Startups: a systematic review of literature and future research directions

Luciana Santos Costa Vieira da Silva
Centro Universitário Municipal de São José (USJ)
email: luvcosta10@gmail.com

Fabíola Kaczam
Universidade Federal de Santa Maria (UFSM)
email: kaczamf@gmail.com

Anderson de Barros Dantas
Universidade Federal de Alagoas (UFAL)
email: anderson.admead@gmail.com

Josenilde Mario Janguia
Universidade Mondlane-Moçambique
email: josenildemario@yahoo.com.br

■ ABSTRACT

The objective of the research is to map the literature based on a Systematic Literature Review on the theme of startups and to highlight some theoretical gaps based on publications of high-reputation scientific journals. The period from 1990 to 2019 was defined for the elaboration of this study. We use the excel spreadsheet, in addition to the HistCite ™, VOSviewer, IRATUMEQ, and R Studio packages. The results show that the typology of the startups evaluated, after reading 68% of the articles, organizations are characterized as a group of new companies, that is, relatively young and inexperienced when compared to the most stable and mature in organizational development. They refer to those that are in the initial stage and are susceptible to the influence of various factors, such as investors, supplier customers, partners, etc., and should think strategically about how to act and, this concerns a group of dynamic startups that deal with innovations.

Key-words: Startups; Systematic Literature Review; Innovation; Entrepreneurship; Theoretical gaps.

■ RESUMO

O objetivo da pesquisa é mapear a literatura de uma Revisão Sistemática da Literatura sobre startups e destacar algumas lacunas teóricas baseadas em publicações de periódicos científicos internacionais. Usamos a planilha Excel, além dos pacotes HistCite ™, VOSviewer, IRATUMEQ e R Studio. Os resultados mostram que pelo tipo de startups avaliadas, após a leitura de 68% dos artigos, as organizações se caracterizam como um grupo de empresas novas, ou seja, relativamente jovens e inexperientes quando comparadas às mais estáveis e maduras em desenvolvimento organizacional e devem pensar estrategicamente sobre como atuar e, que diz respeito a um grupo de startups dinâmicas que operam com inovações.

Palavras-Chave: Startups; Revisão Sistemática da Literatura; Inovação; Empreendedorismo.

1 INTRODUCTION

Startups can be seen as small organizations, which have bold and replicable business models, in the launching process or even with a short time of operation, where they have a high potential in terms of scalability and with a focus turned to the activities of research and development of innovative ideas (Crowne, 2002; Cohen; Feld, 2010; Cooper; Vlaskovits, 2010; Ries, 2012; Nager; Nelsen; Nouyrigat, 2013; Kollmann, Stöckmann, Hensellek, Kensbock, 2016; Spender, Corvello, Grimaldi, Rippa, 2017; Rompho, 2018).

The main contribution of the work is to investigate through a systematic literature review potential theoretical gaps, based on publications in scientific journals contained in the Web of Science (WoS) and Scopus databases on the theme "startup", describing some characteristics (typology) from the textual corpus evaluated.

The originality of this research lies in the fact that empirical evidence containing typologies was not found in systematic literature review articles on this topic, plus suggestions for future research in the aforementioned journal bases. In this sense, the work is justified, because when finding typologies in this regard, it can help managers and researchers to better direct their research and, consequently, find better answers to research questions.

When considering the importance of startups in emerging economies, the general objective of this research is to establish a systematic literature review for startups, based on the following research problem: "What is the state of the art of research on startups? From the previous research question, it unfolds in the following secondary question: What are the directions of future research on the investigated Startups? The Systematic Literature Review is important to point out theoretical gaps and, thus, to be able to suggest future research for the development of the theme about startups.

The article is structured in five sections that can be summarized as follows: the first deals with the introductory part; the second concerns the theoretical framework talking about startups and their characteristics; the third brings the methodological procedures; the fourth refers to the presentation and analysis of results and the fifth concerns future research considerations and directions.

2 LITERATURE REVIEW - CONCEPT AND CHARACTERISTICS OF *STARTUPS*

Startups refer to organizations that arise with high maneuverability in terms of adapting to changes in the market, as well as flexibility and dynamism supported by technological tools that have revolutionized business forms, product concepts and, mainly in the provision of these services (Kim, 2005; Spender *et al.*, 2017).

Startups are not just technology companies; but any companies in the process of being set up; (Hermanson, 2011; Longhi, 2011; Blank and Dorf, 2012; Perin 2016; Kohler, 2016). This type of entrepreneurship happens more in the area of technology because the costs are lower to create a software company than an industry.

It is worth noting that there is a distinction between a small joint venture and a Startup for the authors Blank and Dorf (2012). They are not smaller versions of large companies. For these authors, a start-up company, which is not focused on product or service innovation and, therefore, dispensed with the risks of uncertainties, is not a startup.

For Padrão and Andreassi (2013), startups aim at the growth in sales to have the return on investment applied in the product development phase, as it is the way these companies try to survive in their initial stages.

The acceleration of technological change in a dynamic organizational context emphasizes the need to develop innovative capacities. From the first to the fourth Industrial Revolution what is observed is that competitiveness depends less on the adoption of new technologies and more on the development and effective use of technology to create value (Schwab, 2017).

Based on the concepts presented, it is possible to understand this type of enterprise as those with a short time in the market, present a business model of the rapid application, economic growth in a short period, development of products generated from the practice of innovation and, finally, investment targeting for Research, Development, and Innovation (RD&I).

Finally, uncertainty and innovation are also characteristics inherent to the development of the smart industry, which draws attention to the prox-

Luciana Santos Costa Vieira da Silva • Fabíola Kaczam • Anderson de Barros Dantas • Josenilde Mario Janguia

imity that startups have with enabling technologies, assigning them the role of precursors of the technological transition in Brazil, which leads to the concept of smart startups.

The development of a Systematic Literature Review - RSL, in general, is related to the conduct of research, directing new paths to be investigated; therefore, it requires pre-defined steps by the researcher, such as the definition of the research problem; the search strategy, the inclusion and exclusion criteria, in addition to checking the quality of the selected material (Oliveira, 2007; Sampaio; Mancini, 2007; Tranfield *et al.*, 2003).

3 THE PROCESS OF CONDUCTING SYSTEMATIC LITERATURE REVIEW - RSL

The systematic review seeks to answer the research question defined in the introduction of the work, where the process of developing this type of study requires the realization of a protocol covering three stages: (i) review planning; (ii) conducting the review; and (iii) carrying out the report and disclosing the results.

Based on the identification of the terms and keywords previously created in the first stage in the discussions between the researchers, the subsequent steps must be detailed with the sequence used to guarantee replication.

The choice of databases is justified, due to a large number of journals, in addition to the fact that several databases are indexed. The Scopus database provides an overview of the world's scientific productions, covering the areas of social, biological, health, and physical sciences, indexing the most varied academic titles, conferences, books, among others (ELSEVIER). The WoS database, on the other hand, is considered interdisciplinary, allows access to abstracts and references in all areas of knowledge, and covers around 12,000 journals (Coordination for the Improvement of Higher Education Personnel - CAPES).

