



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Research lines on the impact of the COVID-19 pandemic on business. A text mining analysis

Patricia Carracedo^{a,*}, Rosa Puertas^b, Luisa Marti^b

^a Universidad Internacional de Valencia, Área de empresa, c/Pintor Sorolla, 21, Valencia 46022, Spain

^b Universitat Politècnica de Valencia, Departamento de Economía y Ciencias Sociales, Camino de Vera, s/n 46002 Valencia Spain

ARTICLE INFO

Keywords:
COVID-19
Text mining
Cluster
Business
Marketing

ABSTRACT

COVID-19 has brought about a marked slowdown in global economic development. Companies have been forced to adopt new managerial guidelines to adapt to the difficult conditions and to survive in this “new normal”. The recent and still scarce literature in this field seeks to provide suitable solutions to prevent irreparable disruption and help strengthen business, but does not apply advanced statistical methods to that end. The aim of this paper is to identify the current research lines developed around COVID-19 and their impact on the business environment, applying text mining methodology. The analysis, which uses statistical software R, focuses on systematic review of studies published in prestigious journals of business and marketing areas. In light of the results obtained, three different areas of intervention were identified. The common thread that runs through all of them is the need to introduce new forms of action to improve citizens’ quality of life.

1. Introduction

The rapid spread of COVID-19 meant that in little over three months the whole world was struggling to curb not only the disease but also the harsh economic consequences of the measures taken to do so. Governing authorities in almost all countries were forced to limit the free movement of the population both nationally and internationally, bringing the economy to a total standstill (Nicola et al., 2020). This situation has underscored the fragility of the foundations of the 21st century economy, which is characterized by major technological advances and remarkable globalization, centred around globally interconnected production chains seeking maximum profitability.

The lockdown measures led to widespread economic collapse with significant repercussions on production and employment, and a severe impact all branches of activity due to the sharp decline in consumption. The most developed countries have seen their positive economic growth dip into the red with marked rises in unemployment and an increase in social inequalities. Comparisons are inevitable: some go so far as to equate the uncertainty generated by this pandemic with that caused by the Great Depression of 1929–33, and far exceeding that associated with the financial crisis of 2008–09.

Companies face an unprecedented challenge. Their survival depends on the adoption of management strategies that will allow them to

overcome the sharp drop in orders and the pressure of costs stemming not only from rent, wages and taxes, but also those associated with the rise in the price of raw materials given the significant decline in suppliers (Wen, Wei, & Wang, 2020). Moreover, there have been important changes in consumption habits, with people attempting to avoid physical contact in order to prevent possible contagion, leading to a pressing need for a global transformation to take on this new environment (Sheth, 2020). Responsive systems are required to facilitate the rapid adaptation of supply chains to accommodate changing consumer demands (Sjodin, Parida, Kohtamaki, & Wincent, 2020; Gordon, Ramic, Rohrbeck, & Spaniol, 2020).

There has been a rapid emergence of research in this area in an attempt to find alternative solutions and facilitate the transformation of companies aligned with the new scenario, thereby ensuring their survival under the best conditions. Pantano, Pizzi, Scarpi, and Dennis (2020) summarize the challenges faced by retailers, in order to provide guidelines that can enable them to deal with the disruptions caused by the pandemic in both the short and medium term. In this regard, marketing strategies play a key role (Kirk & Rifkin, 2020; Wang, Hong, Li, & Gao, 2020). The current situation should be taken as an opportunity for companies to put in place corporate social responsibility, consumer ethics and a marketing philosophy. This would open up more attractive strategies for organizations and consumers (He & Harris, 2020;

* Corresponding author.

E-mail address: patricia.carracedo@campusviu.es (P. Carracedo).

<https://doi.org/10.1016/j.jbusres.2020.11.043>

Received 24 September 2020; Received in revised form 19 November 2020; Accepted 21 November 2020

Available online 26 November 2020

0148-2963/© 2020 Elsevier Inc. All rights reserved.

Woodside, 2020). Other key aspects such as human resource management implications (Leung, Sharma, Adithipyangkul, & Hosie, 2020) or gender equality (Carnevale & Hatak, 2020) in pandemic situations have also been analysed recently.

The European countries that have been hardest hit by death due to COVID-19, such as, Italy (12,430 deaths at the beginning of April¹), Spain (8,189), France (3,523) or the United Kingdom (2,453), among others have seen how their tourism sector has experienced a total shutdown of activity with a major direct impact on the rest of the economy. Studies such as that carried out by Sigala (2020) highlight the effects of the pandemic on the sector, in line with the guidelines set out by Nepal (2020), which call for this situation to be viewed as a reset opportunity and a chance to provide new sustainable services. Many more examples can be cited of studies in the literature where researchers have focused on COVID-19 and its consequences for companies (Krishnamurthy, 2020; Mullins, 2020; Eggers, 2020; Sharma et al., 2020a; Bacq, Geoghegan, Josefy, Stevenson, & Williams, 2020).

The urgent need to find workable solutions for businesses brings the discipline of data mining to the fore. The ongoing development of research on this issue requires a global analysis of textual data in order to detect patterns from which conclusions can be drawn, and thus create the best possible conditions for orienting decision-making. For this reason, the main objective of this paper is to identify the most current research fields focused on the impact of COVID-19 on business. To that end, the text summarization technique of text mining was first applied, in order to identify the most frequently mentioned terms in business research papers that study COVID-19. Next, a text clustering method using the statistical software R Core Team (2020) was applied to identify groups of most frequent terms or research lines. The subject of the analysis was articles published in business journals indexed in the Web of Science (WoS) database. Specifically, 16 papers published in the Journal of Business Research and Business Horizons have been analysed.

