence on how innovation policies are implemented.

Thus, considering the diversity of researches studying innovation in the context of administrative sciences, the aim of this study is to extract the most used terms in scientific article summaries in which the word 'innovation' is part of the title, throughout the decades between the years of 1960 and 2014, and the dynamics between them, in some databases that emphasize studies associated with administration.

The study is divided into theoretical framework of innovation and text mining, methodological procedures employed, presentation and analysis of results, conclusions and adopted references.

Theoretical framework

This section discusses the basic theoretical definitions of innovation and the context in which this term has been used in researches related to management. Moreover, it is necessary to detail, theoretically, the analytical technique employed in order to facilitate understanding of the analysis process. It is important to clarify that the intention is not to exhaust the subject, but rather to present the theoretical basis that is minimally necessary for an understanding of the study.

Innovation

Although this word is present in the popular vocabulary, media, business circles and the establishment of public policies, innovation, throughout its history, had its concept widely discussed and treated as the object of various theories and definitions (GODIN, 2008, p. 43).

There is convergence on the fact that innovation is key to achieving sustainable competitive advantage in organizations, particularly in small and medium-sized enterprises (CARRER; PLONSKI; CARRER; OLIVEIRA, 2010, p. 18; REJEB; MOREL-GUIMARÃES; BOLY; ASSIÉLOU, 2008, p. 838). Innovation is key to progress, especially in times of crisis, given its capacity to competitively strengthen organizations, promoting the economic growth of a country and the competitiveness of its economy (NIDUMOLU; PRAHALAD; RANGASWAMI, 2009, p. 58).

One of the first (and most replicated) definitions came from Schumpeter (1934, p. 66), in which innovation ranges from the introduction of new products or services to new means of production, markets, suppliers and even to the establishment of a new business organization. Even though innovation is a comprehensive concept both etymologically and historically, many researchers consider it to be necessarily technology, to the point that many studies of 'technological innovation' simply use the term 'innovation' (GODIN, 2008, p. 5).

The concept of innovation may have different approaches in different contexts, and the choice of one depends on the measurement objectives and on the analysis process of the research in question (Organisation for Economic Co-operation and Development, 2005, p. 19). For example, Vacaro, Jansen, van den Bosch and Volberda (2012, p. 29) highlights the concept of managerial innovation, which deals with the creation and implementation of management practices, processes, organizational structures and techniques, with a high degree of novelty compared to state of the art, so that organizational goals are achieved more quickly. In turn, Liao, Chang, Hu, and Yueh (2012, p. 59) propose the definition of organizational innovation, which is formed by elements associated with the innovation of products, markets and even behavior and strategy. Prestes and Bozac (2012, p. 2) understand that innovation, when related to goods and services, can be divided into incremental (improvement and continuous modifications) or radical (technological breakthrough); it may also be associated with processes (changes in production systems) and the organization itself, when changing communication and reward systems, and even work assignments.

Anyway, while organizations have traditionally adopted internal and closed strategies to develop innovation, virtually no interaction with the external environment, dialectically a new scenario, has emerged in recent decades. The number of organizations that acquire external technology to complement its internal knowledge and accelerate the innovation process has grown increasingly (LICHTENTHALER, 2011, p. 75). This notion of interaction with the external environment in which the organization can and should use external and internal knowledge to accelerate the innovation process is part of the concept of open innovation (CHESBROUGH, 2003, p. xxiv). Grizendi (2011, p. 51) believes that through this, the company can leverage its innovative capacity, maximizing the use of existing opportunities outside the organization, aligning their technology infrastructure to other external structures. Open innovation is a trend that

has been gaining strength in many sectors, notably the industrial, as a sustainable way to do it (LICHTENTHALER, 2011, p. 89).

In this scenario it is interesting (even necessary) to evaluate the formation pattern in studies involving innovation, given its importance, both in the economic context (MUSTEEN; AHSAN, 2013, p. 423; LASAGNI, 2012, p. 311; MIERES; SÁNCHEZ; VIJANDE, 2012, p. 403) as in the social one (AUTANT-BERNARD; CHALAYE, MANCA; MORENO; SURIÑACH, 2010, p. 199; CARRER; PLONSKI; CARRER; OLIVEIRA, 2010, p. 17; BUCIC; NGO, 2012, p. 3). One of techniques capable of identifying unknown patterns in a large body of data, specifically textual nature, is the text mining (CHEN, 2012, p. 493). Noteworthy is the fact that this technique has particular application in scientific research (MOEZZI; GOINS, 2011, p. 171).

Text mining

In general, it can be said that the primary goal of text mining is to identify hidden knowledge in text and present it in a concise, consistent format (FARO; GIORDANO; SPAMPINATO, 2011, p. 62). According to Wang, Chang and Kao (2010, p. 494), it is a conceptual structure developed for obtaining valuable information in large quantities of unstructured documents. Recent studies, using this technique, have been made in order to identify subjects, summarized content, and to view associations between documents (YOON; PHAAL; PROBERT, 2008, p. 54).

To Gajzer (2010, p. 223-224), is possible to identify four distinct steps in the text mining process:

- Transformation: it is the conversion of the original document to text format, format-free, eliminating and replacing (if applicable) symbols and characters irrelevant to the analysis;
- Tokenization: consists of the separation of words. According to Soares, Prati and Monard (2008, p. 3) this process is responsible for identifying important elements of a text, breaking the continuous stream of characters, excluding textual elements (scores, syllables, marks and numbers);
- Stemming: to Soares, Prati and Monard (2008, p. 4) it is the linguistic normalization in which the variant forms of a word are reduced to a common form known as stem from the extraction of prefixes and suffixes that make up the tokens;
- Matrix frequencies: to Fan, Wallace, Rich and Zhang (2006, p. 79) this step categorizes the stems identifying its frequencies. The results of this stage allow inferences about the semantic distances and terms related to the stems.

Besides these stages, Soares, Prati and Monard (2008, p. 7) note that, in the stemming step, it is important to establish the frequency of tokens formed by adjacent tokens (n-gram), since there is the possibility, in this operation, of obtaining more relevant meanings than the analysis of isolated terms.

The use of text mining techniques has already been appointed as a trend in data analysis. To Weiss, Indurkha and Zhang (2010, p. 189-194), at least three applications can now be considered routine: summarization, in which it is possible to create summaries from different texts, but involving the same topics; active learning, from the

automatic document classification, and learning that emerges in the analysis of unstructured data.

