The Sociological and Ontological Dimensions of the Knowledge Organization Domain on Google Scholar Citations

Marcin Roszkowski

Wydział Dziennikarstwa, Informacji i Bibliologii, Uniwersytet Warszawski, Nowy Świat 69, 00-927 Warsaw, Poland, Piętro III, pokój 311,
<m.roszkowski@uw.edu.pl>

Marcin Roszkowski is an assistant professor at the Faculty of Journalism, Information and Book Studies, University of Warsaw (Poland). He received his PhD degree in library and information science. His research interests include bibliometrics, conceptual modeling, linked data, metadata, and networked knowledge organization systems. Dr. Roszkowski is co-editor-in-chief of the peer-reviewed *Zagadnienia Informacji Naukowej - Studia Informacyjne* (Issues in Information Science. Information Studies), a leading scientific journal in Poland related to information science.

Roszkowski, Marcin. 2020. "The Sociological and Ontological Dimensions of the Knowledge Organization Domain on Google Scholar Citations." *Knowledge Organization* 47(2): 160-172. 35 references. DOI:10.5771/0943-7444-2020-2-160.

Abstract: This study aims to identify the profiles of researchers in the knowledge organization domain on Google Scholar Citations (GSC) and investigate its sociological and ontological dimensions. The sociological dimension is related to GSC users who declared research interests that fall within the scope of the knowledge organization domain. The ontological dimension is based on the study of these concepts. Domain analysis was used as a methodological framework for this study. A search was conducted on GSC using keywords in order to create a list of scholars who declared the knowledge organization domain as one of their research interests in their Google Scholar Profiles (GSPs). Next, the search for GSPs of authors who had published their papers in the *Knowledge Organization* journal from 2000 to 2019 was conducted. The results showed that there were 379 publicly available GSPs. Analysis of the affiliated institutions showed that the majority of them were based respectively in the USA, Brazil, and then in India. The ontological dimension of the knowledge organization domain on GSC was examined by studying keywords attached to GSPs. The most frequently used keywords were identified and using network analysis five clusters that represented the main areas of interest were extracted.

Received: 16 September 2019; Revised: 23 December 2019; Accepted: 31 January 2020

Keywords: knowledge, knowledge organization, domain analysis, Google Scholar, research

1.0 Introduction

Knowledge organization (KO) is a field of research and practice which, according to Hjørland, (2008), can be understood in two senses. In the narrow sense, KO is about describing, representing, filing, and organizing documents and document representations as well as subjects and concepts both by humans and by computer programs (Hjørland 2016, 475). In the broader meaning, KO is about the social division of mental labour, where a distinction can be made between the social organization of knowledge, and the intellectual or cognitive organization of knowledge. The broad sense is thus both about how knowledge is socially organized and how reality is organized (Hjørland 2008, 86-87). The narrow definition puts KO inside the library information science (LIS), while the broader one implies connections with disciplines such as cognitive sciences, linguistics, philosophy, and the sociology of science. Having these two perspectives on KO doesn't mean that they should be considered as isolated areas of study. On the contrary, Hjørland (2008, 87) argues that one cannot develop a fruitful body of knowledge without considering KO in the broader perspective.

For Smiraglia (2013, 2), KO is the science of the order of knowledge, and the domain of KO is a discourse community in which rigorous, self-conscious inquiry takes place concerning that which is known, and its various orderings or sequences, both those that are natural or heuristic, and those that are imposed. These orderings take the form of knowledge organization systems—tools that have been designed to support the organization of knowledge and information in order to make their management and retrieval easier (Mazzocchi 2018, 54). From this point of view, there are two main items that characterize KO:

knowledge organization processes (such as abstracting, indexing, cataloguing, subject analysis, and classifying) and knowledge organization systems.

It is also argued that KO is an interdisciplinary field of study (Hjørland, 2003; Ridenour and Smiraglia 2016; Smiraglia 2014; Szostak, Gnoli, and López-Huertas 2016), however, López-Huertas (2013) claims that the changes that have taken place over the last few decades in terms of knowledge production, culture, society, and epistemological positions call for the application of a more open, transdisciplinary approach to knowledge organization and knowledge organization systems, which is more sensitive to social demands and social welfare.