Currently, many researchers are turning their full attention to COVID-19. Lopez, Vasu, and Gallemore (2020) use natural language processing, text mining and network analysis to explore the policies implemented to manage the COVID-19 analyzing information transmitted via Twitter. Cheng, Cao, and Liao (2020) demonstrate that information specialists can support health and medical community by applying text mining technique with latent Dirichlet allocation procedure to perform an overview of a mass of coronavirus literature. Similarly, Ulm and Nelson (2020) use text-mining and manual curation approach to comb and summarize the information from existing clinical trials and previous efforts to develop therapies against related betacoronaviruses. We emphasize that none of them performs a systematic literature review and employs a text clustering method using the statistical software R.

Therefore, it can highlight three novel contributions: (1) text-mining has been used in one of the areas most punished by the pandemic, the company, in order to be able to guide the guidelines that allow to solve the hard consequences caused by the confinement, (2) do a systematic review of business papers of high impact papers indexed in the first quartile of WoS so that the analysis of informative and sometimes sensationalist texts were avoided, (3) Free statistical software programming language R is proposed as a statistical tool.

The rest of this paper is organized as follows. Section 2 presents the database or corpus built from the selected papers, and explains the text summarization technique used to identify the most frequent terms followed by the text clustering method used to determine the groups of the most frequently mentioned terms. Section 3 outlines the main results of the research. Finally, Section 4 presents the most important conclusions of the study.

2. Data and methods

To carry out the empirical research, a selection must be made of recent studies in the literature that meet the analysis criteria. The new measures taken and their direct impact on the business world have brought to light the need for updated management policies to address the difficult economic situation, the first signs of which have not taken long to emerge. The journals for the database have been selected according to the following criteria:

- Journals indexed in the WoS database. WoS is maintained by the Thomson Reuters Institute for Scientific Information (ISI) and contains more than 21,100 journals dating back to 1990, most of them written in English with a high impact factor (Reuters, 2019).
- Journals in the Business Management and Accounting area and Marketing category. These are two business fields that play a key role in ensuring companies' survival. The development of management and marketing measures that facilitate companies' adaptation to the instability of the environment is essential in the process of overcoming the consequences of the pandemic.
- Journals ranked in the top quartile (Q1).
- Journals whose topic of interest is exclusively Business. Selecting these involved reading the titles, aims and scope.

These are generic criteria that have made it possible to narrow down the literature under study, so that the application of these filters yielded the following journals: Journal of Business Research, Journal of World Business, International Business Review and Business Horizons.

In order to select papers published in these journals whose topic of interest was coronavirus, a search was conducted for articles containing the words *covid*, *coronavirus* or *pandemic* in their abstract, title or keywords. Table 1 shows the business journals with the number of papers selected, ordered from highest to lowest impact factor (IF).

It can be seen that the Journal of Business Research has published the greater number of articles on COVID-related research (87.5% of all selected papers), followed by Business Horizons (12.5%). Journal of World Business and International Business Review have not published any research papers that deal with the COVID-19 pandemic, which highlights a focus of academic attention for business researchers.

Applying all the above search criteria has yielded a corpus consisting of 16 research papers published in two business journals which are indexed in the WoS database. Table 2 shows a summary of each individual article, providing the bibliographic citation, objective, keywords and most important conclusions.

Text analysis, also referred to as text mining, is an interdisciplinary field of activity that combines elements of data mining, linguistics, computational statistics, and computer science (Meyer, Hornik, & Feinerer, 2008). The main difference between data mining and text mining is that data mining usually deals with structured data (stock information, geolocation data, etc.) while text mining works with unstructured or semi-structured data (articles, documents, etc.). In recent years, the latter technique has been a focus of interest because of the increasing amount of text data in social networks, web pages, research papers, and other information applications (Aggarwal & Zhai, 2012). Its principal aim is to obtain high-quality information from processing large volumes of textual data. In this study, the summarization and clustering

Table 1
Business journals in the WoS database.

Journal	Frequency	%	IF
Journal of World Business	0	0.00%	7.28
Journal of Business Research	14	87.50%	7.16
International Business Review	0	0.00%	5.60
Business Horizons	2	12.50%	5.54
Total	16	100.00%	

¹ Data published by European Centre for Disease Prevention and Control.

Table 2
Database: Selected Articles.

Adhikary	Adhikary	Adhikary	Adhikary
Bacq et al. (2020)	Provide evidence from participants and mentors showcasing the value of the time-compressed virtual idea blitz in accelerating social entrepreneurial action.	COVID-19 Entrepreneurial hustle Grand challenges Social entrepreneurship Virtual idea blitz	The results of the event suggest the concept is transferable to a number of different contexts. Universities can play a valuable role in linking professionals, students, and researchers in shared endeavours that can achieve substantial societal benefits.
Carnevale and Hatak (2020)	Discuss the implications of COVID-19 for human resource management.	Human resource management Employee adjustment Well-being Crisis COVID-19	There is little reason to believe its impact on organizational life will be short-lived. Some implications of COVID-19 for employee adjustment and well-being are highlighted.
Donthu and Gustafsson (2020)¹	Introduce special issue about the effects of COVID-19 on business and research	–	People have become more conservative and protective after the pandemic outbreak. Countries are stockpiling things like food, equipment and medicine, or are preparing to produce them locally.
Eggers (2020)	Conduct a literature study on 69 manuscripts that studied SMEs in previous crises and propose ways to overcome economic downturns in the areas of finance, strategy and the institutional environment	Small firms SMEs Crisis Strategy COVID-19	Small firms face a liability of smallness, while crises create additional resource availability and liquidity problems. An SME's potential for more flexible decision-making and closeness to its customer base is beneficial in this regard.
He and Harris (2020)	Examine how the COVID-19 pandemic can influence the development of CSR, consumer ethics and marketing philosophy.	COVID-19 Corporate social responsibility Marketing Consumer ethical decision making Marketing philosophy Business ethics	Whatever the changes, it seems highly likely that the ways marketing has operated in the past will need to change and will do to meet the new reality.
Kirk and Rifkin (2020)	Document some of the many unusual consumer behaviour patterns that came to dominate the early days of the COVID-19 pandemic	Coronavirus COVID-19 Pandemic Hoarding Social distancing Do-it-yourself Digital technology Consumer	Three phases of consumer behaviour is addressed: reacting, coping and longer-term adapting.

