
DOMAIN ANALYSIS OF THE LITERATURE ON COVID-19 IN INFORMATION SCIENCE'S BRAZILIAN CONTEXT

Isadora Victorino Evangelista (1),

Marcos Oliveira da Costa (2),

Samuel Santos da Rosa (3),

Thiago Henrique Bragato Barros (4),

Rita do Carmo Ferreira Laipelt (5),

Luciana Monteiro-Krebs (6)

(1) Universidade Federal de São Carlos, Brazil, isadora.evangelista@ufscar.br

(2) Universidade Federal do Pará, Brazil, marcos.costaoliver@gmail.com

(3) Universidade Federal do Rio Grande do Sul, Brazil, samuel.sdrosa@gmail.com

(4) Universidade Federal do Rio Grande do Sul, Brazil, bragato.barros@ufrgs.br

(5) Universidade Federal do Rio Grande do Sul, Brazil, rlaipelt@gmail.com

(5) Universidade Federal do Rio Grande do Sul, Brazil, luciana.monteirokrebs@gmail.com



Abstract

In the past two years, science has been mobilized around studies centered on the new pandemic phenomenon known as COVID-19, a variation of the coronavirus with higher transmission capacity and higher mortality. Thus, as a way of investigating the state of the art of this topic in the field of Information Science, this research sought to answer the following question: what is the Brazilian scientific production in Information Science regarding the COVID-19 pandemic? As main justification, we understand that systematically identifying and disseminating those studies may help the construction of an appropriate theoretical analysis on the theme in the national scope. To support our research theoretically and methodologically, we reserved a section to explore Domain Analysis (DA). Proposed by Hjørland and Albrechtsen (1995) around Information Science, this approach allows us to analyze domains of knowledge from their social interactions as discursive communities. Then, we describe the context of information in

Evangelista, Isadora Victorino, et al. Domain analysis of the literature on COVID-19 in Information Science's national context. *Brazilian Journal of Information Science: Research trends*, vol. 16, Dossiê Análise de Domínio, 2022, e02137. DOI: 10.36311/1981-1640.2022.v16.e02137

health, specifically the COVID-19 pandemic context. This is a quali-quantitative research: we made a textual analysis of the keywords found in the articles that make up the corpus and applied bibliometric criteria to identify occurrence patterns of the most productive authors and journals. Being a descriptive-exploratory research about COVID-19 in Brazil Information Science literature, we aimed to describe how the pandemic phenomenon influenced textual production in this context, seeking to better understand the theme and looking for reflections about the subject. It was possible to infer that the scientific dissemination of knowledge around this theme is closely related to research in information in health, and it even includes a new term in this perspective: “infodemia” (infodemic), a term that is related to the vastness of information made available via the web, especially on social media. In addition to that, the incidence frequency of terms such as comunidade (community) and competência em informação (information competence) enhances the significance of the area of IS in studies oriented toward knowledge organization and representation in face of the overload of information.

Keywords: Domain Analysis; Covid-19; Information in Health.

1 Introduction

In the past two years, science has been mobilized around studies centered on the new pandemic phenomenon known as COVID-19, a variation of the coronavirus with higher transmission capacity and higher morbidity. This mobilization directly affected areas such as Information Science (IS), which comprises the study of informational phenomena and the society’s behavior when faced with those changes.

Thus, as a way of investigating by an exploratory analysis about this topic in the field of Information Science, this research sought to answer the following question: what is the Brazilian scientific production in Information Science regarding the COVID-19 pandemic? As main justification, we understand that systematically identifying and disseminating those studies may help to be aware what is the main issues researched, the phenomenons analyzed in this scope and to what extent the Brazilian literature in Information Science is dedicating time to this theme. Specifically, we sought to identify which journals are the most receptive to this theme, who are the authors who have been writing about the theme, and which keywords are related to it.

This paper is organized in sections. Initially, we expose the main theoretical premises used in the research: we discuss domain analysis, the methodology that supports this research, in addition to dedicating a subsection about information in health, especially with regard to COVID-19. Subsequently, the methods and research tools are described – the database used, the number of papers that make up our *corpus*, the exclusion criteria, as well as all other components of the

methodological framework. Finally, we present and discuss the results in light of the literature. We close the paper with our final considerations.

2 Domain Analysis in the Context of COVID-19

To support our research theoretically and methodologically, we reserved a section to explore Domain Analysis. Proposed by Hjørland and Albrechtsen (1995) in the area of Information Science, this approach allows us to analyze domains of knowledge from their social interactions as discursive communities. As a theoretical choice, we seem to focus on Hjørland research about Domain Analysis, considering his important and well-known works about this methodology. Besides that, it is crucial to cite for future research the following works: Tennis (2003; 2012), Smiraglia (2012; 2015a; 2015b), Castanha and Gracio (2014), Oliveira and Grácio (2014), López-Huertas (2015), Guimarães and Tognoli (2015) and Barros (2021). Then, we describe the context of information in health, specifically the COVID-19 pandemic context.

2.1 Domain analysis

Domain Analysis as proposed by Information Science is originated from a publication by Hjørland and Albrechtsen (1995), in which the authors sought not to define its concept, but present it as a possible methodological approach for the development of works in the scope of IS.

To the authors, the best way to comprehend information's concepts and relations in IS is through the study of domains of knowledge, also understood by the researchers as discursive communities, which are part of the social perspective of professions. Thus, knowledge and individual informational needs are analyzed based on the domain's behavior (Hjørland and Albrechtsen 1995).