Different perspectives on knowledge organization and the complexity of the subject matter have resulted in multiple epistemic stances and, therefore, methods and theories applied in this domain. Hjørland (2016) calls them approaches or research traditions developed inside and outside KO. The former includes the practicalist and intuitivist, consensus based, facet-analytic, user-based/cognitive, and domain-analytic/epistemological approaches. These research traditions are focused on the nature of knowledge organization processes and give the theoretical and methodological framework for the design of knowledge organization systems. In the second group, there is bibliometric and information retrieval (IR) approach. The first one has a strong affiliation with LIS, and the second is mainly related to computer science. Knowledge organization processes and systems should also be considered as well as the relation to the information systems for which they are designed. Hjørland (2016) calls this area of application technological platforms and identifies physical libraries, archives, museums, bibliographic databases, and the web including semantic web technologies.

Castanha and Wolfram (2018, 14) argue that: "the domain of knowledge organization is in continuous development … it interfaces with other subject areas and is concerned with issues of a theoretical and methodological nature that contribute to the systematization, production, organization, dissemination, representation and retrieval of information in different scholarly contexts." According to them this context is continuously transformed because of its practical application and scientific communication. This calls for more complex processes to be used when studying the structure and boundaries of the knowledge organization domain.

The study of the knowledge domain known as domain analysis was first proposed by Hjørland and Albrechtsen (1995) as a paradigm within information science. The goal of domain analysis is to study knowledge domains as thought or discourse communities that are parts of society's division of labour (400). Smiraglia (2012, 111) proposed a thorough discussion on what is a domain and started with the social aspect arguing that it must be a group with a coherent ontology, which implies also a shared epistemology. Defining the domain in terms of a community we can consider at least three types: discourse community, invisible college, and epistemic community. A discourse community is a community in which an ordered and bounded communication process takes place. This communication is structured by a conceptual structure, by institutional enclosure, and by governance of discourse fora (Hjørland 2002b, 258). Smiraglia (2012) defines an invisible college and refers to Zuccala (2006, 155), who argues that it is "a set of interacting scholars or scientists who share similar research interests concerning a subject specialty, who often produce publications relevant to this subject and who communicate both formally and informally with one another to work towards important goals in the subject, even though they may belong to geographically distant research affiliates." According to this definition, there are three conditions upon which we can describe a set of actors as an invisible college: 1) they share common research interests; 2) they publish the results of their research; and, 3), they interact with each other. Finally, an epistemic community is a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area (Cross 2013, 142). Therefore, what bonds members of an epistemic community is their shared belief or faith in the verity and the applicability of particular forms of knowledge or specific truths (Haas 1992, 3). These three concepts are distinct and according to Smiraglia (2012, 112) there is little agreement in the knowledge organization field about the distinctions among and definitions of domains, discourse communities, and invisible colleges. He argues that the functional parameters of the three are divergent: the concept of "domain" suggests intellectual boundaries, the concept of "discourse community" suggests an active exchange of information, and the term "invisible college" has been used to suggest both intellectual commonality and active discourse is taking place in a socially-structured unit. The concept of an epistemic community is based on the assumption that it is a network of experts who persuade others of their shared causal beliefs and policy goals by virtue of their professional knowledge (Cross 2013, 142).

In this paper, Hjørland and Hartel's (2003) conceptualization of a domain is followed. They argue that domains are basically comprised of three kinds of theories and concepts: 1) ontological theories and concepts about the objects of human activity; 2) epistemological theories and concepts about knowledge and the ways of obtaining knowledge, implying methodological principles about the ways objects are investigated; and, 3) sociological concepts about the groups of people concerned with the objects. They claim (242) that quality research in the spirit of do-

main analysis should begin with a rigorous and interpretive study of a subject or community of interest, which will then uncover the interests underlying different conceptions of the area. Hjørland (2002a) proposed a methodological framework for domain analysis that consists of eleven approaches. They include the study of; 1) literature guides; 2) special classifications and thesauri; 3) indexing and retrieving systems; 4) empirical user studies; 5) bibliometric studies; 6) historical studies; 7) document and genre studies; 8) epistemological and critical studies; 9) terminological studies; 10) studies of structures and institutions in scientific communication; and, 11) domain analysis in professional cognition and artificial intelligence. Each of them is focused on revealing information about the domain regarding the ontological, epistemological, or sociological aspects. Tennis (2003) proposed a different perspective on domain analysis. His methodological construct consists of two axes; "areas of modulation" and "degrees of specialization." The first axis is focused on the extension of the domain and states what is included, what is not included, and what the domain is called. The second axis qualifies and sets the intension of a domain and states where the domain is positioned against other domains.