Table 2 (continued)

Adhikary	Adhikary	Adhikary	Adhikary
		behaviour	
Krishnamurthy (2020)	Provide a framework to understand business schools' worldwide transformation by recognizing the changes in the university, the business world and the student.	COVID-19 Business school Digital transformation AI-led innovation	The university system has been forced into an unprecedented reshape. Business schools will rise to the occasion and will adopt a leadership role within the university
Leung et al. (2020)	Research the impact of gender equity on public health outcomes using the ongoing COVID-19 pandemic as its research setting	COVID-19 Gender equity Human development Human environment Public health expenditure Public health outcomes	The importance of women's role in managing public health outcomes, with strong positive effects of gender equity and the proportion of women in legislature on public health expenditure, which in turn has a significant impact on the number of diagnosed and critical cases.
Pantano et al. (2020)	Summarize the challenges that retailers are facing during the COVID-19 emergency, from the perspective of both consumers and managers.	Retailing Consumer behaviour COVID-19 Emergency Retail strategy Pandemic	Retailers should focus on the following main areas: rethinking agile retailing, a new role for retailers in society, putting consumers at the core, digital communication
Mullins (2020)	Analyse four tools to help any business owner: (1) Sources and uses of cash analysis, (2) Line-by-line margin analysis of the income statement, (3) Cash-days analysis of the balance sheet, (4) Hidden-cash analysis.	Cash-flow tools Managing cash Recession Cash-days analysis	Recessions can be an opportune time to pick up new customers as well as new employees. If cash flow is well managed and pockets are full, businesses will be well positioned to do these things and to weather the current economic storm.
Sharma et al. (2020a)	Conduct an extensive review of the international business literature to understand the types of uncertainty in businesses and develop strategies to deal with it.	COVID-19 Globalization International business Risk Trade Uncertainty	This crisis has shown the importance of informational uncertainty and the growing role and influence of direct communication and social media, with inconsistent news and communication from different sources causing confusion and panic
Sharma et al. (2020b)	Offer strategic insights in terms of major issues firms are facing and strategic options firms are	COVID-19 Supply chain Strategic insights Twitter	Firms are facing challenges in terms of demand-supply mismatch, technology, and development of a

(continued on next page)

Table 2 (continued)

Adhikary	Adhikary	Adhikary	Adhikary
	contemplating, using text analytics tools.		resilient supply chain.
Sheth (2020)	Examine the impact of the COVID-19 pandemic on consumer behaviour.	COVID Pandemic Consumer habits New regulations for shopping Customer experience	Digital technology which facilitates work, study and consumption will modify existing habits.
Sigala (2020)	Critically review past and emerging literature to help professionals and researchers alike to better understand, manage and valuate both the tourism impacts and transformational affordance of COVID-19.	Tourism COVID-19 Impacts Recovery Resilience Crisis	COVID-19 it can be seen as a transformational opportunity for reforming mindsets in designing and conducting research and for tourism institutions to reset their standards and metrics for motivating and evaluating the purpose, role and impact of tourism research.
Wang et al. (2020)	Develop a typology of the marketing innovation strategies of firms under crisis management in two dimensions: motivation for innovations and the level of collaborative innovations	COVID-19 crisis Firms in China Typology Marketing innovation strategies Dynamic capabilities Resource dependence	When a firm is exposed to a greater degree of external impacts, it is reasonable to choose the responsive strategy that requires the firm to have stronger reconfiguration capabilities and independently optimize existing business
Woodside (2020)	Apply the “ultimate broadening of the concept of marketing” for designing and implementing interventions in public laws and policy, national and local regulations, and everyday lives of individuals	COVID-19 Intervention Experiment Marketing Ultimate broadening	Coordinated, major, national government actions are necessary to mitigate and eliminate pandemics, civil rights violations, guns, corruption, extreme income inequality, misogyny, and national disasters.

¹As this is an introductory article, it does not have any keywords.

techniques of text mining were applied to a collection of previously processed papers, representing a collection of text documents in R. In this context, a document is a major element of text mining that constitutes a group of discrete textual data within a collection (Acharya, 2018).

Given the technical requirements of the method used, the selected articles have to be filtered to avoid terms that could distort the results. Below, we show step-by-step the actions needed to prepare the text documents for analysis.

Step 1. Download the papers in pdf format and convert to txt format for processing.

Step 2. Manually analyse hyphenated words. If they are terms divided at the end of a line, remove the hyphen. If they are hyphenated words, leave the hyphen in.

Step 3. In order to focus the study on the main research topic of the papers analysed, manually remove the reference section and text citation of each one.

Step 4. Convert all terms to lowercase so there are no differences between the same terms.

Step 5. Remove punctuation marks.

Step 6. Delete digits.

Step 7. Remove stopwords which are commonly used in a language but that do not provide information in the text analysis. Some examples in English are *that, then, the, a, an, and*, among others.

Step 8. Stem the corpus to group related terms under their common root.