In that sense, domain analysis paradigm is primarily considered a social paradigm, enabling an understanding of aspects related to psychology, sociolinguistics, and the sociology of knowledge and of science in the scope of a social science such as IS. In addition, DA has a functionalist paradigm; it seeks to understand the implicit and explicit functions of information and communication in order to better visualize the informational behavior of individuals. Finally,

DA presents a philosophical paradigm, investigating the bases of those behaviors and the individuals' perceptions of the domain (Hjørland and Albrecthsen 1995).

At the time, the authors presented the concept as something not necessarily new but as something that can be understood as a new way to analyze both the margins and the core of studies in Information Science. To Hjørland and Albrecthsen (1995), Domain Analysis can be seen as an approach that can contribute to contemporary studies in a way that, so far, had not been observed in other theories.

Hjørland (2011) seeks to present in a practical way eleven approaches in which the use of domain analysis proves to be methodologically productive in investigations in the scope of Information Science. The main motivation for the author's initiative is the fact that, even though information professionals work directly with different specialties, they do not necessarily become specialists in these themes, even if they must act as such for the proper functioning of the professional activities developed. A way to meet this demand stems from domain analysis' approaches, in which it is possible to analyze the whole through the contribution of its parts. It is also important to state that to Hjørland it is primordial to have a combination of two or more of these approaches to proceed a complete analysis.

The eleven approaches presented by Hjørland (2011) are: the production of literature guides and thematic portals; construction of classifications and thesauri; indexing and retrieving information; user studies; bibliometric studies; historical studies; studies on documentary genres; critical and epistemological studies; terminological studies; languages for specific purposes, semantics in databases, and discursive communities; studies on scientific structures and institutions; and, finally, studies related to artificial intelligence and scientific cognition. In this research, we focused on the bibliometric approach and on terminological and semantic studies, as well as on studies on discursive communities.

Regarding bibliometric aspects, Hjørland (2011) emphasizes that this area of study is used as a science evaluation method and can be employed as a domain analysis tool or method in a number of ways. Bibliometric studies have proven to be an important approach in DA, for they enable the representation of real connections between individual documents. One of the ways is through bibliometric maps or visualizations of scientific areas through co-citation analysis, in

which connections of dependence and recognition between articles, authors, institutions, etc., are made explicit.

Some factors influence the way in which bibliometric studies are effectively performed, such as which database is chosen and the journals it indexes. That is the information that will provide the research data, and it is important that we have in mind that these tools have limitations and that they cannot be considered unbiased tools, since they are not likely to be homogeneous in the types of documents and themes they work with. When a given journal selects a discipline or domain, it does so at the expense of other disciplines, and that reveals some inferences about the area. That perspective feeds into a cycle: the theoretical perception of a domain encompasses what is selected by journals of the area, which, in turn, determines which authors will be the most influential in that area (Hjørland 2011).

Another factor that influences bibliometric analyses is the domain's citation behavior pattern. The reason why a researcher cites another may be negative – a criticism to their previous work, a disagreement with the ideas that author supports. And there is the occurrence of what Hjørland (2011) calls “the uncited”, which happens when an author makes a reference to a theory or specific tool without referencing the researcher responsible for creating it.

There is also the fact that many theories and methods are defined based on their convenience or time consumption, which can influence the non-representation of other theories and methodologies whose quality is known, but which are considered more demanding by the researchers. This “choosing” of theories also happens because at certain times, certain authors end up being more popular than others due to their influence in different areas of knowledge – this popularization may occur due to a change of paradigms, as proposed, for example, by Thomas Kuhn (Hjørland 2011).

In that sense, bibliometric approaches prove to be an important route to domain analysis – given that its empirical data are based on the analysis of relationships and connections between different individual documents – as long as its limitations and biases are considered, so that the analysis can be effective (Hjørland 2011).

When it comes to the ninth approach – terminological studies, studies of languages for specific purposes, semantic studies, and studies of discursive communities – Hjørland (2011) states that information professionals have always had a close relationship with linguistic problems, whether in the construction of thesauri, in the effective retrieval of information through the translation from natural language to controlled language, among others. Based on that, linguistics can guide information professionals through the terms used in a certain specialty, and an analysis of those terms can be considered a domain analysis.

This construction of vocabularies proves to be problematic because, as shown in the bibliometric approach, the preference for certain terms occurs in detriment of others, and that standardization will favor or marginalize certain points of view. To Hjørland (2011), each area has specific terminological problems – the ones faced by the representation of chemical compounds are not the same as those faced by the different schools of thought in psychology – however, what they have in common is their lack of neutrality and objectivity in this process.

According to Hjørland (2011), languages for specific purposes such as controlled languages are determined, on the one hand, by different forms of communication between different groups, and, on the other, by a principle of economy that avoids the use of redundant information. Considering that users make use of terminologies to retrieve information, and that, according to Hjørland, there is a strong correlation between linguistic knowledge and substantial knowledge – the more specialized an individual is, the bigger their lexicon –, studies with this approach can help us understand better the problems that arise in the search for and retrieval of information.

In the context of Information Science, Hjørland (2011) establishes four basic premises for studies with this approach: signs and their meaning are formed from their social application in discursive communities – these groups share terms that have a semantic context that is different from the one used by the rest of society; different communities make up different types of documents, with different structures; these discursive or epistemic communities are also influenced by norms and trends, which will influence the construction of syntactic and semantic contexts as well; when documents are marginalized in databases, not being effectively represented and affecting the informational retrieval, contexts are lost.