Castanha and Wolfram (2018, 15) argue that using domain analysis, it is possible to assess what is actually important or significant in a scientific field, so that aspects such as trends, patterns, processes, dominant thoughts, agents, and their relationships can be identified and analysed. Smiraglia (2015, 51) argues that domain analysis in knowledge organization involves extracting knowledge bases from functioning discourse communities. This means that the subject of this inquiry will be conceptual structures that represent different aspects of the domain (ontological, epistemological, sociological) reconstructed by the application of the methodological framework introduced by, for example, Hjørland (2002a). The literature on the application of domain analysis in knowledge organization is extensive and in depth. Smiraglia (2015) presents a detailed review of these studies using Hjørland's framework. However, the majority of these studies were based on a bibliometric approach (e.g., Araújo, Ferneda, and Guimarães 2017; Beak et al. 2013; López-Huertas 2008; Ibekwe-SanJuan SanJuan 2010; Smiraglia 2013), which involves studying data extracted directly from publications, journals, conference proceedings, or by the means of reliable information sources like the Web of Science or Scopus. This document-centric approach is both epistemologically and methodologically justified for journal articles, books, and conference proceedings because subjects of inquiry in domain analysis are understood as research artefacts that represent information exchange in the scientific discourse in the domain.

In this paper, an alternative approach is proposed for a domain analysis of the knowledge organization domain based on data extracted from the academic social network—Google Scholar Citations (GSC). This is the actor-based approach, where it is assumed that a study can be undertaken of the knowledge organization domain using the research interests of Google Scholar users as concepts that constitute the ontological dimension of the domain. These concepts are being expressed by the means of keywords attached to Google Scholar Profiles by their owners. The goal is then to identify a set of these concepts that are relevant to the knowledge organization domain and to describe this conceptual structure. At the same time, the set of actors related to these concepts will constitute the sociological dimension of this domain on this social networking site. The decision was made to use GSC for two main reasons. The first one is related to the presence of social networking sites for researchers like Academia.edu, ResearchGate, Mendeley, or GSC in the science communication landscape. These tools improve social participation, the sharing of papers, and the seeking of new collaborators (Ortega 2015, 1). Although these platforms are being questioned in terms of their reliability for bibliometric studies (Delgado López-Cózar, Robinson-Garcia, and Torres-Salinas 2013; Orduna-Malea, Martín-Martín, Thelwall, and Delgado López-Cózar 2017; Ortega 2015; Yu, Wu, Alhalabi, Kao, and Wu 2016), they can be an interesting and valuable source of information for research on the social participation of researchers within a networked environment. Martín-Martín et al. (2018, 125) argue that: "the fact that GSC is linked to Google Scholar, currently the most comprehensive academic bibliographic database" available makes this service an important source of information for researchers. The second reason for choosing GSC was the possibility to study the declared research interests of the researchers who contributed to the knowledge organization domain as representatives of the ontological dimension of this domain. The general purpose of this paper is to perform domain analysis on the knowledge organization domain within Google Scholar Citations.

2.0 Google Scholar Citations

Google Scholar Citations (GSC) was launched in 2011, seven years after Google released the Google Scholar service (Jacsó 2012, 126), as a response to Microsoft Academic Search (Ortega and Aguillo 2012). The main feature of GSC is to create and curate personal profiles (Google Scholar Profile) and keep track of the citations of its own publications (Ortega 2014, 125). These profiles are created directly by the user, who has full freedom to include precise information on his/her interests, current place of work and the specific publications authored by him/her and not by other authors with similar names. This should make these profiles more reliable, because they could be