In line with the research objectives, after processing the database, a list of terms was obtained with the number of times each one appeared in the set of selected papers. The most frequent terms on the list were then analysed to identify the main research topics of interest in the corpus. Second, text clustering was applied to the most frequent terms detected. Clustering is a well-known statistical classification technique used in exploratory data analysis, which has been applied in a variety of scientific disciplines (Täuscher & Laudien, 2018; Han, Zhang, Pirbhulal, Wu, & de Albuquerque, 2019; William & Chang, 2019). The main difference between hierarchical and non-hierarchical clustering methods is that the number of clusters is not predefined in the former. This study applied a hierarchical cluster analysis, where the number of groups was objectively determined. Specifically, the analysis was implemented using the statistical software R Core Team (2020), applying Ward’s minimum variance method with Chi-squared distance. More details about the clustering process can be found in Husson, Lê, and Pagès (2017).

The more terms two papers share, the more similar they are to each other (Nguyen, 2013). The clustering process is based on maximizing the similarity between papers and, consequently, minimizing the Chi-squared distance between them. Chi-squared distance is a Euclidean distance but weighted by the inverse of the prevalence of each term in the whole corpus and controlling for the different lengths of the documents (Bouchet-Valat & Bastin, 2018). Each resulting cluster has a certain within-cluster variance, such that the more similar the papers the lower the internal variance.

In order to choose the number of clusters objectively, four popular clustering validity indices were used in this study:

Average silhouette width. A measure that considers how closely related objects are within the cluster and how clusters are separated from each other. The average silhouette width provides clustering validity and is widely used to select the best number of clusters (Rousseeuw, 1987). Its value ranges between 1 and -1, with a value of 1 indicating a very good cluster (Yu-Wei, 2015). For more information about this index, see Kaufman and Rousseeuw (1990).

Average between clusters. A measure of the average distance between cluster centres. It is influenced by the geometry of the cluster centres and increases with the number of clusters. The higher the value, the greater the separation between groups. A more formal and rigorous description of this statistic can be found in Halkidi, Batistakis, and Vazirgiannis (2001).

Dunn index. High values of this metric indicate compact and well-separated clusters where the means of different groups are sufficiently far apart, taking into account the internal variance of cluster. For more detailed information, see Halkidi, Batistakis, and Vazirgiannis (2002) and Dunn (1974).

Entropy Index. Measures the quantity of disorder (variation) is found in clusters. Low values of this index means better clustering. For more information, see Meilă (2007).

3. Results and discussion

The functionalities of R have made it possible to create and examine the corpus formed by the 16 papers. To that end, we built a document

term matrix of dimension $n \times t$ where n is the number of research papers (rows) and t the number of unique terms (columns). Each cell represents the absolute frequency of a term in a document.

Fig. 1 shows the output of the document term matrix. It can be seen that the corpus contains 4,743 different terms in the 16 business research papers. The non-/sparse entries are necessary to obtain the sparsity measure, which represents the frequency of the terms in the corpus. A high sparsity value means terms are not repeated often among the different documents (Dinov, 2018) which implies that the documents that make up the corpus are not very similar. The longest terms, with 17 characters, are interdisciplinary and multidisciplinary.

For the sake of brevity, Table 3 shows the 31 most frequent terms in the 16 papers. It should be noted that the name of each term indicates the root obtained in the stemming process (step 8). Terms that appear at least 135 times in the corpus were considered the most frequent; in the papers analysed, covid (571), crisi (407) and firm (400) were the most commonly mentioned terms.

Once the most frequent terms had been detected, a hierarchical cluster analysis was applied, in which the number of groups was determined objectively. Table 4 shows the result of validation statistics for the maximum number of clusters, 15. The best number of clusters will be the one with the highest value in the average silhouette width, average between and Dunn index measures, and the lowest in the entropy index. It can be seen that the best results are registered by 2 and 3 clusters. Although the results are very similar for both, we opted for 3 clusters to facilitate the interpretation of the groups.

Table 5 below shows the composition of each cluster into which similar papers have been grouped on the basis of the most frequent terms they share. The columns present the number of the cluster; the within-cluster variance; the Chi-squared distance to the cluster centroid, on the basis of which each paper is assigned to the closest cluster; document clustering; and finally, term clustering. Document clustering groups papers with the smallest Chi-squared distance to the average vocabulary of the cluster. Term clustering groups terms whose observed frequency in the document has a probability below 10% under a hypergeometric distribution, based on their overall frequencies in the corpus and on the number of occurrences of all terms in the considered cluster (Bouchet-Valat & Bastin, 2018). The terms *respons*, *provid*, *experi* and *custom* are not shown because they are not significant (p-values ≥ 0.10) for clustering.

Clusters are comprised of similar papers with common terms. Fig. 2 represents the 3 obtained clusters with the terms that make up each one on a two-dimensional plane. Each term is represented by a point, with the distance to the closest one indicating the similarity between them (Lancia, 2008). The distance between terms and the origin (0,0) measures the quality of the terms on the two-dimensional plane. The terms that lie farther from the origin will be more discriminatory and well represented on map cluster (Kassambara, 2017). The clustering of these terms highlights the most pressing concern surrounding the pandemic, revealing patterns that could guide companies in implementing measures to mitigate the consequences of the pandemic.

3.1. Cluster 1: Impact on the public health management model

The first cluster consists of a single research paper; that is, 6.25% of the papers analysed. It differs from the other documents in the sample in that it approaches COVID-19 as a scenario for studying gender equity in public systems, specifically in health organizations. The key terms that

Table 3
Most frequent terms in the corpus.

Term	Frequency	Term	Frequency	Term	Frequency
covid	571	strategi	241	develop	163
crisi	407	impact	238	financi	152
firm	400	time	221	suppli	151
consum	369	manag	211	respons	151
busi	335	health	201	custom	150
research	299	organ	196	public	150
market	279	innov	193	experi	149
pandem	273	chang	184	provid	147
social	261	effect	184	technolog	142
tourism	246	work	181		
new	245	global	179		

Table 4
Selection of the number of clusters.