4 DATA COLLECTION

For this study, the following keyword was defined as the search strategy: "Startup*", as it covers the research problem of this RSL. Regarding the choice of journal bases, two were used: Scopus and Web of Science (WoS) and the time period from 1990 to 2019 was stipulated, since it is from 1990 that the academy started to effectively study the topic of startups. Based on the research question, a comprehensive search string was created for the two Scopus and WoS databases, leading to the textual corpus of this RSL, as shown in Table 1.

Due to the volume of articles found in the two bases of journals, a flowchart was created, as shown in Figure 1. This covers the selection criteria of the articles and is based on the recommendations of Almeida and Goulart (2017), supporting the authors this RSL regarding theoretical and empirical evidence, as well as assisting in the decision-making process in order to minimize selection bias.

Table 1 Search strings and number of articles according to the bases

Bases	Search String	Number of Articles
Scopus	TITLE-ABS-KEY (*"Startup*"*) AND (LIMIT-TO (PUBYEAR > 1989 AND PUBYEAR < *2020* AND (LIMIT-TO (SUBJAREA, *"BUSI"*) OR LIMIT-TO (SUBJAREA, *"ECON"*)) AND (LIMIT-TO (PUBSTAGE, *"final"*)) AND (LIMIT-TO (DOCTYPE, *"ar"*)) AND (LIMIT-TO (LANGUAGE, *"English"*)) AND (LIMIT-TO (SRCTYPE, *"j"*))	1.300
Web of Science	TÓPICO: (*"Startup*"*) Refinado por: ANOS DE PUBLICAÇÃO: (>1989 and < 2020) AND CATEGORIAS DO WEB OF SCIENCE: (MANAGEMENT OR BUSINESS OR ECONOMICS OR BUSINESS FINANCE) AND TIPOS DE DOCUMENTO: (ARTICLE) AND IDIOMAS: (ENGLISH) Tempo estipulado: Todos os anos. Índices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	1.000

Source: The authors (2020).

5 RESEARCH PROTOCOL

The flowchart presented in Figure 1, shows the steps through its four filters, based on the research protocol used for the development of RSL, containing the inclusion and exclusion criteria that were adopted.

Figure 1 comprises the stages of the work selection process through filters related to the research bases, types of documents, language, research areas, and reputation of the journals.

Thus, we found 24,802 documents in the Scopus database, by filtering by types of documents „Article", we obtained a total of 1,726 papers and, later, when we defined the language as „English", 1,300 articles remained. In the WoS database, a total of 11,268 documents were found; when we applied the filters by type of documents „Article", 1,394 articles were found and, after the definition of the language as „English", 1,000 articles remained. It is worth mentioning that the English language was chosen because it is part of all the journals belonging to the Scopus and WoS databases and also because the few journals with other languages were not shown in the selection quartiles.

With the unification of the results of the two databases, a total of 2,300 articles were obtained. As an inclusion criterion, we chose to select articles published only in journals classified in the first, second and third citation quartiles (Q1, Q2, and Q3) from the following contextual areas of knowledge of the SJR (2018): Business and Management; Business, Management and Accounting (miscellaneous) and Economics. Through this selection, 1,098 articles from the Scopus database and 891 articles from the WoS were excluded, leaving a total of 341 documents that were previously selected regarding the adherence of the contextual area and research question, involving the reading of titles, abstracts, and keywords of the same. 113 duplicate articles were still excluded, resulting in 228 articles to compose the final textual corpus of this RSL.

To assess the quality of the studies, we chose to classify them considering the reputation of the journals, as in the view of Tranfield *et al.*, (2003), researchers in the management area place confidence in the classification of specific journals, with implicit quality. We used the Scimago Journal & Country Rank (SJR) indicator as a selection criterion, which can measure the influence of a specific journal through the average of the citations of the last 3 years before the year in which it is desired to analyze, which are weighted according to the journal's prestige and knowledge area (SJR, 2018).

The survey of selected articles was carried out on January 4, 2020, and the final selection took place from January 12 to 14, of the same year, with the participation of four management experts who have already prepared systematic literature reviews. The documents were downloaded from the bases between 17 and 19 January 2020.

Finally, the Excel electronic spreadsheet, HistCite ™ software, which is an implementation of algorithmic histography, was used in the data analysis and helps researchers to visualize the results obtained in the bibliographic searches, in addition to the VOSviewer packages and the IRAMUTEQ that is anchored in the R software that allows other different ways to analyze the textual corpus (Camargo; Justo, 2013), in addition to the R Studio package.

6 THE TEXT MINING ANALYSIS TECHNIQUE

Text Mining, also known in the literature as Intelligent Text Analysis, Text Data Mining or Knowledge Discovery in Text (KDT-Knowledge Discovered in Texts) emerges as an approach to the acquisition of useful and non-trivial information from unstructured data from texts (Figueiredo; Catini; Mendes, 2018).

In general, grouping strategies can be organized into two types: partial grouping and hierarchical grouping. In the partial grouping, the document collection is divided into a simple partition of k groups, also known as grouping by optimization. The objective is to iteratively divide the set of objects into k groups, where k is usually a value previously informed by the user. Document groups are formed to optimize the compression and / or separation of the grouping.