In summary, Hjørland (2011) affirms that studies on terminologies must be object of study in Information Science because they affect how documents are represented in databases and how the searches are performed. For these studies to take place, there needs to be a functional base, which can be found in DA, especially when combined with other approaches such as bibliometric studies.

A few years later, Hjørland (2017) sought to better understand the concept of domain analysis and its approaches in Information Science. To the author, these approaches show that the objects of study in IS are socially and theoretically defined, surpassing the cognitive perspective accepted up until then. These eleven proposed perspectives were formed by processes and theories present in the activities of information professionals, especially in relation to knowledge organization (KO), combining sociological and epistemological conceptions.

To the author, domain analysis centers its studies in the knowledge about a subject, which is also a premise intrinsic to studies related to knowledge organization. According to Hjørland (2017), knowledge about subjects has been institutionalized in information units such as libraries, given that different researchers with a high level of expertise are used to develop representation tools and processes, such as classification languages or indexing in databases. From this perspective – information professionals’ multiple facets – Hjørland attributes to Information Science the character of meta-science.

According to Hjørland (2017), domain analysis moves away from the understanding of an “information professional who is complete” in all existing specialty lanes, in which it is preferable that there be professionals with general knowledge instead of professionals with a high level of specialization in a given domain. To the author, DA contributes to diversity, as it influences the increase in the quality of the information service by assigning professionals who represent different specialties in the informational unit.

Thus, Domain Analysis values the specialization of professionals in specific areas instead of a shallow understanding of the whole. By aiming to understand how a certain domain works, according to the necessary focus for the representation of information organization, DA can help in the effective retrieval of information and in the adequate development of information services (Hjørland 2017).

In short, the author asserts that information and knowledge organization activities are acts of mediation between the user and the documents, and those acts take place in view of a need for information through documents produced by specific individuals about specific knowledge in a specific area of knowledge. For this mediation to occur, a reasonable level of knowledge about that subject is necessary, a level that will depend on the specialty of the information unit. In that sense, although information professionals are not trained in that specific subject, they will have to acquire informational skills on the domain, which characterizes IS as a meta-science. Domain analysis is the methodology in IS which considers the optimization of information systems and services from the perspective of their specialty contents (Hjørland 2017).

Seeking to define how Domain Analysis understands *domain*, Hjørland (2017) proposes that the concept be understood as a discipline, but that it does not have to be seen that way – it can encompass a number of different disciplines or even be a hobby – as long as there are subjects in common. Thus, domains can be understood as the specialization of the division of cognitive labor, which is theoretically defined and socially institutionalized. In this perspective, domains are not static units in time and space, on the contrary, they are dynamic, they change and depend on the context in which they are inserted, and on the theories proposed in the domain.

Another matter that deserves attention is that, at the same time that a domain exists, it is also built by its individuals. Even though the subjects and documents to be analyzed are already available and already constitute a domain, it is necessary that studies be performed for the cycle of construction of the science to be realized, ensuring the growth of the area. This analysis cannot escape its subjective characteristic: a domain is analyzed based on the researcher's previous knowledge about that subject, and their perceptions may change during the analysis. In addition, there is a dual nature between domains: on the one hand, there is an intellectual perspective, and, on the other, their sociological approach, which can be seen as aspects of documentary content on the one hand, and their institutional characteristics on the other (Hjørland 2017).

In short, to Hjørland (2017) domains are a body of knowledge socially and theoretically defined and researched by a group of individuals who share epistemological and ontological commitments. Although the individuals who belong to the domain have characteristics in common, such as the object of study, their level of knowledge, and terms specific to their area of expertise,

it is not required that their studies reach a consensus, even if a certain level of stability is expected in the moment of analysis. It is also important that we have in mind that different disciplines and areas of knowledge constitute their domains based on different theories and social interests.

In Information Science, Hjørland (2017) seeks to define the concept of domain analysis under two perspectives: more generally, in studies performed through methodologies such as bibliometrics and facet analysis – or any of the eleven approaches used individually – in disciplines or other domains. And more specifically, domain analysis when performed based on different theories, paradigms, or traditions common to that domain. In this perspective, the domain is not “given” to the researcher, it is studied from different perspectives, objectives, values, and interests that constitute the analyzed domain.

Based on that, this study will perform a domain analysis of the literature about the COVID-19 pandemic produced in Information Science. Thus, the analyzed domain is the literature that composes the *corpus*. We will use two approaches proposed by Hjørland (2011) to analyze these documents methodologically – the bibliometric approach, and the studies of terms specific to the specialty, of semantic context, and of languages for specialized subjects. This analysis will encompass both a broad perspective – based on the retrieved documents and application of bibliometric aspects – and a more specific perspective – seeking to identify perspectives, objectives, and values in the mining analysis of texts carried out subsequently.

The use of domain analysis in information in health is not so explored in Information Science - that is one of the innovative characteristics of this research. In a quick research with the keywords "domain analysis" and "health" in PubMed, no one of the articles retrieved cites Domain Analysis as we understood in Information Science and the most of them follow a different perspective of the methodology - the one developed in systems engineering and that has as purpose the information that is retrieved and collected to be reusable to create new systems. As examples, we have Okada et al. (2001) and Atkison and Abu El Haj (1996).