Methodological procedures

From a scientific research classification proposed by Silva and Menezes (2005, p. 19-22), this study is framed as basic research by nature; quantitative with respect to the approach to the problem; exploratory in relation to the object and, taking into account the technical procedures, it can be classified as bibliographic research. Data was collected on January 26, 2015, from the Business Source Premier, Web of Knowledge, SCOPUS e Science Direct data bases.

It is important to clarify that (1) the investigated bases were chosen for convenience (for being at the disposal of the institution to which the authors are crowded), but taking into account those articles focusing on the business area / administration; (2) the years 1960 to 2014 were established as a time frame for the analysis; (3) the study does not intend to establish any kind of generalizations – therefore, the results should be taken only as an overview.

The search for articles that contained the term 'innovation' in the title without restriction period achieved a return of 80,817 records, distributed as follows: Business Source Premier (15,214), Web of Knowledge (26,035), SCOPUS (30,901) and Science Direct (8,667). Of the overall total, 40,383 resumes were validated, as follows (by decades): 1960-1969, with 120 articles; 1970-1979, with 446; 1980-1989, 1.502; 1990-1999, a total of 4,675; 2000-2009, 15,791 and; 2010-2014, with 17,849 articles. 'Valid summaries' means those that do not repeat title, author and year of publication, according to criteria adopted by EndNote® X6 software (Thomson Reuters, 2013). In addition, those without the year of publication have been removed, as well as those with identical resumes (albeit with different title, author or year of publication). The research followed the established protocol in Table 1:

Table 1 – Characteristics of the organization

Stage	Goals	Materials
1 Identification	Recover articles containing the term 'innovation' in the title.	Bases de dados <i>Business Source Premier, Web of Knowledge, SCOPUS e ScienceDirect</i>
2 Selection	Remove duplicate articles (found in more than one database) without summary or uninformed publication year.	EndNote® X6 (Thomson Reuters, 2013)
3 Conversion	Separate and convert articles, XML format (generated by EndNote® X6 software) to txt format.	PHP application developed by the authors
4 Text mining	Identifying and grouping terms found in the converted articles.	PreText 2 (Laboratory of Computational Intelligence, 2008).
5 Pre-analysis of results	Identify and remove <i>stems</i> naturally recurring (example: 'abstract', 'research', 'study', 'paper' ...).	
6 Analysis of results	Describe similarities and differences in the top 20 most common terms, in order: 1-gram, 2-gram and 3-gram.	

Source: adapted from Bezerra and Guimarães, 2014, p. 137.

It is emphasized that, in Step 4, there was a need to adjust parts of the source code of PreText 2 software (Laboratory of Computational Intelligence, 2008) once it registered the occurrence of some tokens, with the same radical, are forming distinct stems – a situation that had already been foreseen by authors such as Rijsbergen (1979, p. 4) and Martins, Godoy, Monard, Matsubara and Amandi (2003, p. 5). In addition, it was decided (by physical space issues) to report only the first twenty most frequent terms in each decade – if more than one term has the same frequency, they were ordered alphabetically.

It is important to clarify that once the evidence found are being presented, they will be subjected to empirical reflection, in order to provide information needed to produce hypotheses to be tested in further research.

Presentation and analysis of results

Initially, Figure 1 shows the quantity of items, with the term 'innovation' in the title, distributed over the years:

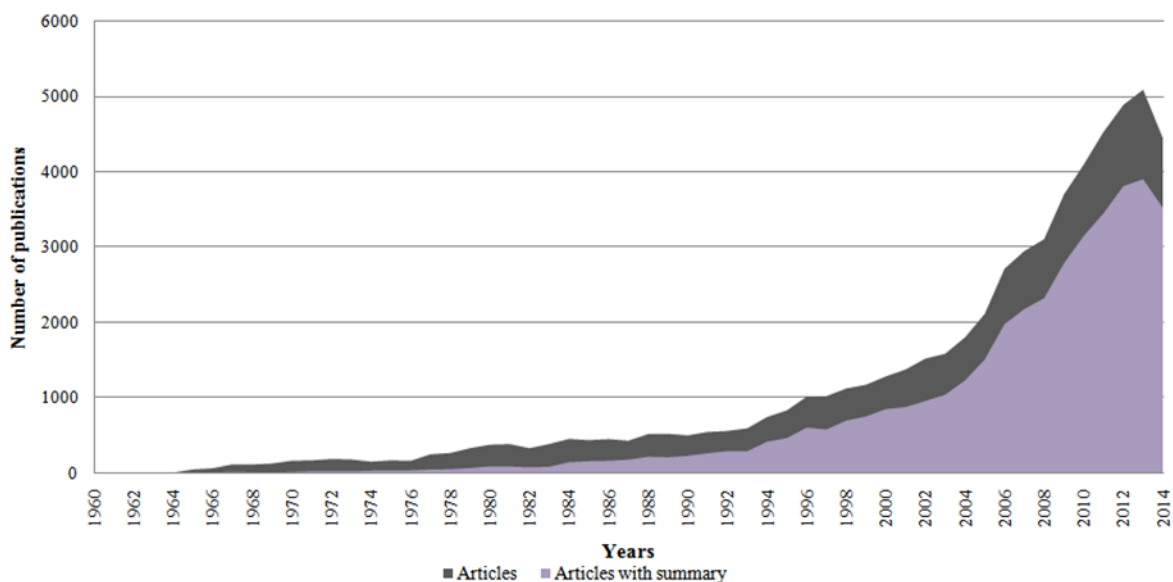


Figure 1 – Publications on innovation

Source: elaborated by the authors.

It is necessary to clarify that there were articles from the year 1897 - but the quantities of articles, in this period until 1960, are very small, so it was decided to not include them in figure. Still, it is worth noting that the first article found, according to the search criteria, was published under the title "Die Innovations-Verhältnisse von *Phaseolus coccineus* L. (= *Ph. multiflorus* Willd.)". The first article ever with summary ("Innovation in teaching elementary accounting") was found in 1936.

It is noticeable that the number of publications on the subject intensifies in the 1990s, and the first year to overcome a thousand records was 1994. However, it is only since 2000 that over a thousand articles with abstracts per year have been found. Alt-

though expected, the difference between the number of articles with summary (average of 365.83 with a standard deviation of 933.77) and the total number of items found (average of 602.08 with a standard deviation of 1,207.29) has decreased in the last decades. If in 1994 the ratio between the two groups was 46.35%, since 2008 it has remained above 78%. Another indication of the growth of scientific literature related to the theme is the increase in the average and the decrease in the standard deviation for the last 20 years studied (1994-2014). During this period, the average of all articles found was 2,854.37 (standard deviation 1599.82) while the average of articles with summary was 2,053.11 (standard deviation 1,409.73).