In this section we have the results obtained from the textual corpus of this research. The analyzes carried out are supported by the research problem, where the productivity of authors is evaluated, the productivity of journals and the measurement of the
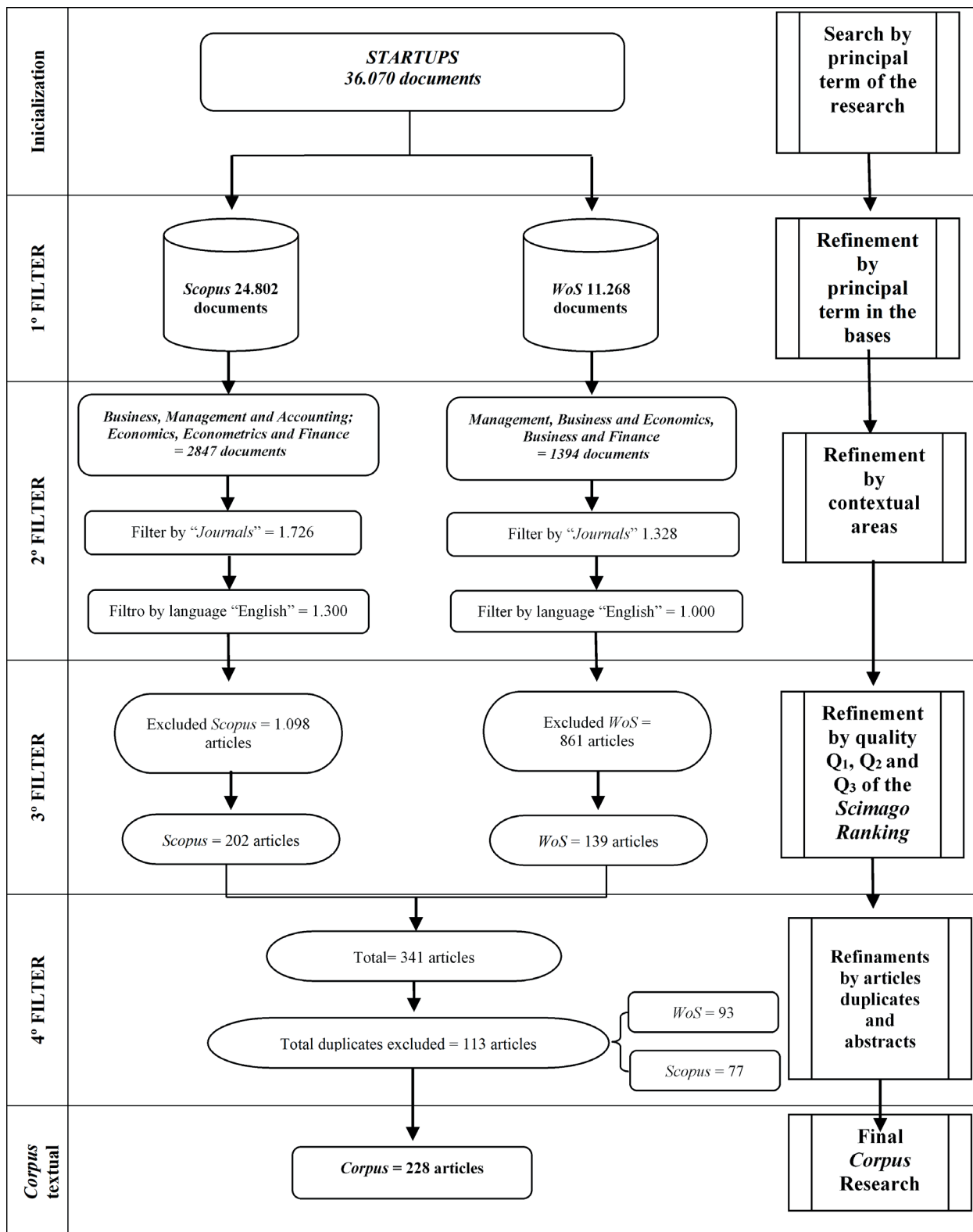
Luciana Santos Costa Vieira da Silva • Fabíola Kaczam • Anderson de Barros Dantas • Josenilde Mario Janguia

Inicialization	***STARTUPS*** *36.070 documents*	**Search by principal term of the research**
1° FILTER	*Scopus* 24.802 documents *WoS* 11.268 documents	**Refinement by principal term in the bases**
2° FILTER	***Business, Management and Accounting; Economics, Econometrics and Finance* = 2847 documents** ***Management, Business and Economics, Business and Finance* = 1394 documents** Filter by "*Journals*" = 1.726 Filter by "*Journals*" 1.328 Filtro by language "English" = 1.300 Filter by language "English" = 1.000	**Refinement by contextual areas**
3° FILTER	Excluded *Scopus* = 1.098 articles Excluded *WoS* = 861 articles *Scopus* = 202 articles *WoS* = 139 articles	**Refinement by quality Q₁, Q₂ and Q₃ of the *Scimago Ranking***
4° FILTER	Total= 341 articles Total duplicates excluded = 113 articles *WoS* = 93 *Scopus* = 77	**Refinaments by articles duplicates and abstracts**
***Corpus* textual**	***Corpus* = 228 articles**	**Final *Corpus* Research**

Figure 1 Systematization of the Research Protocol

Source: The authors (2020).

frequency of occurrence and co-occurrence of certain words in the corpus.

7 TEMPORAL AND GEOGRAPHIC DISTRIBUTION OF THE CORPUS

The articles in the corpus (228 articles) comprise 78 journals and approximately 496 authors and co-authors identified in the Web of Science and Scopus databases. In the period from 1990 to 2019, an average of 64.21 citations per article occurred over the years, which portrays an average of 0.46 documents per author, and a total of 2.18 authors per document. At the same time that the corpus has an annual growth rate of scientific production estimated at 15.05%.

There is a significant increase in terms of the number of documents produced annually, with its peak from 2015, with approximately 48.43% of the production on this theme, which occurred between the years 2015 and 2019. At the same time that the number of citations received over time by articles had a decreasing trend.

It was also noticed that approximately 90% of the articles in the corpus obtained at least 10 citations by the year 2009, while in the last decade (2010 to 2019) only 8.8% of the documents were cited, which can be evidenced as something natural if we consider that the time window for citing a scientific paper takes at least two years to occur, as advocated by Leydesdorff (2009) and Campanario (2015).

Regarding the geographical distribution of published articles, 313 authors and co-authors distributed in 45 countries were identified, who declared that they have their institutional ties in more than one country, as can be seen in Figure 2.

Based on Figure 1, we can highlight the presence of the United States with 37.38% of the sample, followed by Germany with 8.31%, Italy with 6.71%, United Kingdom with 4.47%, Holland with 3, 51% and South Korea with 3.19%. These six countries have a total of 63.58% of the institutional bonds of authors and co-authors. The remaining thirty-nine countries, which represent 36.42% of the sample, have less than 8 authors and co-authors who represent less than 2.5% of the corpus that research this theme.
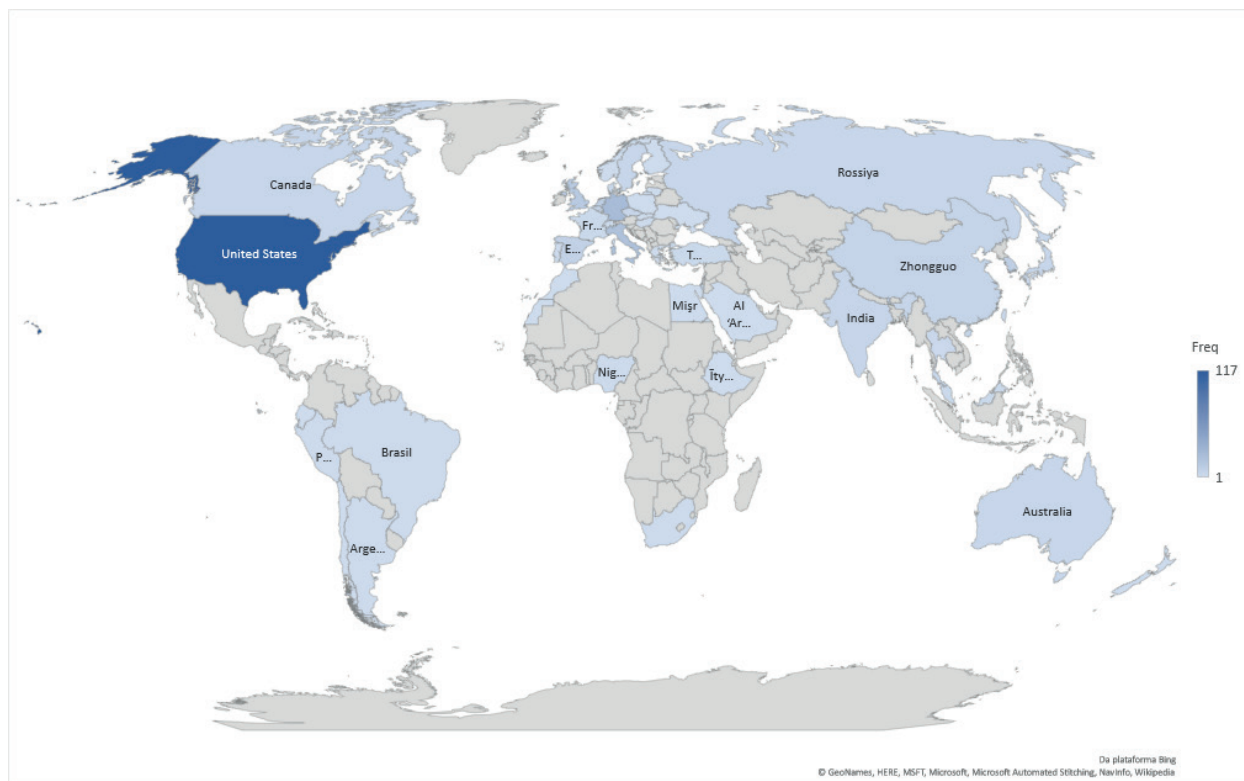