2.2 Information in health: COVID-19's semantic field

The language is constructed through the collective interaction of individuals (natural language, both written and spoken), being full of variations that depend on the historical, political,

and social context (Orlandi 2007; Cipriano 2016). The development of specialty languages, in turn, occurs as a natural consequence of the specialization of human activities, of the use of language in an environment of specialists who use the same register but with interlocutors of different hierarchies and degrees of specialization in different levels of formality, such as is the case of the language used in health, given its high terminological density (Krieger and Santiago 2014).

Currently, with the COVID-19 pandemic, themes related to health have attracted a greater interest from society and have started to systematically integrate the contents of written and spoken media. At the same time, we observe a mobilization, as highlighted Zeng et al. (2020), around the theme COVID-19 from international and national health organizations, as well as of governmental agencies in regard to knowledge organization systems to establish an official name for this new disease. All of that happened with the purpose of granting access to information in a way that is organized, standardized, updated, and shared with everyone, so that there can be exchange and communication for the advancement of the fight against this disease (Zeng et al. 2020).

This new context is enhanced by the fact that news from around the world is being constantly updated, revised, and shared, providing users with an overload of new information, in addition to significantly expanding the circulation of new terms related to the area of health (Domingues 2021). This information overload can be translated, according to Fiorillo and Gorwood (2020), as “infodemic”, a term that denotes a significant increase in the volume of information, be that information true or false. Another challenge related to terminology in health, which Zeng (2020 p. 149) highlight is the “semantic conflict” that can occur, according to the authors:

[...] within any data and information communication process. A relevant example is the naming of a new disease, including the reuse of previous names that may share some similarities, the adoption of a known accepted name which may carry different meanings at different times of history, and the inclusion of names of particular groups of people, places, or animals based on the cases reported earliest.

Thus, we can see the importance of the area of Terminology – responsible for the shaping of domains – which is part of the studies on knowledge organization and representation for the construction, for example, of controlled vocabularies and other specialized technical-scientific organization systems. This is an essential field of knowledge given the emergence of new terms

and concepts, constantly updated, reviewed, and shared so that scientific communication can become more efficient and consistent in its search for answers, as well as in how it informs society, regardless of geographic, cultural, political, and linguistic borders, to ensure its right to information and knowledge (Krieger and Santiago 2014).

Among the several new terms and concepts, we have as the center of the pandemic the term that refers to the disease caused by the coronavirus and that, before being given its official name, “COVID-19”, underwent changes both in the scientific community and in society as a whole. Initially, the virus had several names linked to the place where the first cases occurred and to its country of origin. Among these names, Zeng et al. (2020 p. 149) cite the “Chinese virus”, “Wuhan Coronavirus”, “Wuhan SARS” in which a search, with such terms, carried out on Google Scholar in April 2020 by the authors, returned more than 1.280 (one thousand two hundred and eighty) items.

These terms bring with them xenophobic characteristics and other forms of prejudice against certain nationalities, generated by fear of contamination and false information. On the other hand, in the scientific community, we have names related to the disease, to the virus, and its species, such as, for example, SARS-CoV-2, the official name of the virus that causes COVID-19 (initially called n-Cov) (World Health Organization 2020a; International Committee On Taxonomy Of Viruses 2020).

The work done by Zeng et al. (2020) offers us, in addition to an understanding of knowledge organization in the pandemic, a rich terminological and chronological overview that makes it possible for us to understand the evolution and constitution of the term “COVID-19” both in relation to its term and to its classification, which we summarize in Table 1.

Table 1 – COVID-19's Term and Classification

Date	Event	Term/Classification
Jan 30, 2020	World Health Organization (WHO) declares the outbreak of the new coronavirus disease in 2019, a public health emergency of international concern	The term "2019-nCoV" was used instantly in scientific articles
Jan 31, 2020	The Classification and Statistics Advisory Committee, in WHO's Family of International Classifications (WHO-FIC) network, calls for an emergency meeting to discuss the creation of a specific code for this new type of coronavirus	ICD-10 establishes a new emergency code ("U07.1, 2019-nCoV, acute respiratory disease")
Feb 11, 2020	WHO officially announces the name of the disease, COVID-19, an acronym for "coronavirus disease 2019"	<p>The term "COVID-19" was used promptly all around the world</p> <p>ICD-10 is updated with two emergency codes: "U07.1 COVID-19, virus identified" for a diagnosis of the COVID-19 disease confirmed by laboratory tests, and "U07.2 COVID-19, virus not identified", for a clinical or epidemiological diagnosis of COVID-19, in which the laboratory confirmation is inconclusive or not available</p> <p>In the same day, a new ICTV research group baptized the new virus as "severe acute respiratory syndrome coronavirus 2" or SARS-CoV-2</p>

Source: Adapted from Zeng et al. (2020 p. 151-152).

The WHO and the actions of the International Classification of Diseases, Tenth Revision (ICD-10), together with the International Committee on Taxonomy of Viruses (ICTV) stand as the main institutions in the search for establishing the name of diseases and, in this case, of COVID-19. Their importance is due, according to Zeng et al. (2020 p. 151):

[...] allow the world to compare and share data in a consistent and standard way—between institutions, across regions and countries, and over a period of time. They facilitate the collection and storage of data for analysis and evidence-based

decision-making. Together they are contributing to the actions of eliminating semantic conflicts and avoiding information overload in real-world healthcare systems.