Before presenting the most recurrent terms during the decades analyzed, it is important to clarify how the data were arranged. The terms have been grouped into tables showing, in each column (from left to right): the analyzed decades; the stems classified by the occurrence of percentage relative to the amount of summaries of that decade; the position (i) in the current decade; the number of times (f) in which the steam was found on the summaries; the percentage in relation to the amount of abstracts found (%); the position that the steam occupied in the previous decade (i-1), and the arrows indicate whether the steam kept (↔), increased (↑) or decreased (↓) its position in relation to the current decade and; the number of times the steam has emerged in the top twenty positions (#).

Thus, Table 2 shows the individual stems (1-gram) that are the most frequent:

Table 2 – 1-gram

década	stems	i	f _i	%	i-1	#	década	stems	i	f _i	%	i-1	#
1960-1969	<i>technolog</i>	1	32	26,67		1	1990-1999	<i>technolog</i>	1	2070	44,28	↔1	4
	<i>proc</i>	2	29	24,17		1		<i>product</i>	2	1519	32,49	↔2	4
	<i>economic</i>	3	27	22,50		1		<i>proc</i>	3	1285	27,49	↔3	4
	<i>chang</i>	4	24	20,00		1		<i>system</i>	4	1140	24,39	↔4	4
	<i>develop</i>	5	19	15,83		1		<i>develop</i>	5	1075	22,99	↑15	4
	<i>busin</i>	6	18	15,00		1		<i>industri</i>	6	995	21,28		3
	<i>system</i>	7	18	15,00		1		<i>firm</i>	7	984	21,05	↓6	4
	<i>management</i>	8	17	14,17		1		<i>management</i>	8	742	15,87	↓8	4
	<i>growth</i>	9	16	13,33		1		<i>individual</i>	9	709	15,17	↓5	2
	<i>industr</i>	10	16	13,33		1		<i>model</i>	10	655	14,01	↑12	3
	<i>firm</i>	11	15	12,50		1		<i>information</i>	11	628	13,43	↑13	4
	<i>practic</i>	12	15	12,50		1		<i>users</i>	12	621	13,28		1
	<i>product</i>	13	13	10,83		1		<i>market</i>	13	554	11,85	↓11	3
	<i>techniqu</i>	14	13	10,83		1		<i>economic</i>	14	509	10,89	↓9	4
	<i>administrat</i>	15	12	10,00		1		<i>chang</i>	15	485	10,37	↓8	4
	<i>information</i>	16	12	10,00		1		<i>busin</i>	16	455	9,73		2
	<i>market</i>	17	12	10,00		1		<i>progr</i>	17	424	9,07		2
	<i>organizat</i>	18	12	10,00		1		<i>organizat</i>	18	413	8,83		3
	<i>structure</i>	19	12	10,00		1		<i>practic</i>	19	389	8,32		2
	<i>diffusion</i>	20	11	9,17		1		<i>work</i>	20	386	8,26		1
1970-1979	<i>technolog</i>	1	143	32,06	↔1	2	2000-2009	<i>technolog</i>	1	7313	46,31	↔1	5
	<i>industr</i>	2	136	30,49	↑10	2		<i>develop</i>	2	5643	35,74	↑5	5
	<i>organizat</i>	3	125	28,03	↑18	2		<i>product</i>	3	5256	33,28	↓2	5
	<i>proc</i>	4	124	27,80	↓2	2		<i>proc</i>	4	4571	28,95	↓3	5
	<i>product</i>	5	101	22,65	↑13	2		<i>system</i>	5	4476	28,35	↓4	5
	<i>develop</i>	6	96	21,52	↓5	2		<i>firm</i>	6	3973	25,16	↑7	5
	<i>firm</i>	7	83	18,61	↑11	2		<i>individual</i>	7	3744	23,71	↑9	3
	<i>chang</i>	8	82	18,39	↓4	2		<i>management</i>	8	3318	21,01	↔8	5
	<i>progr</i>	9	75	16,82		1		<i>accuracy</i>	9	3301	20,90		1
	<i>management</i>	10	67	15,02	↓8	2		<i>compani</i>	10	2873	18,19		1
	<i>economic</i>	11	66	14,80	↑3	2		<i>model</i>	11	2470	15,64	↓10	4
	<i>diffusion</i>	12	61	13,68	↑20	2		<i>knowledg</i>	12	2355	14,91		1

	1980-1989						2010-2014						
	Rank	Stem	Count	Avg	Change	Pos	Rank	Stem	Count	Avg	Change	Pos	
	13	<i>model</i>	59	13,23		1		<i>busin</i>	13	2354	14,91	↑16	3
	14	<i>system</i>	52	11,66	↓7	2		<i>market</i>	14	2236	14,16	↓13	4
	15	<i>techniqu</i>	48	10,76	↓14	2		<i>information</i>	15	2015	12,76	↓11	5
	16	<i>policy</i>	47	10,54		1		<i>economic</i>	16	1875	11,87	↓14	5
	17	<i>servic</i>	47	10,54		1		<i>performanc</i>	17	1863	11,80		1
	18	<i>information</i>	45	10,09	↓16	2		<i>policy</i>	18	1831	11,60		3
	19	<i>adoption</i>	44	9,87		1		<i>servic</i>	19	1821	11,53		2
	20	<i>government</i>	44	9,87		1		<i>design</i>	20	1777	11,25		2
	1	<i>technolog</i>	549	36,55	↔1	3		<i>technolog</i>	1	6345	35,55	↔1	6
	2	<i>product</i>	480	31,96	↑5	3		<i>proc</i>	2	5529	30,98	↑4	6
	3	<i>proc</i>	387	25,77	↑4	3		<i>firm</i>	3	5332	29,87	↑6	6
	4	<i>system</i>	351	23,37	↑14	3		<i>management</i>	4	4265	23,89	↑8	6
	5	<i>individual</i>	289	19,24		1		<i>accuracy</i>	5	3838	21,50	↑9	2
	6	<i>firm</i>	265	17,64	↑7	3		<i>model</i>	6	3467	19,42	↑11	5
	7	<i>management</i>	222	14,78	↑10	3		<i>knowledg</i>	7	3350	18,77	↑12	2
	8	<i>chang</i>	195	12,98	↔8	3		<i>high</i>	8	3138	17,58		2
	9	<i>economic</i>	186	12,38	↑11	3		<i>busin</i>	9	3118	17,47	↑13	4
	10	<i>high</i>	186	12,38		1		<i>performanc</i>	10	2763	15,48	↑17	2
	11	<i>market</i>	183	12,18		2		<i>system</i>	11	2613	14,64	↓5	6
	12	<i>model</i>	181	12,05	↑13	2		<i>product</i>	12	2580	14,45	↓3	6
	13	<i>information</i>	159	10,59	↑18	3		<i>practic</i>	13	2554	14,31		3
	14	<i>technical</i>	141	9,39		1		<i>industr</i>	14	2552	14,30		4
	15	<i>develop</i>	138	9,19	↓6	3		<i>market</i>	15	2417	13,54	↓14	5
	16	<i>effect</i>	138	9,19		1		<i>level</i>	16	2389	13,38		1
	17	<i>diffusion</i>	135	8,99	↓12	3		<i>servic</i>	17	2368	13,27	↑19	3
	18	<i>policy</i>	135	8,99	↓16	2		<i>design</i>	18	2323	13,01	↑20	3
	19	<i>design</i>	133	8,85		1		<i>economic</i>	19	2320	13,00	↓16	6
	20	<i>larg</i>	131	8,72		1		<i>information</i>	20	2261	12,67	↓15	6