Figure 2 Scientific production of authors and co-authors from 1990 to 2019.

Source: The authors (2020). Estimated by the RStudio package.

Luciana Santos Costa Vieira da Silva • Fabíola Kaczam • Anderson de Barros Dantas • Josenilde Mario Janguia

Starting from the perspective that scientific progress depends on the interaction or collaboration between scientists, we tried to understand how scientists behave, they report, organize, and how they transmit information among themselves, (Kretschmer; Liming; Kundra, 2001). In this context, we use scientific collaboration as a proxy for co-authorship, seeking to identify the existing relationships between researchers and countries as an advantage of data from other authors. Figure 3 shows the behavior of the scientific collaborations of the various researchers and their respective countries, based on the published documents on the theme explored.
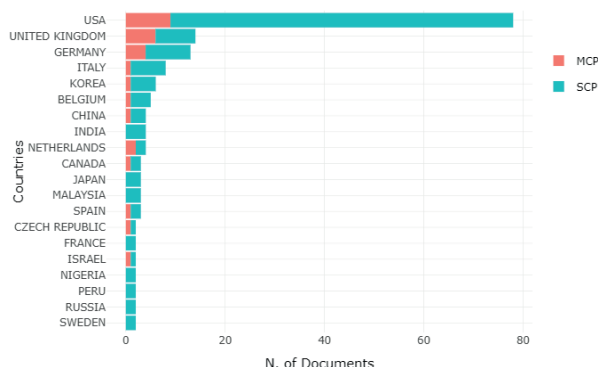

Figure 3 Scientific collaboration between countries in the corpus (1990-2019).

Source: The authors (2020). Estimated by the RStudio package.

Through Figure 3, it is possible to notice that scientific collaborations were divided into two types: those articles that were published by authors from a single country (independent production) - identified by the acronym SCP (Single Country Publications) in blue; and productions with other countries (joint production) - identified by the acronym MCP (Multiple Country Publications) displayed in red. Note that the countries: United States (45.10%; 29.03%), United Kingdom (7.61%; 19.35%), Germany (7.07%; 12.90%), Italy (4.35%; 3.23%) and South Korea (3.26%) ; 3.23%) are those that, respectively, carry out the highest percentage of independent production and joint production of the entire corpus.

Notably, in a group of five countries with low collaboration, they represent 64.05% of the authors of the countries evaluated. Another aspect to be noted refers to the fact that 25 countries do not have any type of scientific collaboration, which we can

deduce that there is room for this theme to be further explored in scientific terms, especially if related to other transversal themes.

It is worth noting that the collaborative networks between research institutions are also relevant for the area of startups, given that the position of the authors and co-authors provide information about who they published with and also highlight the thematic proximity between the authors and research networks.

8 ANALYSIS OF AUTHORS' PRODUCTIVITY

The productivity of authors and co-authors was evaluated based on the component citations of the textual corpus. For this, the bibliometric indicator quantity of citations received was used, as suggested by Zupic and Cater (2015), as a way of showing the measure of influence, since the authors cite the documents they consider to be important. Figure 4 shows the authors' productivity analysis that refers to the use of the quote in a temporal manner.
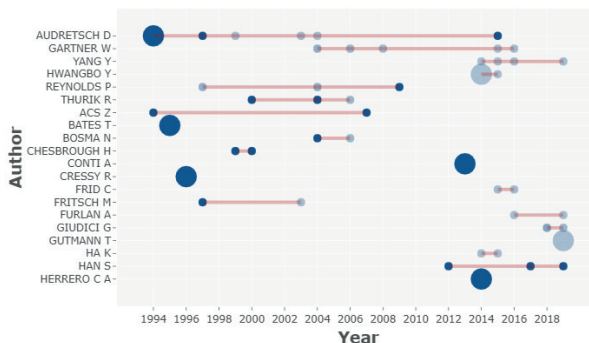

Figure 4 Articles most cited over time.

Source: The authors (2020). Estimated by the RStudio package.

Perform through Figure 4 that the largest the size of the spheres, the greater the frequency of occurrence in which a given author is evidenced in that year. Note that the authors Audretsch (1994), Bates (1995), Cressy (1996), Conti (2013), Herrero (2014), Hwangbo (2014), and Gutmann (2019) have a frequency of occurrence equal to 2 in the different years. The intensity in terms of the sphere's colors, on the other hand, shows that the more intense the color, the greater the number of citations received

by the author over time. Figure 5 shows the 20 most cited articles in the corpus according to the number of citations received by the authors.
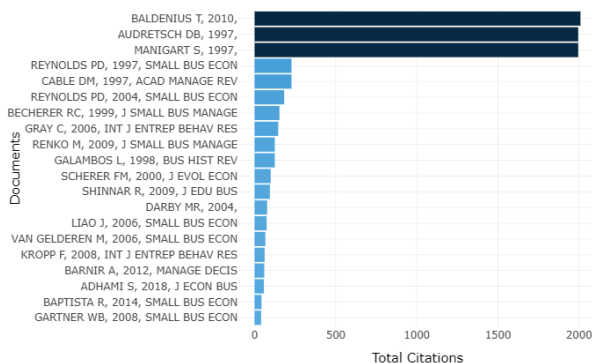

Figure 5 Most cited articles and total citations (1990 - 2019)

Source: The authors (2020). Estimated by the RStudio package.