Thus, when establishing the official name of this new disease, it is also important to emphasize, given the actions taken to eliminate semantic conflicts, that in May 2021 the WHO announced changes in the names of the identified variants of SARS-CoV-2 (new strains), substituting the names of the countries where they were identified for the Greek alphabet. These changes to the nomenclature of the variants aim to assist in scientific discussions and mainly by non-scientific audiences. This is in order to avoid semantic conflict and also for the pandemic to be associated with and serve as a pretext for xenophobia and other forms of prejudice against certain nationalities (World Health Organization 2020b; 2021).

To define the nomenclature for these variants, WHO convened experts from the Technical Advisory Group on Virus Evolution (committee created by WHO in 2020 and composed of 25 scientists of excellence in different specialties and from different regions of the planet), the network of reference laboratories of the WHO COVID-19, representatives and experts in virological, microbial nomenclature and communication from various countries and agencies for the purpose of establishing easy-to-pronounce, non-stigmatizing labels for naming and tracing SARS-CoV-2 genetic strains. As a result, this group of experts has so far recommended the use of letters from the Greek alphabet (Alpha, Beta, Gamma, Delta). As an example, we have the variant of the virus identified in the United Kingdom, which is now called “Alpha” (World Health Organization 2020b; 2021; Zeng et al. 2020).

We see these challenges related to the terminology in the area of health in the current pandemic context, such as the term “COVID-19”, not only as an issue of the present, but also of the past and future. Because, when we analyze and describe the use of language in a specialized discourse, especially in health, it is essential that we consider the development and creation of new terms and concepts, as well as the terminological variations that can occur in the constitution of the specialized context, consequently providing an understanding of the communication flow in a way that is more efficient to the whole society.

This shows us the importance of terminological standardization and control, mainly in the health area, which involves, as in the case of COVID-19, the emergence of new diseases requiring

that, in a global way, there is a scientific collaboration so that there are researchers, institutions, countries, systems and society as a whole generating discussions and research in a common language.

3 Methodology

This is a quali-quantitative research: we made a textual analysis of the keywords found in the articles that make up the *corpus* and applied bibliometric criteria to identify occurrence patterns of the most productive authors and journals. Being a case study, we aimed to describe how the pandemic phenomenon influenced textual production in the context of Information Science, seeking to better explore the theme and look for possible understandings and reflections about the subject.

First, we sought to structure the methodology used in the research – Domain Analysis – by situating the approaches that were applied: indexing and information retrieval, bibliometric studies and studies on terminology, semantics, and discursive communities. In addition, in order to support this domain, we described the main concepts in information in health, especially with regard to the emergence of a new pandemic. The information is presented in the previous sections of this work.

Empirically, bibliographic data were collected and analyzed. According to methodological criteria, the data were necessarily structured as scientific papers, exclusively produced in the scope of Information Science and related to studies developed in Brazil. BRAPCI (*Base de Dados Referencial de Artigos em Ciência da Informação*) was the source for the retrieval of this *corpus*. The database's main objective is the dissemination of studies in IS, providing an overview of the scientific production in the area in several contexts. The database encompasses 52 journals, of which 40 are still being published, from 1972 until the present.

Considering that the database was already within the scope of IS, we used the term “COVID-19” for our search, with inverted commas, and retrieved, thus, scientific production about that term in Information Science. We opted for that term with basis on the WHO standardization realized in February 2021, which determined COVID-19 as the term to be used when referring to the disease caused by the new coronavirus. We searched for the term in the papers' areas of greater

informational content, namely: the title, abstract, and keywords. Because it is a relatively recent informational phenomenon, the articles were all from 2020 and 2021.

Initially, we retrieved a total of 340 papers. After reading the abstracts of with the objective of identifying the studies' main question, research objectives and main results, a new cut was performed to discard the papers that did not refer to studies carried out in Brazil or that did not belong in the scope of Information Science, thus constituting a *corpus* of 89 papers selected for analysis.

To attribute a classification to these papers within the scope of Information Science in relation to their thematic areas, we used the *Tesouro Brasileiro em Ciência da Informação* (Brazilian Thesaurus on Information Science), published in 2014 by IBICT and authored by Lena Vânia Pinheiro and Helena Dodd Ferrez. With this tool, it was possible for us to distribute the documents in thematic lanes and identify which subareas in Information Science have produced more studies that have as their object of study the pandemic phenomena related to COVID-19.

Thus, we opted for analyzing these documents in three lanes. Initially, we sought to understand which were the journals that published the most about the theme, about the theme, special issues and which institutions the journals were affiliated to. Subsequently, we sought to identify who the most productive authors were: main research themes, affiliation and contributions to the researched universe. Finally, we analyzed the text considering not only the most used words, but also the rarity of distinctive words. Distinctive words are defined based on a text mining technique that selects the words that are, at the same time, the most frequent within a paper, and the least frequent within the *corpus*, which constitutes the concept of rarity. In other words, distinctive words are those that distinguish each paper within the *corpus*.

In order to do so, we used the tool Voyant, an open-source web-based application that performs text mining. Initially, we started from a general and textual perspective of the *corpus* provided by the tool's function *Summary*, which offers different kinds of information about the *corpus*. Among that information we have a survey of the most representative words, that is, those that best describe (most frequent) the covered themes, and information related to the list of distinctive and characteristic words in each paper, making it possible for us to visualize, in a summarized way, what each paper in the *corpus* brings to the table that is different from the rest.