considered almost like a personal curriculum vitae, although normalized and structured by a unique data model (Ortega 2014, 134). GSC allows its registered users to create and edit their personal records, which consist both of manually edited and automatically generated metadata. The former includes name, affiliation, research interests (maximum five keywords), e-mail, and personal web page, and the latter includes publications registered on Google Scholar, a list of co-authors, and bibliometric indicators (h-index, i10-index, total citation counts). GSC users can also manually add their publications not listed in their profiles by filling in a form. Ortega (2014, 125) argues that GSC is poorly integrated within Google Scholar, which may cause failures and errors with assignments of publications from Google Scholar to Google Scholar Profiles, and this may create an unrealistic ranking in GSC. Ortega and Aguillo (2012, 2370) argue that the major novelty of GSC when compared with the traditional citation databases is that it is focused on the author instead of the journal. Another feature of GSC is that all the metadata is expressed using a natural language. This may be seen as an advantage when expressing the actual research interest by the means of uncontrolled keywords, but also raises the problems of synonymy and spelling variants, which makes it difficult to use labels for querying the GSC database and making use of them for research purposes. Another problem with this approach is related to affiliations where the name of the same organization may be written in multiple different forms (Ortega 2015, 4). Having a Google Scholar Profile does not mean that the personal information will be publicly available. Making the profile public gives the opportunity for other users to have insight into one's publications, metrics, and research interests.

3.0 Related work

A literature search was performed on the Web of Science, Scopus, Library and Information Science Abstract, and Google Scholar using the phrase "Google Scholar Profiles" in order to find publications using data form GSP for the study of domains and scientific disciplines for methodological considerations. Only a few studies on this topic were found. The query was then expanded by adding the phrase "knowledge organization" in order to find related publications. As far as can be ascertained, there have not been any studies on the knowledge organization domain related to this social site. However, there have been a small number of research papers where authors studied other domains. Martín-Martín et al. (2018) developed a methodological framework for analysing scientific disciplines on Google Scholar Citations called Multifaceted Analysis of Disciplines through Academic Profile (MADAP). Their method was based on a multi-step process related to authors' identification of the bibliometrics domain of using Google Scholar Profiles. They used an iterative snowball process. They extracted keywords from journal articles and conference proceedings from bibliometrics in order to find the most frequently used and representative words in the discipline. After obtaining the list of terms, they checked for the existence of GSC profiles in which the authors had selected one or more of these terms as their areas of interest. They were able to construct the map of the discipline, taking into account both actors and their relationships based on co-citations and concepts based on co-occurrence in the profiles. Tetsworth et al. (2017) studied the growth of Google Scholar public profiles in orthopaedics over a twelve-month period with ninety-day interval queries. They identified members of this community using a keyword search in the Google Scholar Profiles database. They used a set of keywords they found relevant to the scope of the orthopaedics domain. They argued that although this approach might not be considered as an exhaustive data collection technique, it allowed for capturing the vast majority of those public profiles associated with the domain. The results of their study (Lande and Andrushchenko 2016) developed an algorithm of co-authors network formation on Google Scholar Citations for the physical optics domain. They applied a keyword search as the main method for research community members identification. However, their main contribution was to develop and analyse the structure of the co-authors network.

4.0 Objectives

The main objective of this study was to investigate the ontological and sociological dimensions of the knowledge organization domain based on Google Scholar. The sociological dimension relates to the community of researchers on GSC, whose research interests, declared in their profiles, fall into the scope of knowledge organization. In other words, what qualifies to be a member of this research community is to share similar research interests within the scope of the knowledge organization. The considerations whether this community meets the criteria of discourse community or epistemic community is beyond the scope of this study. The ontological dimension refers to the concepts of research activity undertaken by members of this domain. These concepts are expressed by keywords attached to Google Scholar Profiles by their users. Three research questions were formulated to detail the objectives of the study:

- RQ1: What is the extent of the knowledge organization research community on GSC?
- RQ2: What is the affiliation of community members?
- RQ3: What is the topical distribution of the research interests of community members?

Due the critique of Google Scholar Citations as a source for bibliometric studies (as reported in Section 2.0), the decision was made not to use citation metrics for authors in the knowledge organization domain on GSC. In our dataset, some inconsistencies were found concerning automatic matching publications with Google Scholar Profiles, and, therefore, the false value of citation indicators. This approach would give an interesting insight into the sociological dimension of the domain; however, it would require manual data inspection.