Notice in Figure 5 that the article that deserves more emphasis is Baldenius and Meng (2010), with the work entitled "Signaling firm value to active investors" published in the Review of Accounting Studies and which received a total of 2,009 citations. The authors show that investors provide risk-sharing and value-added efforts in the form of consultancy, networking, monitoring, etc., and that there is a conflict that can lead to stable and economically significant balances for startups. They also show that more established companies, with access to various types of the mentioned values, can achieve both objectives and still decide to renounce the efforts of the investor. They concluded that by identifying the conditions under which the types of value-added signs increase well-being, this helps to guide investors' efforts towards more advantageous ventures.

Another prominent work seen in the same Figure refers to that produced by Audretsch and Elston (1997), entitled "Financing the German Mittelstand" published in Small Business Economics and received a total of 1,995 citations. The work aimed to demonstrate how small and medium-sized companies in Germany are financed. The authors show how German success is partly attributed to a finance system widely used by complementary institutions designed to meet the financial needs of large and small companies. Finally, they showed that even with financial liquidity constraints, the German system appears to

be particularly deficient in channeling funds to new start-ups in newer industries.

The third work presented in Figure 5 that deserves to be highlighted was published by Manigart and Struyf (1997), entitled "Financing High Technology Startups in Belgium in Explorative Study", was published in Small Business Economics and obtained a total of 1,995 citations. This is an exploratory study on the financing of 18 Belgian high-tech startups. This study shows how the most important sources of financing at the beginning of the business are entrepreneurs and banks, and the sources that most provide the largest amounts of financing are venture capital companies and private investors, with the government, other companies and universities have a more limited role.

It is worth mentioning that in the analysis of this textual corpus, a total of 8,540 citations from all authors were considered and only the three authors highlighted above have a representation of 70.25% of all scientific production analyzed in the period considered.

Finally, when verifying the validity of the Lotka authors' productivity law, it was realized in the light of the data collected that 463 authors representing 93.35% of the total produced even a scientific article, while 6.65% of the investigated authors produced more than one job.

9 ANALYSIS OF PERIODICALS PRODUCTIVITY

The search for evaluating the productivity of scientific journals, based on their citation indicators, is gaining more and more important in the academic community because an overview of a specific thematic area or area of knowledge is drawn up. In this context, Bradford's Law is invoked making it possible to investigate the degree of relevance of journals and, at the same time, that superior quality of journals is correlated with many articles published on a given topic (Machado Junior & Souza, 2016).

It should be noted that the 75 journals in this study were distributed in three zones in order of decreasing productivity and each zone has 1/3 of the total number of articles (228/3), approximately 76

Luciana Santos Costa Vieira da Silva • Fabíola Kaczam • Anderson de Barros Dantas • Josenilde Mario Janguia

articles per zone. In the first zone, we find 4 highly productive journals, with emphasis on Small Business Economics with 46 published articles from the corpus; in the second zone there are 17 journals where we highlight the Journal of Small Business and Enterprise Development with 9 published articles and, in the third zone, we find 54 journals with less productivity, with only 2 published articles, with the Academy of Management Review as the highlight.

Note that 83 articles in the corpus that represent 36.4% of the sample are in Zone 1; in Zone 2, 74 articles are allocated, making a total of 32.46%, and in Zone 3, 71 articles are allocated, representing 31.14% of the total articles comprising the corpus. Notably, there is no direct relationship between the Bradford zones and Scimago quote quartiles, given that there are journals and, consequently, articles in the most different quote quartiles that have been allocated in the three zones or vice versa. It should also be noted that 73.3% of the analyzed articles are in journals located in the first two citation quartiles (Q1 and Q2) according to SJR (2019), which represents a highly qualified representative sample.

10 ANALYSIS OF THE WORDS OF THE TEXTUAL CORPUS

The word cloud aims to organize in a graph the words that appear most frequently in the abstracts of articles in the corpus. Such words are grouped in a cloud that facilitates the understanding of lexical content and demonstrates their representativeness in the sample, as mentioned by Ratinaud (2009) and Camargo and Justo (2013).

The word cloud was built using the RStudio package, while similarity was estimated using the IRAMUTEQ program (Interface for R pour les Analyzes Multidimensionnelles de Textes et de Questionnaires), which is hosted by the software R (Ratinaud, 2009), as shown in Figure 6.

Figure 6 Word cloud of the textual corpus

Source: The authors (2020). Estimated by the RStudio package.

In Figure 6 we have the main words present in the abstracts of studies on Startups, considering only the active forms and the occurrence for the formation of the cloud. For Camargo and Justo (2013), the more centralized and the larger the size of the word, the greater the evocation by the subjects, in a similar way, the smaller and further away from the center it is, the less it is evoked.

A total of 10,088 words were evidenced in the construction of the word cloud, which stands out: Startup with a frequency of 447 words that represent 4.43%, Business with 339 that represents 3.36% of the total words analyzed, Firm with 267 words, making 2.65% of the total; Study with 216 words, with 2.14% of words, Entrepreneur with 195 words that represent 1.93%, Entrepreneurial with 179 representing about 1.77% of words, Research with 166 that represent 1.65% of words, and Paper with 155 that makes up 1.54% of the words evaluated.

In addition to the previous investigation, similarity analysis was carried out, as can be seen in Figure 8, which is based on graph theory, being able to identify its co-occurrences between words. The results of similarity lead to connections between the main words and make it possible to identify the structure of the textual corpus through the built network, (Camargo and Justo 2013; Ratinaud 2009).

the frequency of occurrence of a particular keyword, while the strength of the relationship between the nodes can be assessed by the proximity between them.

Clustering allows you to assess the relationship between words. In this sense, the parameters defined in the VOSviewer software were: "co-occurrence", for the type of analysis; "All keywords" for the analysis unit, and "full counting" for the counting method, assigning the same weight to each link between words. For the formulation of the keyword network, a minimum frequency of 5 occurrences was assigned, which resulted in 27 nodes with good visualization, distributed in 4 clusters, as shown in Figure 8.
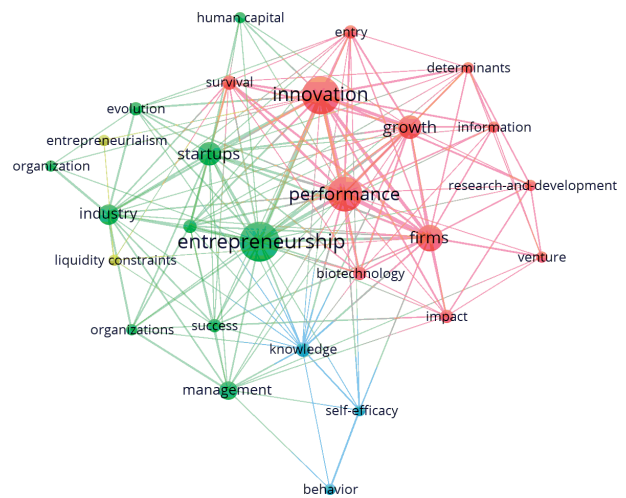

Figure 7 Similarity of the textual corpus

Source: The authors (2020). Estimated by the RStudio package.