So, we observed the most frequent words (in the whole *corpus*) and the distinctive words (the five most frequent) of each paper. Next, we analyzed the distinctive words that represented each paper, and those sets were grouped into thematic clusters. As an example, we have the cluster formed by the words “disinfodemic”, “disinformation” and “infodemic”. Subsequently, we used *Voyant*’s function *Collocates Graph* to create a network visualization (Figure 1). To do so, we added the identified words to the analysis of distinctive words, together with the most frequent words in the *corpus*, all of that through the options offered by the tool. We also excluded the words that were irrelevant to our analysis, such as articles and some prepositions, through the option *Stopwords*. It is worth mentioning that this visualization reproduces the “relationships” automatically, based on the criterion of close proximity between keywords and terms driven by the force of contextual proximity. The term COVID-19 was manually removed from the visualization because, given that it was used as a search query, it would be disproportionately larger than the other words present in the image. That removal aimed to improve the quality of the image presented and the quality of the analysis by allowing less frequent terms to be analyzed as well.

The information obtained will be presented in the following section, which brings us the main results of this study.

4 Results analysis and discussion

Initially, quantitative analyses of the most productive journals and authors will be presented. The most prominent research sources in this universe are presented in Figure 1.

Table 2 – Most productive journals on the corpus theme

JOURNALS	PUBLICATION	PERCENTAGE
Liinc em Revista	24	9%
Revista Fontes Documentais	19	9%
Folha de Rosto	5	6%
Informação & Informação	4	6%
AtoZ	3	6%
P2P Inovação	3	6%
PBICIB	3	6%
Ágora	2	6%
Brajis	2	6%
Em Questão	2	6%
Reciis	2	6%
Bibliomar	2	6%
TOTAL	71	100%

Source: The authors.

The papers recovered were published between 2020 and 2021, given that the analyzed phenomenon is recent in science. Two journals were responsible for more than half of the total publications in the *corpus*: the journal *Liinc em Revista*, with 34% of the total publications, and *Revista Fontes Documentais*, responsible for 27% of the total publications in the *corpus*.

Liinc em Revista is a journal produced by IBICT – Instituto Brasileiro de Informação, Ciência e Tecnologia (Brazilian Institute of Information, Science, and Technology) and its mission is the dissemination of scientific production that reflect about the production and appropriation of scientific knowledge, especially when that knowledge is related to the changes of the contemporary world. In the 2nd issue of its 16th volume, published in the second semester of 2020, the journal had as its main theme the informational challenges faced during the pandemic, which can explain why there was an increase in the publication on this research theme in the scope of that journal. It is important to notice that the editorial policy of journals tends to induce certain themes, as was observed in this specific case.

The journal *Revista Fontes Documentais* is a publication by the Instituto Federal de Sergipe (Federal Institute of Sergipe), more specifically, it is affiliated to the Grupo de Pesquisas e Estudos

em Histórias das Bibliotecas de Ensino Superior (Research and Study Group on Histories of Higher Education Libraries). The journal publishes studies on papers, experience reports, and other studies carried out in the state of Sergipe and in other regions of Brazil. The journal was responsible for publishing papers originated from the V Colóquio Internacional “A Medicina na Era da Informação” (5th International Colloquium “Medical Sciences in the Age of Information”), organized by the Universidade Federal da Bahia together with the Universidade do Porto, in Portugal. That is a possible explanation for the high number of articles on the topic of information and pandemic retrieved by the research *corpus*.

Beside is known that the journal *AtoZ: novas práticas em informação e conhecimento* published a special issue about COVID-19 with at 10 articles in this subject, using the keywords determined by this research and in the database analyzed, only 4 of them was recovered.

A set of other 18 journals published only one article on the theme and, thus, were not considered representative enough to be included in the chart. A second quantitative analysis carried out by this study is related to the most productive authors, as shown in Figure 2:

Table 3 – Most productive authors in the corpus

AUTHORS	PUBLICATION	PERCENTAGE
NEVES, Bárbara Coelho	3	9%
SILVA, Rafaela Carolina da	3	9%
SOUZA, Amanda Damasceno de	2	6%
SANTOS, Beatriz Rosa Pinheiro dos	2	6%
OLIVEIRA, Bernardina Maria Juvenal Freire de	2	6%
LIMA, Clóvis Ricardo M.	2	6%
MOTA, Francisca Rosaline Leite	2	6%
FREIRE, Gustavo Henrique de Araújo	2	6%
BRITO, Jean Fernandes	2	6%
SILVA, Márcio Bezerra da	2	6%
FERNANDES, Mariana Ribeiro	2	6%
MELLO, Mariana Rodrigues Gomes de	2	6%
LIMA, Paulo Ricardo Silva	2	6%
LUNARDELLI, Rosane Suely Alvares	2	6%
SOUSA E SILVA, Dulce Elizabeth Lima de	2	6%
FRANÇA, André Luiz dias de	2	6%
TOTAL	34	100%

Source: The authors.

As we can see, the data in the second figure is more distributed than in the first, with little dispersion of the number of articles published in the group of the most productive authors. Two authors published three papers that are part of the *corpus*, more than all other authors; they are Bárbara Coelho Neves and Rafaela Carolina da Silva.

Neves is an assistant professor in the Universidade Federal da Bahia (UFBA) and contributes to the Programa de Pós-Graduação em Ciência da Informação (Graduate Program in Information Science) of the Universidade Federal de São Carlos (UFSCar). Her main research themes are related to the contribution of educational processes in Information Science, especially educational technologies. She has also been studying fake news in the current context and the use of artificial intelligence in health during the pandemic.