5.0 Methods

The foundations of domain analysis have been described in the introduction section of this paper; however, both our approach and methodological decisions need to be explained in detail. A three-dimensional conceptualization of the domain of knowledge was used (as introduced by Hjørland and Hartel 2003). The subject of this inquiry was thus the sociological dimension of the knowledge organization domain—the community of Google Scholar Citations users who share the similar research interests that fall within the scope of this domain—and the ontological dimension—the set of research interests of members of this domain that constitute the conceptual structure of the knowledge organization domain on GSC.

Referring to the methodological construct of eleven approaches to domain analysis introduced by Hjørland (2002a), it is difficult to place the approach of this study in only one of them. On the one hand, this study falls into user studies. Although the context of the research is Google Scholar Citations as a networked information system, the subject of these investigations are its users, who constitute a research community devoted to knowledge organization. On the other, there is an interest in their research interests, which are expressed using keywords. This situates the study within Hjørland's approach to language and terminology.

The basic unit of analysis was the publicly available record of a researcher (Google Scholar Profile) on Google Scholar Citation. The identification of researchers related to the knowledge organization domain was made using a two-step process, namely a keyword search and a personal name search.

5.1 Keyword search

A search was conducted on GSC using keywords in order to create a list of scholars who declared the knowledge organization domain as one of their research interests in their Google Scholar Profiles. GSC allows for searching profiles by keywords representing research interests. This type of query pattern is based on the prefix "label:" followed by any keyword. So the simplest solution for retrieving the list of researchers related to knowledge organization with publicly available profiles was label:knowledge_organization. However, this approach had serious limitations, and the decision was made to apply a simple query expansion based on different spelling variants. After manual inspection of a small sample of selected profiles, the following keywords were also used: knowledge organisation, knowledge organization system, knowledge organization systems, and KO. The rationale behind this decision was to achieve a higher level of recall in the search results and, therefore, identify more relevant Google Scholar Profiles. The assumption was made that GSC users might use different spelling variants, an abbreviation of the name of this domain, or refer to knowledge organization systems in their research interest as one of the core concepts in this domain. The profiles selected by using the keyword "KO" were further manually inspected in order to confirm the relevance to the knowledge organization domain. Due to the fact that on GSC users can use their native language for expressing their research interest, the decision was made to include the translation of the label "knowledge organization" in to the languages of ISKO chapters and use them as an additional keyword search.

5.2 Author search

Searching on GSC for profiles only using the label "knowledge organization" with spelling variants and the narrow term "knowledge organization systems" excludes all those researchers who use other keywords in their profile to represent their research interests relevant to the knowledge organization domain. Due to the fact that the *Knowledge Organization* journal is a major communication platform for the KO community, it was decided to use the authors' index when searching Google Scholar Profiles. The Web of Science was used to obtain the list of authors who had published their research papers in *Knowledge Organization* journal between 2000 and 2019. All were included, irrespective of how many papers they had published during this period. With multi-author papers, the names of all co-authors were extracted. Using the list of 615 authors, their profiles were manually searched for on GSC.

5.3 Information extraction

The first stage of the data collection process resulted in the compilation of a list of researchers whose research interest on GSP was relevant to the knowledge organization domain with the URL of their publicly available Google Scholar Profile. The next step was to extract relevant information from their profiles. Google Scholar does not offer ready to use software nor a web application for extracting data from their database based on profile ID, so the decision was made

to use a the simple solution—Google Spreadsheet and XPath queries for scraping Google Scholar Profiles based on their ID. Research data was collected automatically between June 3 and June 7, 2019 by extracting structured information from Google Scholar Profiles. Having in mind the research questions, the following metadata was extracted from GSP: name, affiliation, and keywords.

A quantitative approach was used in data analysis. For affiliations, a standard statistical analysis was performed based on the frequency of occurrences in the research dataset. For keywords, the same approach was applied, which produced a list of the most frequently used labels and allowed for a visualization of keywords distribution across the dataset. A network analysis was then applied using the VOSviewer (Van Eck and Waltman 2019) application for calculation of the keywords co-occurrence in Google Scholar Profiles. This facilitated the identification of keyword clusters that represented the areas of research interests and at the same time the structure of the ontological dimension of the knowledge organization domain on Google Scholar Citations.