Based on the similarity analysis of the words presented in Figure 6, it is possible to notice the structure of the construction of the textual corpus, as well as the most important terms, which are located close together and form a nucleus, coinciding with the words generated in the word cloud. , which are: Startup, Business, Firm, Study, Entrepreneur, Entrepreneurial, Research, and Paper. It is important to note that the thickness of the line arranged between the words indicates the intensity of the connection between them, while the Jaccard index shows the proportion between two words to the total sample (Kent and Coker 1992).

For Kent and Coker (1992), the Jaccard index estimated with a value greater than 0.50 shows a high similarity between the words. Following this assumption, the analyzed similarity can be considered low among the different words evaluated; proof of this was the highest value found for the Jaccard index between the words Startup and Business, which was 0.07. This low similarity can be justified because these words are presented in isolation in the articles that make up the textual corpus.

Regarding the grouping of words, this type of analysis aims to map the possible research themes related to the "startup" theme. It is noteworthy that the size of the node to be estimated is proportional to

Figure 8 Corpus keyword network with "nodes" calculated according to the frequency of occurrence

Source: The authors (2020). Estimated by the RStudio package.

In Figure 8, the lines that link the most relevant keywords within the same grouping, and the strength of the words are represented by connections. It is worth noting that the different colors of the spheres represent the respective clusters, while the size of the spheres indicates their importance when compared to the universe presented. The same goes for common themes that tend to get closer and closer. Thus, it is possible to affirm that the startup theme is, in fact, comprehensive with wide possibilities of approach in the academic sphere.

In the case of the relationship between Innovation and Performance, for example, which were the two words that stand out the most, it is clear that technological innovation (product and process innovations) can improve the performance of an

organization, since the innovation can be seen as a mechanism for competitiveness and economic development, serving as a driving force for creating distinctive value for products, services, and brands, as attested by Rivasa *et al.* (2019).

Two other words that stand out in the formulation of clusters and their respective links refer to startups and entrepreneurship. This can be explained due to the favorable context for the growth of entrepreneurship with the whole ecosystem of startups, mainly with universities where there is a large part of the scientific literature on this subject.

In this context, despite a high range of profiles for startups, it is a consensus among researchers (Dahlstrand, 2007; Saemundsson, 2004) that startups are closely linked to the concept of entrepreneurship, being an important vector for regional development (Venkataraman, 2003), being considered as techno-enterprises that can play a central role in the transformation of a given region.

In general, in approximately 68% of startups evaluated from the textual corpus, we realized that they can be characterized based on the following peculiarities:

a) They are organizations characterized as a group of new companies, that is, relatively young and inexperienced when compared with more stable and mature companies in organizational development;

b) They are those that are in their initial stage and are susceptible to the influence of several factors such as investors, supplier customers, partners, etc., and should think strategically about how to act;

c) It concerns a group of dynamic startups that work with innovative technologies aimed at the development of disruptive products or services.

The count of keywords used, their citations, and connectivity (Total Link Strength) for the formulation of clusters were also evaluated. Note that of the three most attractive words in terms of quotes were "entrepreneurship", with a total of 37 occurrences, corresponding to 13.45%, followed by the word "Innovation", with 33 occurrences corresponding to 12% of the total of 275 occurrences existing, followed by the

word "Performance", which corresponds to 10.54% and "Companies", with 19 occurrences, which corresponds to 6.91% of the total occurrences received for the formation of the four groups.

In terms of connectivity, that is, those terms that end up being linked to other terms, "Innovation" deserves mention with a total of 91 connections, which corresponds to 12.2% of all 743 existing connections for the formulation of the 4 clusters, followed by the term "Performance" with a total of 78 connections, which corresponds to 12.5% of the existing total, while the themes "Entrepreneurship" and "Companies" obtained 69 connections corresponding to 9.3% of the total terms, "Crescimento" obtained 51 connections corresponding to 6.9% of the total, "Startup" obtained 46 connections corresponding to 6.2% and the term "Industry" totaled 34 connections, which corresponds to 4.6% of the total connections in the corpus. The remaining terms had a total of 273 connections, representing 36.7% of the overall total of existing connections.

In order to validate the clusters that were formulated in terms of proposals and approaches, descriptive validation was carried out with regard to the identification of relevant studies in the databases of consulted journals. The procedure adopted, as suggested by Sandelowski and Barroso (2007), consists in the search for studies by two reviewers, independently.

In this context, the search was again carried out with the string "Startup *", in the bases of journals Web of Science and Scopus, where 10 articles were extracted, randomly, that address this theme in English, in February 2020. From this search it was possible to verify that the terms with the highest recurrence in the clusters, generated with the aid of the VOSviewer software, were identified in other researches that deal with the same themes. This comparison brings the guarantee that the central approach of the present research, evidenced by the keywords, is in the agenda of the main researches that permeate the same focus.

11 DIRECTIONS FOR FUTURE RESEARCH ON STARTUPS

Based on the steps proposed to carry out the Systematic Literature Review (RSL), as well as the results of previous analyzes, the content analysis of the primary studies was carried out to decipher some directions of future research, meeting the objective of this article.

Figure 9 shows the 345 most frequent terms extracted from the abstracts of the analyzed corpus, categorized according to the objective of the study, the applied methodology and the results found, being constructed based on the word incidence matrix, in which the size of the terms it is proportional to its occurrence.

| A – objective | B – methodology | C – results |

Figure 9 Most frequent terms extracted from the abstracts of the *corpus*

It can be seen in Figure 9, that when evaluating the objectives of the articles in the corpus, it is possible to show that the research has a focus on business aimed at startups, with 77 occurrences, which is natural because it is the focus of the research. Regarding the methodology, we noticed that the overwhelming majority of the works are quantitative, with a total of 138 occurrences, that is, 61.33% of the total analyzed works, while the qualitative studies correspond to 38.67%. Among the outstanding metrics, we have the analysis of multiple linear regression, binary or multinomial logistics, and structural equation technique, with a total of 98 articles (43.56%) of the published works. Concerning the results found, we noticed that they portray the terms that are in line with the central theme of the article, in addition to reporting some directions for the construction of future works.

Although the present work presents a systematic review of the exhaustive literature to this field of research, some directions for the elaboration of future research are proposed based on the articles of the textual corpus:

Investigate based on what Szerb and Voros (2019) report, which factors such as confirmation bias or the illusion of control (heuristics) play a role in entrepreneurs' exaggerated growth expectations or change over the business life cycle;

Explore the potential of applying business models to promote user entrepreneurship and social entrepreneurship, or specific forms inhibited by limited competencies and entrepreneurial readiness or significantly influenced by relevant contextual factors, based on the perspective of Del Bosco, Chierici and Mazzucchelli (2019);

Explore in longitudinal terms how the development of new ventures can be affected by organizational factors, such as entrepreneurial cognitions and firm behaviors, as recommended by Sunny and Shu (2019);

Carry out a longitudinal study, as mentioned by Sampedro, Fernandez-Laviada, and Crespo (2014), aiming to examine which variables can give rise to changes in entrepreneurial intention, in addition to replicating the study in different countries with economic, cultural and legal characteristics distinct,

which can affect the perception of advantages and disadvantages of entrepreneurship; examine the differences in the repercussions of knowledge about entrepreneurship (spillovers) in different sectors of economic activity, simultaneously investigating gender differences in entrepreneurship in various dimensions, as advocated by Goel and Saunoris (2017).