Silva is a doctoral student in the Programa de Pós-Graduação em Ciência da Informação of UNESP/Marília and completed her master's and undergraduate degree in Library Science in the same institution. She is a member of research groups related to information mediation and

organizational information, and has dedicated her studies mainly to the concepts of hybridity in libraries.

Considering the classification attributed using the *Tesouro Brasileiro em Ciência da Informação* (Brazilian Thesaurus on Information Science), 46 classes were given. In this amount, five of them that were attributed eight or more times are described below.

Table 4 – Classifications with higher incidence

Class	Incidence
Information technology and communication	21
Information literacy	11
Information access	10
Information use	9
Communication	8

Source: The authors.

Even though the initial objective of the research was to analyze publications about the pandemic and how they can affect processes and tools in knowledge organization, after retrieving the *corpus*, we observed that only four papers worked with this theme in their scope, especially when it comes to the Semantic Web and to conceptual models of knowledge organization. That finding leads us to believe that, although it is recognizably noticeable how this new pandemic crisis will affect the semantic contexts of controlled terms, the need for a new bibliographic classification and the use of tools such as taxonomies and ontologies to retrieve information that is increasingly more immediate and necessary, this concern is still walking at a slow pace when it comes to scientific publications that analyze these changes.

At a qualitative level, the text mining sought to analyze the keywords of the *corpus*, respecting the criterion of rarity in frequency, as better described below.

Thus, we can see that the impacts caused by the pandemic are provoking and prompting a range of studies in Information Science and in line with the area of Health, with health information and data as key knowledge. That becomes evident when we analyze the terms “saúde”, “desinformação”, and “infodemia”, which encompass one of the consequences of the COVID-19 pandemic, namely, the overload of information available (infodemic) in several media, such as social media, government official websites, among others, which can often provoke informational confusion, especially with the dissemination of false information. With that, we see the importance of organizing and representing information and knowledge as a source of study to help the area of Health with accessing and disseminating information, given the need to quickly publish, update and review that information (Zeng et al. 2020).

In addition, the presence of the terms “competência”, “comunidade”, “organização”, “sociedade”, “bibliotecas” and “bibliotecários” enhances the significance of the area of IS in studies oriented toward knowledge organization and representation in face of the overload of information. That is due to the fact that scientific information, data, and results must become more widely accessible and enable their reuse and redistribution to society as a whole, meeting the needs of the population. This, in turn, provides the dissemination of the promotion of and the learning with Science, encouraging scientists, policy makers, and citizens to work together for technological advancement and the advancement of innovations between and within countries, bringing as a key point the return of scientific results and data to society, guaranteeing the right to information and science (Unesco 2021; World Health Organization 2020a).

5 Conclusions

We aimed to evidence and systematize exploratory research of the Brazilian Information Science in studies related to the COVID-19 pandemic through the use of domain analysis. The objectives were reached, given that our intent was to present this information in a structured way, as well as to discover the most productive journals and authors, and the main research tendencies related to the analyzed topic.

As demonstrated in the results section, the most productive journals within our *corpus* are *Liinc em Revista* and *Revista de Fontes Documentais*, affiliated to the IBICT and to the Instituto Federal de Sergipe, respectively. These two journals were responsible for more than half of the publications that made up our analyzed *corpus*. With that in mind, we can infer that these sources embrace the theme researched here and that they have worked toward disseminating studies in this scope.

In relation to the most productive authors, the dispersion was a little smaller than the dispersion of the most productive journals: most researchers were responsible for two publications, except for Bárbara Coelho Neves and Rafaela Carolina da Silva, who authored three articles each in the *corpus*. The first author is an assistant professor in the Universidade Federal da Bahia and works mainly with the interdisciplinarity between information, technology, and education. And Silva is a doctoral student in the Programa de Pós-Graduação em Ciência da Informação of the Universidade Estadual Paulista and works with the concepts of hybridity in libraries.

In the qualitative analysis, we emphasize the relationship between the term COVID-19 with other concepts such as information, data, and pandemic, among others. It was possible to infer that the scientific dissemination of knowledge around this theme is closely related to research in information in health, and it even includes a new term in this perspective: “infodemia” (infodemic), a term that is related to the vastness of information made available via the web, especially on social media. In addition to that, the incidence frequency of terms such as *comunidade* (community) and *competência em informação* (information competence) has highlighted the need for information professionals to act effectively in this informational increase, especially when it comes to countering fake news and searching for reliable sources.

It was also noticeable that the area of knowledge organization and representation is still developing slowly in its analysis and research on how its processes and tools can be used in the informational context of the new coronavirus. That leads us to believe that the area needs to pay closer attention to the changes happening within it, so that it can contribute effectively to terminological changes, to the construction of new bibliographic or social classifications, as well as to tools such as ontologies and taxonomies, to systematize and disseminate information in health.

In future studies, we intend to broaden the research, so that it encompasses the international context, aiming to analyze comparatively how international researchers are working with studies in this scope, and whether the themes related to it are the same as those developed in Brazil or whether there is a cultural bias in these perspectives. We also intend to deepen our research on the identification of which studies in knowledge organization can contribute to the new health informational context.