Source: elaborated by the authors.

In general, the periods surveyed did not register a significant amount of unpublished terms between the stems 1-gram – suggesting stabilization of the terms used in the articles. Among the most recurrent terms, technology and proc always remained in the top four positions throughout the period analyzed, especially the first term, which remained in the top position in all decades. The recurrence of the stems information and economic is noted in the same way. In this context, one cannot ignore the elevated frequency of the stems product and system in the decades of 1970-1979 to 2000-2009.

It is remarkable the presence of stem model, which, on one hand, remained between the intermediate positions, on the other suggests the commitment of articles on the representation of innovation. Also, it is worth noting an increasing occurrence of stems associated with the organization (firms, businesses, industries, management). These stems generally have prominent climbing positions along the decades.

From 2000-2009, the stem chang no longer ranked among the twenty most frequent, indicating that aspects related to change had diminished interest in the analyzed abstracts. Paradoxically, these decades presented the stem knowledg, suggesting the growing interest in the cognitive aspect of innovation. In these same years, the stem accuracy has gained positions, indicating the interest of studies related to aspects involving the measurement of innovation.

As for stems 2-gram, Table 3 presents the top twenty positions along the analyzed decades:

Table 3 – 2-gram

decade	stems	i	f _i	%	i-1	#	decade	stems	i	f _i	%	i-1	#
1960-1969	<i>innovat_proc</i>	1	8	6,67		1	1990-1999	<i>technological_innovat</i>	1	268	5,73	↔1	4
	<i>technological_innovat</i>	2	7	5,83		1		<i>innovat_proc</i>	2	210	4,49	↔2	4
	<i>larg_scal</i>	3	6	5,00		1		<i>proc_innovation</i>	3	193	4,13	↑5	4
	<i>technological_chang</i>	4	5	4,17		1		<i>product_innovati</i>	4	193	4,13	↓3	3
	<i>economic_growth</i>	5	4	3,33		1		<i>cas_studi</i>	5	186	3,98	↑6	2
	<i>production_function</i>	6	4	3,33		1		<i>product_development</i>	6	129	2,76	↑12	2
	<i>busin_enterpris</i>	7	3	2,50		1		<i>long_term</i>	7	112	2,40	↔7	2
	<i>decision_making</i>	8	3	2,50		1		<i>small_busin</i>	8	96	2,05	↑20	2
	<i>larg_firms</i>	9	3	2,50		1		<i>information_technology</i>	9	93	1,99		1
	<i>practic_scientific</i>	10	3	2,50		1		<i>technological_chang</i>	10	83	1,78	↓6	4
	<i>proc_innovation</i>	11	3	2,50		1		<i>decision_making</i>	11	80	1,71	↓8	4
	<i>reduction_practic</i>	12	3	2,50		1		<i>competit_advantag</i>	12	76	1,63		1
	<i>scal_reduction</i>	13	3	2,50		1		<i>innovation_diffusion</i>	13	76	1,63	↓9	2
	<i>schumpeterian_hypothesis</i>	14	3	2,50		1		<i>innovation_management</i>	14	74	1,58		1
	<i>scientific_discov</i>	15	3	2,50		1		<i>busin_med</i>	15	72	1,54		1
	<i>social_structure</i>	16	3	2,50		1		<i>scienc_busin</i>	16	72	1,54		1
	<i>steel_industry</i>	17	3	2,50		1		<i>technology_transf</i>	17	64	1,37	↓15	3
	<i>technical_chang</i>	18	3	2,50		1		<i>economic_growth</i>	18	59	1,26	↓11	3
	<i>administrat_manag</i>	19	2	1,67		1		<i>innovat_activity</i>	19	53	1,13	↓12	2
	<i>assumption_government</i>	20	2	1,67		1		<i>high_technology</i>	20	51	1,09	↓19	2
1970-1979	<i>technological_innovat</i>	1	48	10,76	↑2	2	2000-2009	<i>technological_innovat</i>	1	1080	6,84	↔1	5
	<i>innovat_proc</i>	2	31	6,95	↓1	2		<i>cas_studi</i>	2	1019	6,45	↑5	3
	<i>diffusion_innovati</i>	3	26	5,83		1		<i>innovat_proc</i>	3	915	5,79	↓2	5
	<i>proc_innovation</i>	4	25	5,61	↑11	2		<i>innovation_system</i>	4	878	5,56		1
	<i>product_innovati</i>	5	23	5,16		1		<i>research_development</i>	5	506	3,20		1
	<i>technological_chang</i>	6	22	4,93	↓4	2		<i>innovation_management</i>	6	502	3,18	↑14	2
	<i>decision_making</i>	7	16	3,59	↑8	2		<i>product_innovati</i>	7	468	2,96	↓4	4
	<i>diffusion_proc</i>	8	14	3,14		1		<i>product_development</i>	8	453	2,87	↓6	3
	<i>technology_transf</i>	9	12	2,69		1		<i>long_term</i>	9	451	2,86	↓7	3
	<i>industrial_innovation</i>	10	10	2,24		1		<i>proc_innovation</i>	10	397	2,51	↓3	5
	<i>economic_growth</i>	11	9	2,02	↓5	2		<i>competit_advantag</i>	11	370	2,34	↑12	2
	<i>innovat_activity</i>	12	8	1,79		1		<i>practical_implicati</i>	12	358	2,27		1
	<i>training_progr</i>	13	8	1,79		1		<i>regional_innovation</i>	13	316	2,00		1
	<i>economic_development</i>	14	7	1,57		1		<i>small_med</i>	14	304	1,93		1
	<i>larg_firms</i>	15	7	1,57	↓9	2		<i>innovation_performanc</i>	15	301	1,91		1
	<i>larg_scal</i>	16	7	1,57	↓3	2		<i>decision_making</i>	16	293	1,86	↓11	5
	<i>privat_sect</i>	17	7	1,57		1		<i>innovat_activity</i>	17	281	1,78	↑19	3
	<i>product_proc</i>	18	7	1,57		1		<i>information_technology</i>	18	278	1,76	↓9	2
	<i>social_chang</i>	19	7	1,57		1		<i>economic_growth</i>	19	263	1,67	↓18	4
	<i>technical_chang</i>	20	7	1,57	↓18	2		<i>innovation_policy</i>	20	261	1,65		1
1980-1989	<i>technological_innovat</i>	1	135	8,99	↔1	3	2010-2014	<i>cas_studi</i>	1	1563	8,76	↑2	4
	<i>innovat_proc</i>	2	56	3,73	↔2	3		<i>innovat_proc</i>	2	1294	7,25	↔2	6
	<i>product_innovati</i>	3	56	3,73	↑5	2		<i>technological_innovat</i>	3	1186	6,64	↓1	6
	<i>technological_chang</i>	4	48	3,20	↑6	3		<i>innovation_system</i>	4	933	5,23	↔4	2
	<i>proc_innovation</i>	5	47	3,13	↓4	3		<i>innovation_performanc</i>	5	737	4,13	↑15	2
	<i>cas_studi</i>	6	42	2,80		1		<i>open_innovation</i>	6	674	3,78		1
	<i>long_term</i>	7	38	2,53		1		<i>innovation_management</i>	7	659	3,69	↓6	3
	<i>decision_making</i>	8	36	2,40	↓7	3		<i>product_innovati</i>	8	638	3,57	↓7	5
	<i>innovation_diffusion</i>	9	27	1,80		1		<i>practical_implicati</i>	9	570	3,19	↑12	2
	<i>diffusion_innovati</i>	10	26	1,73	↓3	2		<i>competit_advantag</i>	10	549	3,08	↑11	3
	<i>adaption_innovation</i>	11	24	1,60		1		<i>small_med</i>	11	530	2,97	↑14	2
	<i>product_development</i>	12	24	1,60		1		<i>proc_innovation</i>	12	502	2,81	↓10	6
	<i>diffusion_proc</i>	13	23	1,53	↓8	2		<i>innovat_activity</i>	13	482	2,70	↑17	4
	<i>technical_chang</i>	14	22	1,46	↑20	3		<i>product_development</i>	14	421	2,36	↓8	4
	<i>technology_transf</i>	15	22	1,46	↓9	2		<i>busin_med</i>	15	384	2,15		1
	<i>industrial_innovation</i>	16	19	1,26	↓10	2		<i>decision_making</i>	16	379	2,12	↔16	6
	<i>market_structure</i>	17	19	1,26		1		<i>med_sized</i>	17	362	2,03		1
	<i>technical_innovation</i>	18	19	1,26		1		<i>servic_innovation</i>	18	351	1,97		1
	<i>high_technology</i>	19	18	1,20		1		<i>economic_growth</i>	19	343	1,92	↔19	5
	<i>small_busin</i>	20	18	1,20		1		<i>policy_mak</i>	20	320	1,79		1