5.4 Methodology limitations

Several limitations to this study need to be acknowledged. First of all, the results obtained using this method are limited to the community of researchers who have their profile on Google Scholar Citations. The assumption was not made that every researcher from the field has its profile on GSC. Next, the data extracted for the purpose of this study was based only on publicly available Google Scholar Profiles, which omits the researchers who had declared an interest in knowledge organization in their profiles, but did not make them open to the general public. Thirdly, research data represents the state of Google Scholar Citations information database at a certain point of time, and this social site is a dynamic information system where new accounts may be created or deleted and existing profiles may be updated at any time. This means that results obtained from this study represent the community of researchers as at the time when the data was extracted. The last limitation refers to the epistemic value of users' research interests on GSC. In bibliometric and informetric approaches profiling the expertise of a researcher is objectified by using subject descriptions of works he/she published. Here declared research interests have to be relied upon, which also may represent only the current area of study.

6.0 Results

The results of this study were organized according to the research questions stated in the objectives section.

6.1 The extent of the knowledge organization research community on GSC

This study revealed that there are 379 publicly available Google Scholar Profiles of researchers, who met the criteria for inclusion. Using the keyword search, 172 profiles were identified, which means that less than half (45%) of users in this study's dataset identified themselves with the knowledge organization domain by using keywords referring to the name of this domain. Using the author search, 242 from 615 authors of *Knowledge Organization* journal were found to have a Google Scholar Profile. This means that 39% of the members of *Knowledge Organization* journal discourse community have their profiles on GSC publicly available. Furthermore, only thirty-five authors (9%) of the *Knowledge Organization* journal were also identified in the keyword search.

6.2 The affiliation of community members

GSC gives users full control over the use of language for expressing the names of the affiliated institutions. This, of course, causes the problem of different names for the same institutions, which makes the analysis more difficult. For the purpose of this study, simple data cleaning techniques were applied in order to solve this problem and to proceed with quantitative analysis. The results of this study show that there were 279 institutions to which knowledge organization research community members on GSC expressed their affiliation. In only three cases were affiliation not available. Table 1 shows the top ten institutions for KO community members on GSC.

For each institution, the name of the country was established manually. The results, as shown in Table 1, indi-

Institution	Number of GSPs	Country
Sao Paulo State University	15	Brazil
Federal University of Minas Gerais	9	Brazil
University of Washington	8	USA
University of Copenhagen	7	Denmark
University of Wisconsin-Milwaukee	6	USA
University of Toronto	6	Canada
University of Illinois at Urbana-Champaign	6	USA
University of Warsaw	5	Poland
Kent State University	5	USA
University of Trento	3	Italy

Table 1. Top ten institutions with the highest number of affiliated researchers.

cate that there was a strong representation of institutions from Brazil and the USA regarding the highest number of affiliated researchers. However, further analysis of all affiliated institutions showed that amongst forty-five countries the majority of research institutions were based respectively in USA (70), Brazil (57) and in India (30). Table 2 presents the distribution of the top ten countries for affiliated researchers.

Country	Number of GSPs	Percentage of GSPs
USA	70	19.0
Brazil	57	15.4
India	30	8.1
Spain	20	5.4
France	20	5.4
Canada	19	5.1
UK	15	4.1
Italy	14	3.8
Iran	14	3,.8
Denmark	9	2.4

Table 2. Top ten countries for affiliated researchers.

The distribution of affiliated institutions across continents gave a slightly different picture of the knowledge organization community on GSC (see Table 3). Although the USA and Brazil had been represented by a significant number of researchers, European institutions took almost 40% of the stake.

Continent	Number of GSPs	Percentage of GSPs
Europe	134	36.3
North America	91	24.7
South America	66	17.9
Asia	55	14.9
Africa	11	3.0
Oceania	2	0.5

Table 3. The distribution of affiliated institutions across continents.

6.3 The topical distribution of the research interests

The assumption was made that the topical distribution of the research interest would provide insights into the ontological dimension of the knowledge organization domain on Google Scholar Citations. User-generated keywords attached to Google Scholar Profiles' were used. Due to the fact that GSC users can apply uncontrolled keywords in their profiles, there was the problem of synonymy and spelling variants. In order to conduct our analysis, the decision was made to perform a multi-step process of data cleaning.

This study found that there were 702 unique keywords created by GSC users in order to express their research interests. The first step was to establish the language in which a particular keyword was used by a GSC user. Table 4 presents the results from language recognition based on keywords attached to Google Scholar Profiles.