Number of clusters	Average silhouette width	Average between	Dunn Index	Entropy Index
2	0.34	2.70	1.50	0.35
3	0.30	2.70	1.50	0.40
4	0.16	2.30	1.00	0.92
5	0.14	2.10	0.93	1.20
6	0.16	2.10	1.10	1.40
7	0.14	2.00	0.95	1.70
8	0.12	2.00	0.91	1.90
9	0.12	2.00	0.93	2.00
10	0.10	2.00	1.10	2.10
11	0.10	2.00	1.20	2.20
12	0.10	1.90	1.00	2.40
13	0.09	1.90	1.30	2.50
14	0.05	1.90	1.10	2.60
15	0.04	1.90	1.10	2.70

make up this cluster are *health*, *public*, *develop*, *effect* and *impact*. The terms *health* and *public* have a high frequency in this cluster relative to the total for the corpus. This indicates that they are the most discriminant terms of the cluster; as such, they lie far from the origin in Fig. 2. *Develop*, *impact* and *effect* have a lower frequency, but are closer to one other, reflecting the high similarity between them. It should be noted that the term *gender* does not appear in this cluster because its frequency relative to the total for the corpus is very low.

Leung et al. (2020) show that gender equity and development have a positive impact on women’s representation in the legislature and are consequently positively correlated with public health spending. Furthermore, this expenditure significantly influences the volume of diagnosed cases and the treatment of critical cases, with the urban population having a positive influence on the number of tests and COVID-related deaths. All of this underscores the importance of continuing the struggle for gender equality and of women holding decision-making positions. In extreme situations such as those experienced during the pandemic, there is a need to combine the characteristic features of both genders, toning down the riskier positions more typical of the male gender (Ertac & Gurdal, 2012).

This pandemic has meant the quality of public health systems being called into question in all affected countries. The emerging new scenario is encouraging a shift from public policymakers towards fostering women’s access to leadership positions, which has been shown to be a positive factor in ensuring the quality of public services provided (WHO,

```
<<DocumentTermMatrix (documents: 16, terms: 4743)>>
Non-/sparse entries: 15119/60769
Sparsity           : 80%
Maximal term length: 17
```

Fig. 1. Summary of document term matrix.

Table 5
Summary of hierarchical cluster analysis of 16 research papers.

Cluster	Within-cluster variance	Chi ² distance	Document clustering	Terms clustering
1	0.00	0.00	Leung et al. (2020)	public, health, develop, effect, impact
2	0.91	0.18	Kirk and Rifkin (2020)	consum, technolog, time, pandem, social, new, chang, research
2		0.21	Sheth (2020)	
2		0.33	Pantano et al. (2020)	
3	3.1	0.31	Donthu and Gustafsson (2020)	firm, tourism, financi, strategi, organ, busi, crisi, market, innov, suppli, manag, global, work, covid
3		0.89	Sharma et al. (2020b)	
3		0.92	Woodside (2020)	
3		0.93	Krishnamurthy (2020)	
3		0.98	He and Harris (2020)	
3		1.08	Wang et al. (2020)	
3		1.56	Sigala (2020)	
3		1.91	Eggers (2020)	
3		2.01	Bacq et al. (2020)	
3		3.20	Mullins (2020)	
3		3.31	Carnevale and Hatak (2020)	
3		3.92	Sharma et al. (2020a)	

2017; Mayer & Oosthuizen, 2020). Efforts must be made to ensure the excellence of the management of health systems, with clear policies and priorities set to benefit the population. In addition, it is necessary to ensure correspondence between the effects of the coronavirus pandemic and national development, where access to public health should be a universal right.

3.2. Cluster 2: Changes in society's consumption habits

This cluster includes 18.7% of the articles analysed, and the key theme centres around *consumption, technology, pandemic, time, social, new, change* and *research*. The terms *consumption* and *technology* have a high frequency in the cluster relative to the total for the corpus, and it is these terms are the most discriminating in the cluster. There is a strong connection between the terms *research, change, pandemic, social, new* and *time* which highlights the usefulness of focusing studies on the social changes arising from the new situation that emerged during the pandemic (Fig. 2).

The research carried out and grouped in this cluster has placed great importance on the rapid spread of the disease and the new societal habits established in response. In just under five months this infectious disease has affected all the countries of the world, forcing all the agents in the economy to adapt. Nevertheless, countries' reactions to the uncontrolled spread of the virus have not been homogeneous. The vast majority have opted for public lockdowns, imposing strong restrictions on the national and international mobility of people and products. Social distancing rules are forcing companies to adapt their distribution channels in order to reach consumers, who have had to adapt very quickly to this new scenario.

All this has led to a reorientation of research, with analyses focused on trying to explain the unusual needs emerging in the markets. The "new normal" requires changes that are revolutionizing previous customs and life as we knew it before the pandemic (Kirk & Rifkin, 2020). The use of emerging technologies is now an imperative; society is incorporating them into daily life at a dizzying pace, prompting new patterns of behaviour, which in turn are having a major impact on consumption habits. This calls for a realignment of business strategies to adapt to this new globalized market. It is thus essential to establish good management practices to bring supply in line with consumer demand (Pantano et al., 2020).