Source: elaborated by the authors.

In a similar way to that was found in the stems 1-gram the terms *innovat_proc* and *technological_innovat* prevailed in all decades, never appearing lower than third positions. This indicates the interest related to the technological innovation and innovation processes in the researched articles. In turn, the stem *decision_making*, although present in all the decades analyzed, showed decreasing behavior, indicating diminishing interest in researches involving decision making in this context.

What draws attention in the first two decades analyzed is how often the term *larg_scal* appeared, as well as those related to the organization's size (*larg_firms*, *small_busin*, *busin_med*, *small_med*, *med_sized*). Interestingly, in the decades of 1960-1969 and 1970-1979, the stems related to the size of the companies indicated studies focused to large enterprises, while, from there, the focus shifted to the medium and small firms.

The term *economic_growth* has often been employed; albeit at increasingly lower positions (there was no record of this stem in the top twenty positions in the decade from 1980-1989). Researches on product innovation also continue to recur – except for the decade of 1960-1969, the term *product_innovati* has always appeared in prominent positions. Although present in the first four decades analyzed, the term *technological_chang* has not attracted the same interest in articles published since 2000, indicating a possible exhaustion in studies of technological change.

Since the decade of 1980-1989, the stem *product_development* has frequently been present in the analyzed abstracts, indicating interest in issues associated with the development of innovative products. Interestingly, from this decade on, the research method employed (*cas_studi*) gained positions and was the most used term in the years 2010-2014. As for the decade of 1990-1999, the stems *innovation_management* and *competit_advantag* are always presented in prominent positions, leading to consider the increase of studies about innovation management and innovation as competitive advantage element. Therefore, it is possible to draw a parallel between the use of the term "competitive advantage", coming from innovation, combined with studies on small and medium enterprises (term also highlighted in recent years).

The first twenty positions occupied by the stems 3-gram are shown in Table 4:

Table 4 – 3-gram

decade	stems	i	f _i	%	i-1	#	decade	stems	i	f _i	%	i-1	#
1960-1969	<i>larg_scal_reduction</i>	1	3	2,50		1	1990-1999	<i>national_innovation_system</i>	1	56	1,20		1
	<i>practic_scientific_discov</i>	2	3	2,50		1		<i>small_med_enterpris</i>	2	45	0,96		1
	<i>reduction_practic_scientific</i>	3	3	2,50		1		<i>product_proc_innovati</i>	3	42	0,90		1
	<i>scal_reduction_practic</i>	4	3	2,50		1		<i>med_sized_enterpris</i>	4	22	0,47		1
	<i>assumption_government_agenci</i>	5	2	1,67		1		<i>total_quality_management</i>	5	19	0,41		1
	<i>beli_screening_early</i>	6	2	1,67		1		<i>small_busin_economics</i>	6	18	0,39		1
	<i>careful_selection_projects</i>	7	2	1,67		1		<i>human_resourc_management</i>	7	17	0,36		1
	<i>corporation_examin_probl</i>	8	2	1,67		1		<i>product_development_proc</i>	8	17	0,36		1
	<i>critical_system_selection</i>	9	2	1,67		1		<i>product_innovation_management</i>	9	17	0,36	↓6	2
	<i>deal_uncertainty_national</i>	10	2	1,67		1		<i>high_technology_firms</i>	10	16	0,34		1
	<i>econometric_models_misleadingly</i>	11	2	1,67		1		<i>adaption_innovation_inventory</i>	11	15	0,32	↓1	2
	<i>foregoing_great_opportuniti</i>	12	2	1,67		1		<i>successful_product_development</i>	12	15	0,32		1
	<i>gaps_fundamental_knowledg</i>	13	2	1,67		1		<i>proc_technological_innovation</i>	13	14	0,30	↓11	2
	<i>government_agenci_support</i>	14	2	1,67		1		<i>decision_making_proc</i>	14	12	0,26	↓8	3
	<i>growth_amount_money</i>	15	2	1,67		1		<i>scienc_technology_policy</i>	15	11	0,24		1
	<i>incent_resourc_research</i>	16	2	1,67		1		<i>product_development_projects</i>	16	10	0,21		1
	<i>innovation_united_stat</i>	17	2	1,67		1		<i>regional_innovation_systems</i>	17	10	0,21		1
	<i>invention_commercialization_introduc</i>	18	2	1,67		1		<i>world_war_ii</i>	18	10	0,21	↓13	2