Language	Number of keywords	Percentage of keywords
English	564	80.3
Portuguese	63	9.0
Spanish	33	4.7
French	20	2.8
Polish	8	1.1
Finnish	5	0.7
Danish	4	0.6
German	3	0.4
Russian	2	0.3

Table 4. The language of keywords as research interests.

As shown in Table 4, the majority of keywords were in English, although a significant number of occurrences were also recorded for Portuguese. The latter result corresponds to the distribution of countries for affiliated institutions, where Brazil was the second most represented country.

Next, each keyword was translated into English. The same data cleaning techniques were then applied as for corporate names when analysing affiliations. It was possible to discover and normalize keywords with different spelling variants, abbreviations, and plural/singular forms. This was done automatically by the means of Google Refine features. This process reduced 631 unique labels to 597 normalized keywords. Researchers who expressed their research interests in their native language (other than English) did not use any mixed forms. This means that they used labels only in their native language.

GSC allows users to create up to five keywords representing their research interest. An analysis of keyword distribution per profile was conducted. The research data (Table 5) shows that less than half of Google Scholar Profiles contain a maximum number of keywords. There were also Google Scholar Profiles (17) where users did not use any keyword.

The results of this study show that the majority (73%) of keywords used by GSC users were multi-word labels. The in-depth analysis of the keywords structure showed that almost half (47%) of the keywords created by com-

Number of keywords	Number of GSPs	Percentage of GSPs
0	17	4.5
1	32	8.4
2	37	9.8
3	72	19.0
4	76	20.1
5	145	38.3

Table 5. Keyword distribution in GSPs.

munity members consisted of two words. The results also showed the usage of three (17%), four (5%), five (2%), and six-word phrases (1%). With multi-word phrases over twenty-five characters, the problem with proper display occurs. The phrase is cut after reaching the limit of characters and the symbol "…" is automatically added at the end. Although these keywords are not properly presented on screen, they can be used in full form for searching purposes.

During the analysis of keyword frequency, it was determined that there were only a small number of keywords that were relatively often used by GSC users in their profiles, but the great majority of labels were used only a few times. Figure 1 presents the distribution of keyword occurrences in the dataset.

Most of the keywords (97%) appeared less than ten times, however, the percentage of single usage is very high (77%). Table 6 presents a list of keywords that represent the research interests, which appeared ten times or more in the dataset of this study.

It can be seen from the data in Table 6 that the most frequently used keyword describing the research interest in this community was knowledge organization. However, it was attached to less than half (42%) of Google Scholar Profiles. Relatively low values for other keyword usage didn't allow for general conclusions. What can be observed here is the presence of keywords that refer to the disciplinary affiliation of library and information science. The most frequently used keywords also represented system-oriented research interests rather than user-centered ones.

Keyword	Occurrence
knowledge organization	160
information science	61
metadata	25
information retrieval	25
ontologies	24
Semantic Web	23
library and information science	23
digital libraries	23
classification	23
knowledge organization systems	17
knowledge management	15
information organization	14
indexing	12
thesauri	11
Natural Language Processing	11
artificial intelligence	11
scientometrics	10
library science	10
bibliometrics	10

Table 6. Keywords that were used in GSPs ten times or more.

The results obtained from this analysis can be compared with the outcomes of the study conducted by Beak et al. (2013). Their approach was based on domain analysis in knowledge organization in four countries: Brazil, South