3.3. Cluster 3: Economic effects of covid-19 on business organizations

Of all the papers that make up the corpus, 75.05% analyse the

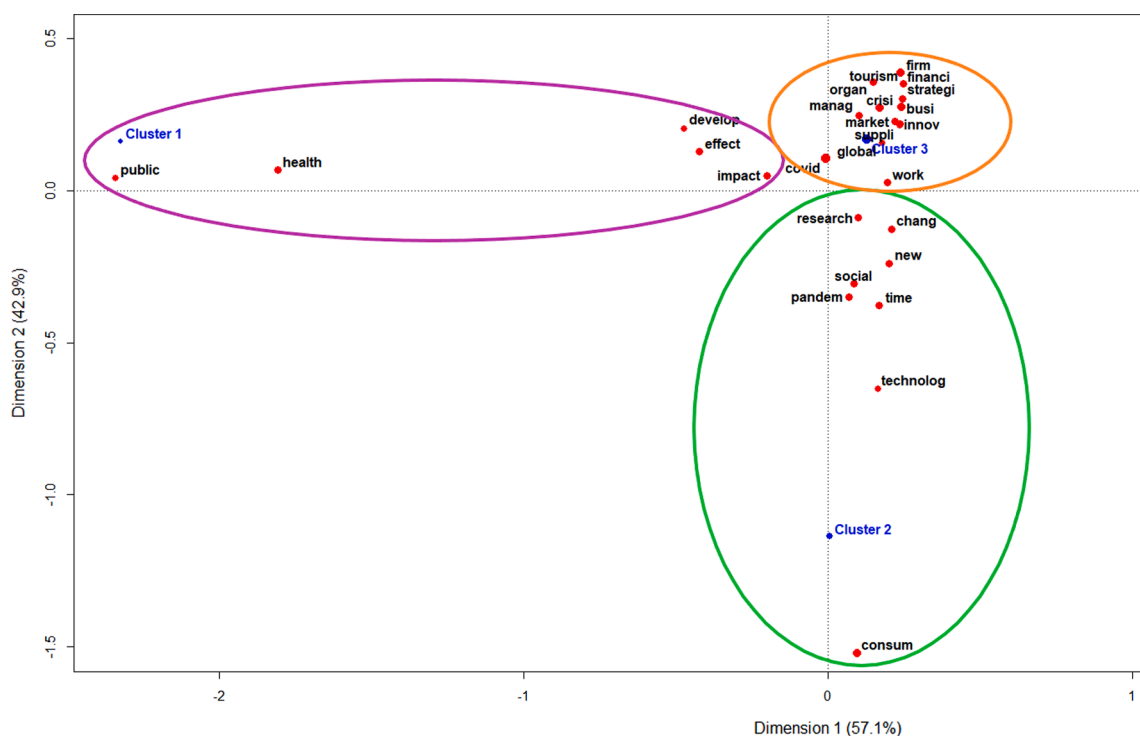


Fig. 2. Hierarchical clustering of most frequent terms in the corpus.

economic effects of coronavirus from different perspectives, but the common thread running through them is the economic and social implications for production networks as a result of the measures taken to curb the contagion. In this respect, the most commonly used terms were *firm*, *tourism*, *business*, *crisis*, *strategies*, *financial*, *organization*, *market*, *innovation*, *supply*, *management*, *global*, *covid* and *work*, with *firm*, *tourism* and *financial* being the most significant. The proximity of all these terms indicates the strong correlation between them (Fig. 2).

Globalization and the interconnectedness of countries has been the main trigger for the worldwide spread of COVID-19. The closure of national borders and the restricted mobility of the population even within their own territory has resulted in economic disruption comparable only to the Great Recession. The resulting global crisis is expected to be unprecedented; the paralysis of the entire productive sector has led to a supply and demand shock, with the consequent collapse of international financial markets. All economic sectors have been affected, with the impact being especially intense in countries that are heavily dependent on tourism.

Faced with this new situation, there are only two real alternatives: to try to get back to the normality we knew before the pandemic, in order to avoid the social impact of high levels of unemployment; or to consider this new scenario as a turning point, marking the start of new ways of working. International bodies and authorities responsible for setting national policy should focus their efforts on helping companies to introduce technological and social innovation to enable them to recover and remain competitive. Without exception, the pandemic has rendered countries' productive systems obsolete. New, more interconnected forms of management are needed, but with less dependence on external agents, which is where virtual markets are irreversibly gaining ground.

4. Conclusions

Affecting all countries around the world, COVID-19 has caused the most severe pandemic of recent times, the economic consequences of which have not taken long to emerge. Business survival will depend on the ability to adapt to the new market, which has been shaped by the restrictions on movement imposed by almost all nations to curb the spread of the disease. Small and medium-sized businesses have been the hardest hit by the pandemic, so they will have to reinvent themselves quickly if they want to survive in this unforeseen scenario. In this new reality, the future of business is uncertain, so making the right decisions is crucial to strengthening companies.

There has been a rush of new research papers on the disease and its economic and social effects, so it is necessary to identify some kind of similarities among them in order to be able to draw conclusions that can be used to orient guidelines for business. This paper proposes the use of text mining methodology to detect lines of research centred around the coronavirus, and the results can be used to mark out policies to be implemented. We analysed a set of papers representative of the area of business and marketing, published in journals indexed in the top quartile of the prestigious database, WoS.

In light of the results obtained, three different areas of intervention can be identified. The common thread that runs through all of them is the need to introduce new forms of action to improve citizens' quality of life. The "nascent" market resulting from the lockdown situation requires profound changes in the management systems of the entire productive network, with public health playing a key role. The standard that existed until now has become obsolete. There is thus a need to adapt to the novel features of consumer habits, where virtual business systems have gained ground they are unlikely to cede. Remote working has become universal due to the pandemic, with many companies adopting it on a long-term or even permanent basis. The survival of businesses will depend on their ability to treat this situation as a turning point; detailed market analysis and the introduction of innovation at all levels will strengthen the foundations of productive activity. Authorities must facilitate and assist in this arduous task—one for which not everyone is

prepared—without neglecting issues such as gender equality and human rights. This is an opportunity for renewal that must be seized in order to promote the sustainability of organizations, regardless of their profit-making purpose. As is well known, the pandemic has caused the largest drop in CO2 emissions in history. Therefore, the time is right to introduce management models that address climate change, which is becoming an international urgency at all levels.