1970-1979	<i>isolation_gaps_fundamental</i>	19	2	1,67		1	2000-2009	<i>technological_innovation_diffusion</i>	19	9	0,19		1
	<i>long_tim_span</i>	20	2	1,67		1		<i>world_wid_web</i>	20	9	0,19		1
	<i>decision_making_proc</i>	1	4	0,90		1		<i>national_innovation_system</i>	1	260	1,65	↔1	2
	<i>adaptation_occurs_stag</i>	2	3	0,67		1		<i>regional_innovation_system</i>	2	211	1,34	↑17	2
	<i>assess_congruent_economic</i>	3	3	0,67		1		<i>small_med_sized</i>	3	187	1,18	↔3	2
	<i>beings_involved_proc</i>	4	3	0,67		1		<i>information_communication_technologi</i>	4	139	0,88		1
	<i>british_industry_oil</i>	5	3	0,67		1		<i>med_sized_enterpris</i>	5	131	0,83		1
	<i>concret_launching_improved</i>	6	3	0,67		1		<i>product_proc_innovati</i>	6	118	0,75	↓3	2
	<i>consequenc_relationship_enterpris</i>	7	3	0,67		1		<i>creativity_innovation_management</i>	7	79	0,50		1
	<i>des_society_chang</i>	8	3	0,67		1		<i>human_resourc_management</i>	8	74	0,47	↓7	2
	<i>economic_forc_impinged</i>	9	3	0,67		1		<i>product_development_proc</i>	9	73	0,46	↓8	2
	<i>enterpris_continu_function</i>	10	3	0,67		1		<i>analys_strategic_management</i>	10	67	0,42		1
	<i>forced_compani_inflation</i>	11	3	0,67		1		<i>technology_analys_strategic</i>	11	67	0,42		1
	<i>impinged_tangibl_intangibl</i>	12	3	0,67		1		<i>knowledg_based_economy</i>	12	62	0,39		1
	<i>influenc_monopoly_product</i>	13	3	0,67		1		<i>economics_innovation_technology</i>	13	60	0,38		1
	<i>innovation_united_stat</i>	14	3	0,67	↑17	2		<i>small_med_enterpris</i>	14	60	0,38	↓2	2
	<i>launching_improved_products</i>	15	3	0,67		1		<i>decision_making_proc</i>	15	55	0,35	↓14	4
	<i>mad_concret_launching</i>	16	3	0,67		1		<i>total_quality_management</i>	16	54	0,34	↓5	2
	<i>monopoly_product_innovation</i>	17	3	0,67		1		<i>critical_succ_fact</i>	17	51	0,32		1
	<i>organizat_consequenc_relationship</i>	18	3	0,67		1		<i>resourc_based_view</i>	18	43	0,27		1
<i>practic_fre_probl</i>	19	3	0,67		1	<i>sem_structured_interviews</i>	19	43	0,27		1		
<i>proc_adaptation_occurs</i>	20	3	0,67		1	<i>foreign_direct_investment</i>	20	42	0,27		1		
1980-1989	<i>adaption_innovation_inventory</i>	1	16	1,07		1	2010-2014	<i>small_med_sized</i>	1	327	1,83	↑3	3
	<i>med_sized_firms</i>	2	10	0,67		1		<i>med_sized_enterpris</i>	2	251	1,41	↑5	2
	<i>small_med_sized</i>	3	10	0,67		1		<i>structural_equation_modeling</i>	3	236	1,32		1
	<i>innovation_diffusion_model</i>	4	9	0,60		1		<i>information_communication_technologi</i>	4	189	1,06	↔4	2
	<i>technology_transf_society</i>	5	8	0,53		1		<i>national_innovation_system</i>	5	180	1,01	↓1	3
	<i>product_innovation_management</i>	6	7	0,47		1		<i>regional_innovation_system</i>	6	176	0,99	↓2	3
	<i>adaption_innovation_theory</i>	7	6	0,40		1		<i>small_med_enterpris</i>	7	149	0,83	↓5	2
	<i>decision_making_proc</i>	8	6	0,40	↓1	2		<i>sem_structured_interviews</i>	8	136	0,76	↑19	2
	<i>innovation_diffusion_proc</i>	9	6	0,40		1		<i>community_innovation_survey</i>	9	119	0,67		1
	<i>comput_aided_design</i>	10	5	0,33		1		<i>resourc_based_view</i>	10	114	0,64	↑18	2
	<i>proc_technological_innovation</i>	11	5	0,33		1		<i>analys_strategic_management</i>	11	98	0,55	↓10	2
	<i>textil_machinery_exhibition</i>	12	5	0,33		1		<i>technology_analys_strategic</i>	12	98	0,55	↓11	2
	<i>world_war_ii</i>	13	5	0,33		1		<i>product_innovation_proc</i>	13	96	0,54		1
	<i>comput_integrated_manufacturing</i>	14	4	0,27		1		<i>busin_model_innovation</i>	14	92	0,52		1
	<i>importanc_technological_innovation</i>	15	4	0,27		1		<i>human_resourc_management</i>	15	87	0,49	↓8	3
	<i>larg_small_firms</i>	16	4	0,27		1		<i>multipl_cas_study</i>	16	76	0,43		1
	<i>long_term_planning</i>	17	4	0,27		1		<i>decision_making_proc</i>	17	75	0,42	↓15	5
	<i>management_information_systems</i>	18	4	0,27		1		<i>innovation_technology_management</i>	18	74	0,41		1
	<i>model_proc_innovation</i>	19	4	0,27		1		<i>social_network_analys</i>	19	72	0,40		1
	<i>national_innovation_policy</i>	20	4	0,27		1		<i>multipl_regression_analys</i>	20	62	0,35		1

Source: elaborated by the authors.