12 CONCLUSIONS

This work had the objective of elaborating a Systematic Literature Review (RSL), from which the theme „Startup *" was mapped. The analyzes were performed based on the three classic laws of bibliometry that measure the productivity of authors; it measures the productivity of journals and it measures the frequency of occurrence and co-occurrence of certain words in a text.

The descriptive analysis of the research corpus revealed 228 articles, with a total of 313 authors and co-authors that were distributed in 25 countries, with a concentration in the United States (37.28%). The period of distribution of the articles comprises the years from 1990 to 2019, with emphasis on the year 2015, with 48.43% of the articles published.

We also realized that the United States once again stands out with the exploration of this theme in terms of scientific collaboration, while 25 countries do not have any type of scientific collaboration, which we can deduce that there is room for this theme to be further explored in scientific terms, especially if we report to other transversal themes.

When evaluating the validity of Lotka's authors' productivity law, it was perceived in the light of the collected data that 463 authors representing 93.35% of the total produced a scientific article, while 6.65% of the investigated authors produced more than one work.

Regarding the distribution of production among the journals, the results confirm Bradford's Law, given that in the first zone (Z1), 4 highly productive journals were found; in the second zone (Z2) 17 journals were seen and in the third zone (Z3) 54 journals with lower productivity were found. We also noticed that 73.3% of the analyzed articles were located in the first two citation quartiles (Q1 and Q2) according to SJR (2019), which is configured as a representative sample of high qualification.

Regarding the grouping of words, translated by the Zipf Law, this type of analysis aimed at mapping possible research themes related to the "startup" theme, which is comprehensive with wide possibilities of approach in the academic sphere. Other words that had a high occurrence were Innovation and Performance.

It was noticed by reading the articles in the corpus that technological innovation can improve the performance of an organization, since innovation can be seen as a mechanism for competitiveness and economic development, serving as a driving force for creating distinctive value. for products, services, and brands, as attested by Rivasa *et al.* (2019).

Two other words that stood out in the formulation of clusters and their respective links refer to startups and entrepreneurship. This was explained due to the favorable context for the growth of entrepreneurship in the whole ecosystem of startups, mainly in universities where there is a large part of the scientific literature on this subject.

In general, it was noticed that approximately 68% of the startups evaluated from the textual corpus could be characterized (Typology) as a group of new companies; those that are in their initial stage and are susceptible to the influence of several factors such as investors, customers, suppliers, partners, etc., and should strategically think about how to act in addition to being a group of dynamic startups and working with innovative technologies aimed at developing disruptive products or services.

Additionally, the 345 most frequent terms extracted from the abstracts of the analyzed corpus were extracted from the articles of the corpus, categorized according to the objective of the study, the methodology applied in the works and the results obtained, is built based on the word incidence matrix, where the size of the terms is proportional to their occurrence. Based on the results found in this analysis, we realized that they portray the terms that are consistent with the central theme of the article, in addition to reporting some directions for the construction of future works.

Although the present work presents a systematic literature review consistent concerning this field of research, some directions for the elaboration of future

research have been proposed based on the articles of the textual corpus, as defended by the academic community (de Almeida Biolchini, Mian, Natali, Conte, & Travassos, 2007; Kitchenham 2004; Tranfield *et al.* 2003): (i) the search string limited the results to publications in the English language; (ii) only journals placed in the first and second quartile of citation according to the SJR were included, so that it cannot be disregarded that relevant publications may have been published in other formats such as books and conference articles. The study on growth harking in startups can also be seen as a suggestion for future work, as it is a growing trend in the technology area.

■ REFERENCES

Almeida, C. P. B., & Goulart, B. N. G. (2017). Como minimizar vieses em revisões sistemáticas de estudos observacionais. *Revista CEFAC*, São Paulo, v. 19, n. 4, p. 551-555, ago. 2017. DOI: 10.1590/1982-021620171941117.

Audretsch, D. B. (1994). Business survival and the decision to exit. *Journal of the Economics of Business*, 1(1), 125-137. DOI: https://doi.org/10.1080/758540503

Baldenius, T., & Meng, X. (2010). Signaling firm value to active investors. *Review of Accounting Studies*, 15(3), 584-619. DOI: https://doi.org/10.1007/s11142-010-9130-7

Bates, T. (1995). Analysis of survival rates among franchise and independent small business startups. *Journal of Small Business Management*, 33(2), 26.

de Almeida Biolchini, J. C., Mian, P. G., Natali, A. C. C., Conte, T. U., & Travassos, G. H. (2007). Scientific research ontology to support systematic review in software engineering. *Advanced Engineering Informatics*, 21(2), 133-151. DOI: https://doi.org/10.1016/j.aei.2006.11.006

Blank, S., & Dorf, B. (2012). The startup owner's manual: The step-by-step guide for building a great company. K & S Ranch. *Inc, California. PubMed Abstract OpenURL.*

Camargo, B. V., & Justo, A. M. (2013). IRAMUTEQ: um software gratuito para análise de dados textuais. *Temas em psicologia*, 21(2), 513-518. ISSN: 1413-389X

Campanario, J. M. (2015). Providing impact: the distribution of JCR journals according to references they contribute to the 2-year and 5-year journal impact factors. *Journal of Infometrics*, v. 9, 398-407. DOI: https://doi.org/10.1016/j.joi.2015.01.005

Conti, A., Thursby, J., Thursby, M. (2013). Patents as Signals for Startup Financing, *The Journal of Industrial Economics*, v. 61, John Wiley & Sons Ltd. DOI: https://doi.org/10.1111/joie.12025

Cooper, B., & Vlaskovits, P. (2010). Entrepreneur's Guide to Customer Development. ISBN-13: 978-0-9827436-0-7

Cressy, R. (1996). Are business startups debt-rationed? *The economic journal*, 106(438), 1253-1270. DOI: https://doi.org/10.2307/2235519

Crowne, M. (2002). Why software product startups fail and what to do about it. Evolution of software product development in startup companies. *IEEE Xplore.* 18-20 Aug. Doi: 10.1109/IEMC.2002.1038454.