The whole world is experiencing an unprecedented situation that requires additional analysis and efforts in all economic sectors. The researches of the last few months is aimed at providing solutions to problems that have not ceased to arise, and which will continue until the pandemic is brought under control so as to avoid affecting the economy. It is essential to implement a joint effort between researchers and those responsible for economic and business guidelines with the aim of combining contributions that will allow them to move into the professional field to optimise the solutions found.

The main limitation of this research is the volume of the corpus. Specifically, 16 works published at the time of the empirical analysis have been analyzed, which met the criteria defined in the Section 2. Therefore, the natural extension of this paper is to expand the database in order to be able to establish more specific measures, tailored to each scenario and country. Once the new database has been processed using the programming language provided in R, the methodology presented in this article could easily be replicated.

References

- Aggarwal, C. C., & Zhai, C. (2012). An introduction to text mining. In C. C. Aggarwal, & C. Zhai (Eds.), *Mining text data* (pp. 1–10). New York: Springer.
- Acharya, S. (2018). *Data Analytics using R*. India: McGraw-Hill Education.
- Bacq, S., Geoghegan, W., Josefy, M., Stevenson, R., & Williams, T.A. (2020). The COVID-19 Virtual Idea Blitz: Marshaling social entrepreneurship to rapidly respond to urgent grand challenges. *Business Horizons*. Available online 12 May. <https://doi.org/10.1016/j.bushor.2020.05.002>.
- Bouchet-Valat, M., & Bastin, G. (2018). RcmdrPlugin.temis: Graphical Integrated Text Mining Solution. R package version 0.7.10. <https://rdrr.io/cran/RcmdrPlugin.temis/>.
- Carnevale, J., & Hatak, I. (2020). Employee adjustment and well-being in the era of COVID-19: Implications for human resource management. *Journal of Business Research*, 116, 183–187. <https://doi.org/10.1016/j.jbusres.2020.05.037>.
- Cheng, X., Cao, Q., & Liao, S. S. (2020). An overview of literature on COVID-19, MERS and SARS: Using text mining and latent Dirichlet allocation. *Journal of Information Science*, 0165551520954674. <https://doi.org/10.1177/0165551520954674>.
- Dinov, I. D. (2018). *Data science and predictive analytics: Biomedical and health applications using R*. Switzerland: Springer.
- Donthu, N., & Gustafsson, A. (2020). Effects of COVID-19 on business and research. *Journal of Business Research*, 177, 284–289. <https://doi.org/10.1016/j.jbusres.2020.06.008>.
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1), 95–104. <https://doi.org/10.1080/01969727408546059>.
- Eggers, F. (2020). Masters of disasters? Challenges and opportunities for SMEs in times of crisis. *Journal of Business Research*, 116, 199–208. <https://doi.org/10.1016/j.jbusres.2020.05.025>.
- Ertac, S., & Gurdal, M. Y. (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization*, 83(1), 24–30. <https://doi.org/10.1016/j.jebo.2011.06.009>.
- Gordon, A. V., Ramic, M., Rohrbeck, R., & Spaniol, M. J. (2020). 50 Years of corporate and organizational foresight: Looking back and going forward. *Technological Forecasting and Social Change*, 154, Article 119966. <https://doi.org/10.1016/j.techfore.2020.119966>.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2–3), 107–145. <https://doi.org/10.1023/A:1012801612483>.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2002). Clustering validity checking methods: Part II. *ACM Sigmod Record*, 31(3), 19–27. <https://doi.org/10.1145/601858.601862>.
- Han, T., Zhang, L., Pirbhulal, S., Wu, W., & de Albuquerque, V. H. C. (2019). A novel cluster head selection technique for Edge-computing bases IoMT systems. *Computer Networks*, 158, 114–122. <https://doi.org/10.1016/j.comnet.2019.04.021>.
- He, H., & Harris, L. (2020). The impact of Covid-19 pandemic on corporate social responsibility and marketing philosophy. *Journal of Business Research*, 116, 176–182. <https://doi.org/10.1016/j.jbusres.2020.05.030>.
- Husson, F., Lê, S., & Pagès, J. (2017). *Exploratory multivariate analysis by example using R*. New York: CRC Press, Taylor & Francis Group.
- Kassambara, A. (2017). Practical guide to principal component methods in R: PCA, (M) CA, FAMD, MFA, HCPC, factextra, (Vol. 2). STHDA.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New Jersey: Wiley-Interscience.