It is important to note that in the 1960-1969 decade, only a single stem appeared later: *innovation_united_stat*. During this period it is noticed that the summaries of the articles highlight issues related to scientific practice, production scales and governmental aspects and which subsequently showed less interest.

In the decade of 1970-1979, the stem *decision_making_proc* was the most used in the analyzed abstracts. This term continued to be repeatedly used, indicating the interest of researchers in the decision-making process (confirming the facts found under 2-gram stems). Still in the decade of 1970-1979, there are summaries containing terms associated with adaptation, product launch (or improved products), monopoly products and business relationships.

The decade from 1980 to 1989 was responsible for bringing to the fore studies concerned the size of enterprises, especially small and medium – it is noteworthy that in other decades the same three distinct stems related to the size of organizations were always repeated. It is also highlighted the *adaption_innovation_inventory*, *product_innovation_management* and *proc_technological_innovation* stems, that also ap-

peared in the following decade and, thereafter, no longer consisted in the top twenty positions. Also persistent in the decades of 1980-1989 and 1990-1999, the `world_war_ii` stem indicates the prevalence of studies that addressed innovation in this historical context. Furthermore, it is evident the recurrence of terms that involve models and innovation diffusion transfer processes, or technologies to society. Likewise, it highlights studies involving terms related to projects and computerized production.

The decade of 1990-1999 stands out by the number of articles dealing with regional innovation systems, as well as those associated with product development - besides those associated with small and medium enterprises. Specifically, this decade has presented for the first time the top twenty positions, stems related to human resource management and total quality (which were also found in the following decade). It is interesting to note the emergence of `steam_world_wide_web` in the twentieth position.

The years 2000-2009 have been marked by interest in national and regional innovation systems as well as the persistence of summaries with terms related to the size of companies. In addition, this decade marks the studies (which was repeated the next decade) addressing strategic issues, information technology, resource-based view and research involving semi-structured interviews. Also appears in the first twenty positions, terms associated with creativity and knowledge management.

In the years 2010-2014, what prevailed were the summaries in which `small_med_sized` and `med_sized_enterpris` stems were found. In addition, in a prominent position, remained the stems `national_innovation_system` and `regional_innovation_system`. Also notes the increase (compared to the previous decade) in the use of terms indicating researches using semi-structured interviews as well as studies related to resource-based view. It is noticeable the emergence of quantitative analysis (specifically, structural equation modeling and multiple regression), and social network analysis.

Final considerations

Initially, considering the objective of this study (explain the most commonly used terms in scientific paper abstracts, where the word 'innovation' is part of the title, over the decades between the years 1960 and 2014, and the dynamics between them), it is clear that in the surveyed databases, assuming that it is possible to infer from up to three terms, the published articles had a dynamic in which it is possible to identify movement of studies ranging from theoretical to practical. It can be inferred it is a passage from research addressing innovation in the context of larger companies and government policies for small sized organizations. Case research also stands out as the main research method employed - but in recent years, studies using quantitative methods have been increasingly frequent. More evident than the movement of the most common terms is the fact that the results reinforced innovation as a concept 'co-existing' with the decision-making process, and highlighted its role in the production process - specifically, in products.

The present study is not free from limitations - which, in most cases, are a natural consequence of the very technique used. It does not rule out the possibility that

other patterns can arise when there is a higher estimated amount of terms. It is, however, an easily configurable parameter, according to the aspirations of the researcher. Also, an inherent feature of the technique is the fact that there is an inverse relationship between the size of the stems and the frequency with which they are found, that is, the higher a chain of analyzed words (which brings more meaning to our analysis) the smaller its occurrence is (which weakens the identification of patterns). Moreover, it is important to emphasize that the frequency of a stem is not formed necessarily by separate articles: it is possible that the stem is found more than once in the same summary.

Several future studies can be developed from this research. As the focus of this paper lies in the frequency with which terms occur, it would be interesting to associate these occurrences to the socio-economic events that marked the years. It would also be relevant to relate the terms of the nationality and the institutions with which the authors are associated, seeking geographic and institutional standards for research involving innovation. Similarly, it would be necessary to verify the existence of networking authors related to certain terms. The studies suggested here could contribute to the formation of a reference panel on research innovation and its peculiarities.

In any case, the technique proved to be adequate for identifying trends in this field of study. Furthermore, it allows affirming that the themes found, in general, have more alignment with each other than dissimilarities, indicating consolidation and maturity of the term in the academic context.

References

- ALCOUFFE, S.; GUEDRI, Z. Le rôle des canaux de communication et des caractéristiques perçues de l'innovation dans le processus d'adoption de la comptabilité par activités (méthode ABC). *Comptabilité - Contrôle - Audit*, 14(3), 39-65, 2008.
- AUTANT-BERNARD, C.; CHALAYE, S.; MANCA, F.; MORENO, R.; SURIÑACH, J. Measuring the adoption of innovation: a typology of EU countries based on the Innovation Survey. *Innovation: The European Journal of Social Science Research*, 23(3), 199-222, 2010.
- BEZERRA, C. A.; GUIMARÃES, A. J. Mineração de texto aplicada às publicações científicas sobre gestão do conhecimento no período de 2003 a 2012. *Perspectivas em Ciência da Informação*, 19(2), 131-146, 2014.
- BUCIC, T.; NGO, L. V. Examining drivers of collaborative inbound open innovation: empirical evidence from Australian firms. *International Journal of Innovation Management*, 16(4), 1-25, 2012.
- ÇAKAR, N. D.; ERTURK, A. Comparing innovation capability of small and medium-sized enterprises: examining the effects of organizational culture and empowerment. *Journal of Small Business Management*, 48(3), 325-359, 2010.
- CARRER, C. C.; PLONSKI, G. A.; CARRER, C. R.; OLIVEIRA, C. E. Innovation and entrepreneurship in scientific research. *Revista Brasileira de Zootecnia*, 36(especial), 17-25, 2010.
- CHEN, Z. Database keyword search: beyond its current landscape. *International Journal of Information Technology & Decision Making*, 11(2), 491-500, 2012.

CHESBROUGH, H. W. **Open innovation: the new imperative for creating and profiting from technology** (Vol. 20). Boston: Harvard Business School Press, 2003.

DEMBOWSKI, F. L. The roles of benchmarking, best practices & innovation in organizational effectiveness. **International Journal of Organizational Innovation**, 15(3), 6-20, 2013.