References

- Atkinson, Sarah, and Abu el Haj, Mônica. "Domain analysis for qualitative public health data". *Health Policy and Planning*, vol. 11, no. 4, 1996, pp. 438-444.
- Barros, T. H. B. "Análise Discurso de Michel Pêcheux e a Organização do Conhecimento: possibilidades teórico-meotodológicas". *Ensaio em Organização do Conhecimento*. Organized by I. S. Amorim and, R. de Sales. Editora UDESC, 2021, pp. 61-78.
- Castanha, R. C. G., and Grácio, M. C. C. "Bibliometrics contribution to the metatheoretical and domain analysis studies". *Knowledge Organization*, vol. 41, no. 2, 2014, pp. 171-174.
- Cipriano, R. "Um vírus nem sempre foi um vírus. Palavras que mudaram de significado com os tempos". *Observador*, Lisboa, 2016, <http://observador.pt/2015/07/26/um-virus-sempr-um-virus-palavras-mudaram-significado-os-tempos/>. Accessed 28 jun. 2021.
- Domingues, L. "Infodemia: uma ameaça à saúde pública global durante e após a pandemia de Covid-19". *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde, Rio de Janeiro*, vol. 15, no. 1, 2021, pp. 12-17, <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2237>. Accessed 15 jun. 2021.
- Fiorillo, A., and Gorwood, P. "The consequences of the COVID- 19 pandemic on mental health and implications for clinical practice". *European Psychiatry*, vol. 63, no. 1, 2020, pp. 1-2, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7156565/pdf/S0924933820000358a.pdf>. Accessed 13 jun. 2021.
- Guimarães, J. A. C. and Tognoli, N. B. "Provenance as a Domain Analysis Approach in Archival Knowledge Organization". *Knowledge Organization: Ergon-verlag*, vol. 42, no. 8, pp. 562–569, 2015.

- Hjørland, Birger, and Albrechtsen, Hanne. "Toward a new horizon in Information Science: domain-analysis". *Journal of the American Society for Information Science*, vol. 46, no. 6, jul. 1995, pp. 400-425.
- Hjørland, Birger. "Domain analysis in information science: eleven approaches – traditional as well as innovative". *Journal of Documentation*, vol. 58, no. 4, 2002, pp. 422-462.
- Hjørland, Birger. "Domain analysis". *Knowledge Organization*, vol. 44, no. 6, 2011, pp. 436-464.
- Hjørland, Birger. "Domain analysis". *Encyclopedia of Knowledge Organization*. Organized by Birger Hjørland and C. Gnoli, vol. 44, no. 6, 2017, pp. 20-80.
- International Committee on Taxonomy of Viruses. "Naming the 2019 Coronavirus". 2020.
- Krieger, M. G., and Santiago, M. S. "Terminologia Médica e Variação". *Anais do 17º Congresso Internacional Asociación De Lingüística Y Filología De América Latina: João Pessoa, ALFAL*, 2014, <http://www.mundoalfal.org/CDAnaisXVII/trabalhos/R1134-1.pdf>. Accessed 13 jun. 2021.
- López-Huertas, M. J. "Domain Analysis for Interdisciplinary Knowledge Domains". *Knowledge Organization*, vol. 42, no. 8, 2015, pp. 570–580.
- Okada, Mihoko, et al. "Domain analysis and modeling to improve comparability of health statistics". *MedInfo*, 2001.
- Oliveira, E. F. T., and Grácio, M. C. C. "Studies of author cocitation analysis: a bibliometric approach for domain analysis". *IRIS-Revista de Informação, Memória e Tecnologia*, vol. 2, no. 1, 2014, pp. 12–23, <http://periodicos.ufpe.br/revistas/IRIS/article/view/500>. Accessed 7 jan. 2022.
- Orlandi, E. P. *O que é lingüística*. Brasiliense, 2007.
- Smiraglia, R. "Epistemology of domain analysis." *Cultural Frames of Knowledge*. Organized by R. Smiraglia and H. Lee. Ergon, 2012, pp. 111–124.
- Smiraglia, R. P. *Domain analysis for knowledge organization: tools for ontology extraction*. Elsevier Chandos Pub, 2015a.
- Smiraglia, R. P. "Domain analysis of domain analysis for knowledge organization: observations on an emergent methodological cluster." *Knowledge Organization*, vol. 42, no. 8, 2015b, pp. 602–611.
- Tennis, J. T. "Two axes of domains for domain analysis". *Knowledge Organization*, vol. 30, no. 3/4, 2003, pp. 191–195.
- Tennis, J. T. "What does domain analysis look like in form, function and genre?". *Brazilian Journal of Information Science: Research Trends*, vol. 6, no. 1, 2012, <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/3026>. Accessed 07 jan. 2022.

- Unesco. "Draft text of the UNESCO Recommendation on Open Science". *Intergovernmental Meeting of Experts (Category II)*, 2021, <https://unesdoc.unesco.org/ark:/48223/pf0000376893>. Accessed 13 jun. 2021.
- World Health Organization. "Coronavirus disease (COVID-19) Pandemic". 2020a, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed 15 jun. 2021.
- World Health Organization. "Naming the coronavirus disease (COVID-19) and the virus that causes it". 2020b, [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it). Accessed 15 jun. 2021.
- World Health Organization. "Tracking SARS-CoV-2 variants". 2021, <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Accessed 15 jun. 2021.
- Zeng, M. L., et al. "Implications of Knowledge Organization Systems for Health Information Exchange and Communication during the COVID-19 Pandemic". *Data And Information Management*, vol. 4, no. 3, 2020, pp. 148-170, <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-828345>. Accessed 13 maio 2021.

Copyright: © 2022 Evangelista, Isadora Victorino, et al. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Received: 20/08/2021

Accepted: 24/03/2022