Cohen, D. G.; Feld, B. (2010). Do More Faster: Techstars Lessons to Accelerate Your Startup. *publisher: Wiley*. October. ISBN: 1119583330, 9781119583332

Dahlstrand A. L. (2007). Technology-Based Entrepreneurship and Regional Development: The Case of Sweden. *European Business Review*. Bradford: 2007. Vol. 19, n. 5; p. 373. DOI: https://*doi*.org/10.1108/

Del Bosco, B., Chierici, R., & Mazzucchelli, A. (2019). Fostering Entrepreneurship: An Innovative Business Model to Link Innovation and New Venture Creation. *Review of Managerial Science*, 13(3), 561-574. DOI: https://doi.org/10.1007/s11846-018-0318-8

RCA

Figueiredo, E. B., de Cássia Catini, R., & Mendes, L. M. (2018). *Mineração de Textos*: Análise de Sentimento em Redes Sociais-Revisão Sistemática. Anais do WCF, Vol 5, pp 24-29. ISSN 2447-4703

Goel, R. K., & Saunoris, J. W. (2017). Dynamics of Knowledge Spillovers From Patents to Entrepreneurship: Evidence Across Entrepreneurship Types. *Contemporary Economic Policy, 35(4), 700–715.* Doi:10.1111/Coep.12224.

Gutmann, T. (2019). Harmonizing corporate venturing modes: an integrative review and research agenda. *Management Review Quarterly*, DOI: https://doi.org/10.1007/s11301-018-0148-4.

Hermanson, B. (2011). O que é uma startup. *São Paulo: Mundo Sebrae.*

Herrero C. A., Sampedro, I. R. & Fernández-Laviada, A. (2014). *Entrepreneurial intention: perceived advantages and disadvantages. Academia Revista Latino americana de Administración, 27(2), 284–315.* DOI:10.1108/arla-09-2013-0144.

Hwangbo, K.S. & Seo, W.S. (2014). The current status and improvement strategy of the Korean start-up accelerators, *Academy of Entrepreneurship Journal.* questia.com

Kent M. & Coker P. (1992). *Vegetation Description and Analysis: A Practical Approach.* Wiley, Baffins Lane, Chichester, West Sussex, UK, p. 363.

Kim, L. (2005). *Da imitação à inovação: a dinâmica do aprendizado tecnológico da Coréia.* Unicamp.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33,* 1-26.

Kollmann, T., Stöckmann, C., Hensellek, S., & Kensbock, J. (2016). *European startup monitor 2016.* Universität Duisburg-Essen Lehrstuhl für E-Business. ISBN: 978-3-938338-17-9

Kohler, T. (2016). Corporate accelerators: Building bridges between corporations and startups. Business Horizons, Vol. 59, No. 3, pp. 347-357. DOI: https://doi.org/10.1016/j.bushor.2016.01.008

Kretschmer, H., Liming, L., & Kundra, R. (2001). Foundation of a global interdisciplinary research network (COLLNET) with Berlin as the virtual centre. *Scientometrics*, *52*(3), 531-537. DOI: https://doi.org/10.1023/A:1014268505676

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science and Technology*, 60(7), 1327-1336. DOI: https://doi.org/10.1002/asi.21024

Longhi, F. (2011). A história da revolução das startups. *Imasters. Disponível em: http://imasters. com. br/ artigo/20027/mercado/a-historia-da-revolucao-das--startups. acedido em*, 14.

Machado Junior, C., & Souza, M. T. S. de. (2016). The Laws of Bibliometrics in Different Scientific Databases. *Revista de Ciências da Administração*, *18*(44). DOI: http://dx.doi.org/10.5007/2175-8077.2016v18n44p111

Manigart, S., Struyf, C. (1997). Financing High Technology Startups in Belgium: An Explorative Study. *Small Business Economics* **9,** 125–135. https://doi.org/10.1023/A:1007967722346

Nager, M., Nelsen, C., & Nouyrigat, F. (2013). Startup Weekend: Como levar uma companhia do conceito à criação em 54h.

Oliveira, M. M. (2007). *Como fazer pesquisa qualitativa*. Petrópolis, RJ: Vozes.

Padrão, L. C., & Andreassi, T. (2013). O desempenho de startups de base tecnológica: um estudo comparativo em regiões geográficas brasileiras. *Revista da Micro e Pequena Empresa*, *7*(2).

Perin, B. (2015). *A revolução das startups*: o novo mundo do empreendedorismo de alto impacto. Rio de Janeiro, RJ: Alta Books. 172p.

Ratinaud, P. (2009). *IRAMUTEQ*: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires [Computer software]. Retrieved from http://www.iramuteq.org.

Ries, E. (2012). *A Startup Enxuta*. Leya.

Rivasa, J., Pereab, J.M., De-Pablos-Herederoc, C., Angonb, E., Barbab, C., & García, A. (2019). Canonical correlation of technological innovation and performance in sheep's dairy farms: Selection of a set of indicators. *Agricultural Systems*, v. 176. DOI: https://doi.org/10.1016/j.agsy.2019.102665

Rompho, N. (2018). Operational performance measures for startups. *Measuring Business Excellence*. DOI: https://doi.org/10.1108/MBE-06-2017-0028

Saemundsson R. J. (2004). Technical knowledge-seeking in a young and growing technologybased firm: incentives and direction. *International Journal of Innovation Management.* Vol. 8, n. 4; pp. 399–429. DOI: https://doi.org/10.1142/S136391960400112X

Sampaio, R. F., & Mancini, M. C. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. Brazilian Journal of Physical Therapy, 11(1), 83-89. DOI: https://doi.org/10.1590/S1413-35552007000100013

Sampedro, I. R., Fernández-Laviada, A., & Crespo, Á. H. (2014). Entrepreneurial intention: perceived advantages and disadvantages. *Academia Revista Latinoamericana de Administración*. DOI: https://doi.org/10.1108/ARLA-09-2013-0144

Sandelowski, M., Barroso, J., & Voils, C. I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in nursing & health*, *30*(1), 99-111. DOI: https://doi.org/10.1002/nur.20176

Schwab, K. (2017). *The fourth industrial revolution.* Currency. DOI: https://doi.org/10.5430/ijfr.v9n2p90

Scimago Journal & Country Rank (SJR). 2018. https://www.scimagojr.com/

Spender, J. C., Corvello, V., Grimaldi, M., & Rippa, P. (2017). Startups and open innovation: a review of the literature. *European Journal of Innovation Management*. DOI: https://doi.org/10.1108/EJIM-12-2015-0131

Sunny, S. A., & Shu, C. (2019). Investments, incentives, and innovation: geographical clustering dynamics as drivers of sustainable entrepreneurship. *Small Business Economics*, *52*(4), 905-927. DOI: https://doi.org/10.1007/s11187-017-9941-z

Szerb, L., & Vörös, Z. (2019). The changing form of overconfidence and its effect on growth expectations at the early stages of startups. *Small Business Economics*, 1-15. DOI: https://doi.org/10.1007/s11187-019-00297-9

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207-222. DOI: https://doi.org/10.1111/1467-8551.00375

Venkataraman S. (2004). Regional transformation through technological entrepreneurship. *Journal of Business Venturing*. New York: vol. 19, n. 1; pp. 153-167.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472. DOI: 10.1016/j.jbusvent.2003.04.001

RCA