DURISIN, B.; CALABRETTA, G.; PARMEGGIANI, V. The intellectual structure of product innovation research: a bibliometric study of the Journal of Product Innovation Management, 1984-2004. **Journal of Product Innovation Management**, 27(3), 437-451, 2010.

FAN, W.; WALLACE, L.; RICH, S.; ZHANG, Z. Tapping the power of text mining. **Communications of the ACM**, 49(9), 76-82, 2006.

FARO, A.; GIORDANO, D.; SPAMPINATO, C. Combining literature textmining with microarray data: advances for system biology modeling. **Briefings in Bioinformatics**, 13(1), 61-82, 2011.

FREZATTI, F.; BIDO, D. S.; CRUZ, A. P.; CAMARGO, M. J. O papel do balanced scorecard na gestão da inovação. **Revista de Administração de Empresas**, 54(4), 381-392, 2014.

GAJZER, M. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry. **Technological and Economic Development of Economy**, 16(2), 219-232, 2010.

GEER FRAZIER, B. Complexity leadership generates innovation, learning, and adaptation of the organization. *Emergence*: **Complexity & Organization**, 16(3), 105-116, 2014.

GODIN, B. **Innovation: the history of a category**. Project on the intellectual history of innovation. Montreal: INRS, 2008.

GRIZENDI, E. **Manual de orientações gerais sobre inovação**. Ministério das Relações Exteriores/Departamento de Promoção Comercial e Investimentos/Divisão de Programas de Promoção Comercial. Brasília: Programa das Nações Unidas para o Desenvolvimento, 2011.

HALKOS, G. E.; TZEREMES, N. G. Modelling the effect of national culture on countries' innovation performances: a conditional full frontier approach. **International Review of Applied Economics**, 27(5), 656-678, 2013.

HAVE, S.; HAVE, W.; STEVENS, F.; ELST, M.; POL COYNE, F. **Modelos de gestão**. São Paulo: Pearson Prentice Hall, 2003.

Laboratory of Computational Intelligence. Acesso em 1 de janeiro de 2015, disponível em PreText 2: **a reestruturação da ferramenta de pré-processamento de textos**: <http://sites.labic.icmc.usp.br/pretext2>, 2008.

LASAGNI, A. How can external relationships enhance innovation in SMEs? New evidence for Europe. **Journal of Small Business Management**, 52(2), 310-339, 2012.

LAZZAROTTI, F.; DALFOVO, M. S.; HOFFMANN, V. E. A bibliometric study of innovation based on Schumpeter. **Journal of Technology Management & Innovation**, 6(4), 121-134, 2011.

LIAO, S.-H.; CHANG, W.-J.; HU, D.-C.; YUEH, Y.-L. Relationships among organizational culture, knowledge acquisition, organizational learning, and organizational innovation in Taiwan's banking and insurance industries. **The International Journal of Human Resource Management**, 23(1), 52-70, 2012.

LICHTENTHALER, U. Open innovation: past research, current debates, and future directions. **Academy of Management Perspectives**, 25(1), 75-93, 2011.

MARTINS, C. A.; GODOY, D.; MONARD, M. C.; MATSUBARA, E. T.; AMANDI, A. **Uma experiência em mineração de textos utilizando clustering probabilístico clustering hierárquico**. Instituto de Ciências Matemáticas e de Computação. São Carlos: Universidade de São Paulo, 2003.

MIERES, C. G.; SÁNCHEZ, J. Á.; VIJANDE, M. L. Internal marketing, innovation and performance in business services firms: the role of organizational unlearning. **International Journal of Management**, 29(4), 403-429, 2012.

MOEZZI, M.; GOINS, J. Text mining for occupant perspectives on the physical workplace. **Building Research & Information**, 39(2), 169-182, 2011.

MUSTEEN, M.; AHSAN, M. Beyond cost: the role of intellectual capital in offshoring and innovation in young firms. **Entrepreneurship Theory and Practice**, 37(2), 421-434, 2013.

NIDUMOLU, R.; PRAHALAD, C. K.; RANGASWAMI, M. R. Why sustainability is now the key driver of innovation. **Harvard Business Review**, 87(9), 56-64, 2009.

Organisation for Economic Co-operation and Development. **Oslo manual: guidelines for collecting and interpreting innovation data** (3 ed.). Paris: OECD Publishing, 2005.

ORLICKAS, E. **Modelos de gestão: das teorias da administração à gestão estratégica**, 2 ed. Curitiba: IBPEX, 2011.

PELLICELLI, M.; MEO, C. C.; CIOFFI, A. Organizational change: business process reengineering or outsourcing. Annals of the University of Oradea, **Economic Science Series**, 21(2), 277-293, 2012.

PRESTER, J.; BOZAC, M. G. Are innovative organizational concepts enough for fostering innovation? **International Journal of Innovation Management**, 16(1), 23, 2012.

REJEB, H. B.; MOREL-GUIMARÃES, L.; BOLY, V.; ASSIÉLOU, N. G. Measuring innovation best practices: improvement of an innovation index integrating threshold and synergy effects. **Technovation**, 28, 838-854, 2008.

RIJSBERGEN, C. J. **Information retrieval**. Newton: Butterworths, 1979.

SCHUMPETER, J. A. **The theory of economic development**. Cambridge: Transaction Books, 1934.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. Florianópolis: UFSC, 2005

SOARES, M. V.; PRATI, R. C.; MONARD, M. C. **PreText: a reestruturação da ferramenta de pré-processamento de textos**. Instituto de Ciências Matemáticas e de Computação. São Carlos: Universidade de São Paulo, 2008.

THOMSON REUTERS. Acesso em 20 de janeiro de 2015, disponível em EndNote: <http://endnote.com>, 2013.

VACARO, I. G.; JANSEN, J. J.; VAN DEN BOSCH, F. A.; VOLBERDA, H. W. Management innovation and leadership: the moderating role of organizational size. **Journal of Management Studies**, 49(1), 28-51, 2012.

WANG, M. Y.; CHANG, D. S.; KAO, C. H. Identifying technology trends for R&D planning using TRIZ and text mining. **R&D Management**, 40(5), 491-509, 2010.

WEISS, S. M.; INDURKHYA, N.; ZHANG, T. **Fundamentals of predictive text mining**. London: Springer-Verlag, 2010.

YOON, B.; PHAAL, R.; PROBERT, D. Morphology analysis for technology roadmapping: application of text mining. **R&D Management**, 38(1), 51-68, 2008.

Received: 05/28/2015

Approved: 11/19/2015

303