- Kirk, C., & Rifkin, L. (2020). I'll trade you diamonds for toilet paper: Consumer reacting, coping and adapting behaviors in COVID-19 pandemic. *Journal of Business Research*, 117, 124–131. <https://doi.org/10.1016/j.jbusres.2020.05.028>.
- Krishnamurthy, S. (2020). The future of business education: A commentary in the shadow of the Covid-19 pandemic. *Journal of Business Research*, 117, 1–5. <https://doi.org/10.1016/j.jbusres.2020.05.034>.
- Lancia, F. (2008). Word Co-occurrence and Similarity in Meaning. In S. Salvatore, & J. Valsiner (Eds.), *Mind as Infinite Dimensionality*. Rome: Edizioni Carlo Amore.
- Leung, T. Y., Sharma, P., Adithipyangkul, P., & Hosie, P. (2020). Gender equity and public health outcomes: The COVID-19 experience. *Journal of Business Research*, 116, 193–198. <https://doi.org/10.1016/j.jbusres.2020.05.031>.
- Lopez, C.E., Vasu, M., & Gallemore, C. (2020). Understanding the perception of COVID-19 policies by mining a multilanguage twitter dataset. (accessed November 2nd, 2020) arXiv:2003.10359v1 [cs.SI].
- Mayer, C. H., & Oosthuizen, R. M. (2020). Concepts of creative leadership of women leaders in 21st century. *Creativity Studies*, 13(1), 21–40. <https://doi.org/10.3846/cs.2020.10267>.
- Meilä, M. (2007). Comparing clusterings—an information based distance. *Journal of multivariate analysis*, 98(5), 873–895.
- Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54. <http://www.jstatsoft.org/v25/i05>.
- Mullins, J. (2020). Are your cash-flow tools recession ready? Business Horizons. Available online 23 April 2020. <https://doi.org/10.1016/j.bushor.2020.04.003>.
- Nepal, S. (2020). Travel and tourism after COVID-19 – business as usual or opportunity to reset? *Tourism Geographies*, 22(3), 646–650. <https://doi.org/10.1080/14616688.2020.1760926>.
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., ... Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193.
- Nguyen, E. (2013). Text Mining and Network Analysis of Digital Libraries in R. In Y. Zhao, & Y. Cen (Eds.), *Data Mining Applications with R* (pp. 95–115). Amsterdam: Elsevier Inc.
- Pantano, E., Pizzi, G., Scarpi, D., & Dennis, C. (2020). Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research*, 116, 209–2013. <https://doi.org/10.1016/j.jbusres.2020.05.036>.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. Reuters. T. (2019). ISI Web of Knowledge Web site (accessed February 6, 2019). <http://www.isiwebofknowledge.com>.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Reuters, T. (2019). ISI Web of Knowledge Web site (accessed June 22, 2020). <http://www.isiwebofknowledge.com>.
- Sharma, P., Leung, T. Y., Kingshott, R. P. J., Davcik, N. S., & Cardinali, S. (2020). Managing uncertainty during a global pandemic: An international business perspective. *Journal of Business Research*, 116, 188–192. <https://doi.org/10.1016/j.jbusres.2020.05.026>.
- Sharma, A., Adhikary, A., & Borah, A. B. (2020). Covid-19's impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data. *Journal of Business Research*, 117, 443–449. <https://doi.org/10.1016/j.jbusres.2020.05.035>.
- Sheth, J. (2020). Impact of Covid-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research*, 117, 280–283. <https://doi.org/10.1016/j.jbusres.2020.05.059>.
- Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal Business Research*, 117, 312–321. <https://doi.org/10.1016/j.jbusres.2020.06.015>.
- Sjodin, D., Parida, V., Kohtamaki, M., & Wincent, J. (2020). An agile co-creation process for digital servitization: A micro-service innovation approach. *Journal of Business Research*, 112, 478–491. <https://doi.org/10.1016/j.jbusres.2020.01.009>.
- Täuscher, K., & Laudien, S. M. (2018). Understanding platform business models: A mixed methods study of marketplaces. *European Management Journal*, 36(3), 319–329. <https://doi.org/10.1016/j.emj.2017.06.005>.
- Ulm, J. W., & Nelson, S. F. (2020). COVID-19 drug repurposing: Summary statistics on current clinical trials and promising untested candidates. *Transboundary and Emerging Diseases*. <https://doi.org/10.1111/tbed.13710>.
- Wang, Y., Hong, A., Li, X., & Gao, J. (2020). Marketing innovations during a global crisis: A study of China firm's response to COVID-19. *Journal of Business Research*, 116, 214–220. <https://doi.org/10.1016/j.jbusres.2020.05.029>.
- Wen, H., Wei, Y., & Wang, S. (2020). Survey on Impact of COVID-19 No. 2: Report on the impact of COVID-19 crisis on the business management. <https://www.sohu.com/a/385568115.476872>.
- WHO. (2017). *Statement by Dr Zsuzsanna Jakab, WHO Regional Director for Europe, for International Women's Day*.
- William, D., & Chang, B. (2019). Analysis of energy consumption of selected cluster technique. *Engineering Science & Technology Journal*, 1(2), 37–44.
- Woodside, A. (2020). Interventions as experiments: Connecting the dots in forecasting and overcoming pandemics, global warming, corruption, civil rights violations, misogyny, income inequality, and guns. *Journal of Business Research*, 117, 212–218. <https://doi.org/10.1016/j.jbusres.2020.05.027>.
- Yu-Wei, C. D. C. (2015). *Machine learning with R cookbook*. Birmingham: Packt Publishing Ltd.

Patricia Carracedo is graduated in Actuarial and Financial Sciences (2007) and Business Studies (2004) at the Universitat de València and has a Master's Degree in data analysis engineering (2013) at the Universitat Politècnica de València (UPV). PhD in Statistics and Operational Research at UPV. She is Director of two degrees in Business Administration and Management and Economics at the Valencian International University and since 2014, lecturer of Statistics and Econometrics in both degrees. She was awarded at the congress "XXV ASEPUMA Conference - XIII International Meeting 2017" and obtained a research project of the Fundación Mapfre as main researcher. Her research has been developed using the statistical package R-studio, getting to have a good command of it.

Rosa Puertas. PhD in Economics and Business. Professor at the Universitat Politècnica de València, Department of Economics and Social Sciences. She has carried out research related to various fields: International Trade, Efficiency, Input-Output and Logistics. Currently, she has focused on issues related to innovation, sustainability and construction of synthetic indexes, which are leading to publications in internationally prestigious journals. Member of Group of International Economics and Development.

Luisa Marti is PhD in Economics and Business Studies from the University of Valencia (2001). For the last 20 years I have been a professor at the Polytechnic University of Valencia, where I have taught subjects related to the field of applied economics. Since 2019 I have been a university lecturer. My main areas of research are trade, economic impact analysis, quantitative data analysis in companies, and efficiency analysis. Having a six-year research and numerous publications in prestigious journals of impact Journal Citation Report.