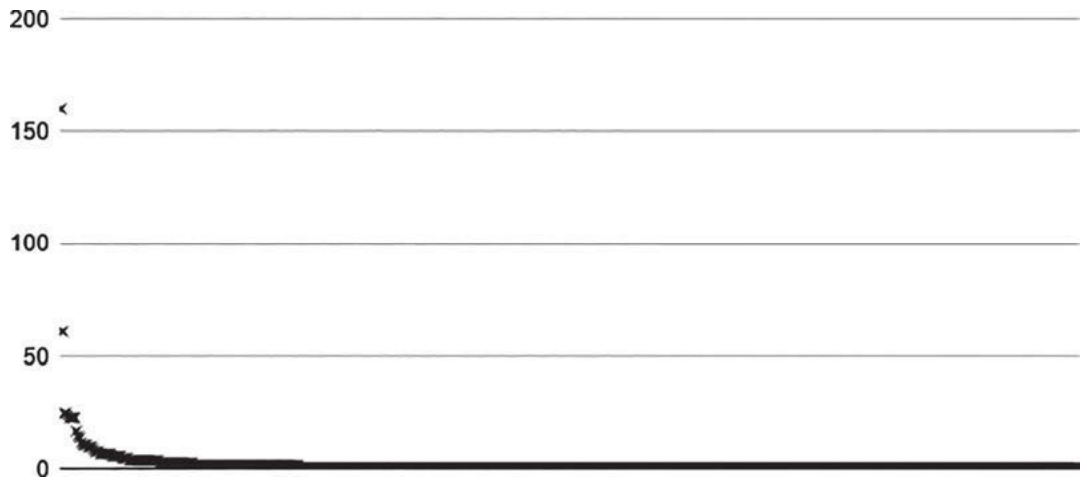

Figure 1. *The distribution of keyword frequency.*

Korea, Spain, and the USA. They identified the most frequently used keywords in papers published in two KO-specialized journals from each country over a five-year period (2007-2011). The decision was made to compare only Brazil and the USA because of their significant input to the research data in this study (Table 7).

Table 7 shows the overlapped keywords among these two studies. Although some minor similarities can be observed, the profiles of each country in these two studies are different. For the USA, there are matching keywords that refer to rather general topics (information science, knowledge organization), and there is a common interest in classification issues. A similar situation applies to Brazil. Here there are overlapping keywords referring to information organization and information science and two specific keywords pointing at indexing and ontologies. However, it is important to bear in mind that a comparison is being made between the declared area of interest of a researcher on GSC and the actual topic of a paper published in a scientific journal.

The next step with the analysis of keyword distribution was to create clusters with frequently co-occurred labels in GSPs. A dataset was prepared where a list of keywords had been attached by the user for each profile. VOSviewer (http://www.vosviewer.com) was used for network analysis and visualization. For the purpose of keyword co-occurrence analysis, labels that appeared at least five times in research data were included. The results obtained from network analysis are shown in Figure 2.

Due to a high number of occurrences, the keyword "knowledge organization" was established as a central node in this network. It was possible to extract five clusters of keywords based on a high level of association in GSPs (Table 8). These were extracted with VOSviewer using default parameters for association strength. Total link strength attribute was used as a standard weight parameter in VOSviewer for calculation of the role of the keyword in a particular cluster. This attribute indicates the total strength of the links of an item with other items in the cluster (Van Eck and Waltman 2019, 7). The order of keywords inside a cluster reflects their connectivity with others based on the value of total link strength attribute.

Cluster one consisted of thirteen keywords from which digital libraries had the highest value for total link strength, which means that it was the most linked keyword in this cluster. This cluster represents the area of research interests related to information processing and information systems with the application of particular IT technologies. The presence of the keyword information literacy was also observed, which was strongly connected to information retrieval, digital libraries, and knowledge organization systems. Many of the concepts in this cluster refer to the information retrieval approach identified by Hjørland (2016) as one of the approaches developed outside KO.

Cluster two was strongly related to core concepts in the knowledge organization domain but with the library and information science perspective. It consisted of nine keywords, where the central concept was knowledge organization with library and information science as the second most linked keyword in this cluster. An interest in different types of knowledge organization systems was observed, as well as the context of their application in the indexing process.

Cluster three consisted of nine keywords and it was related to networked knowledge organization systems and their application in the semantic web. There were three keywords that organized the network of relations here: semantic web, ontologies, and metadata. An interest in the issues related to information interaction (human-computer interaction) and information visualization was noticed. Although these two keywords did not play an im-

Brazil		**USA**	
Google Scholar Citations	Beak et al.	Google Scholar Citations	Beak et al.
bibliometrics	**automatic indexing**	**classification**	**classification**
documentary languages	classification systems	information organization	information
epistemology	**information organization**	**information science**	information retrieval
indexing	**information science**	**knowledge organization**	**information science**
information organization	knowledge management	library and information science	**knowledge organization**
information science	knowledge management instruments	metadata	model
knowledge organization	knowledge representation	ontologies	retrieval
library	online catalog	Semantic Web	science
ontologies	**ontologies**	social media	systems
Semantic Web	**ontology**	scholarly communication	web

Table 7. The comparison of keywords from Google Scholar Profiles and journal articles.

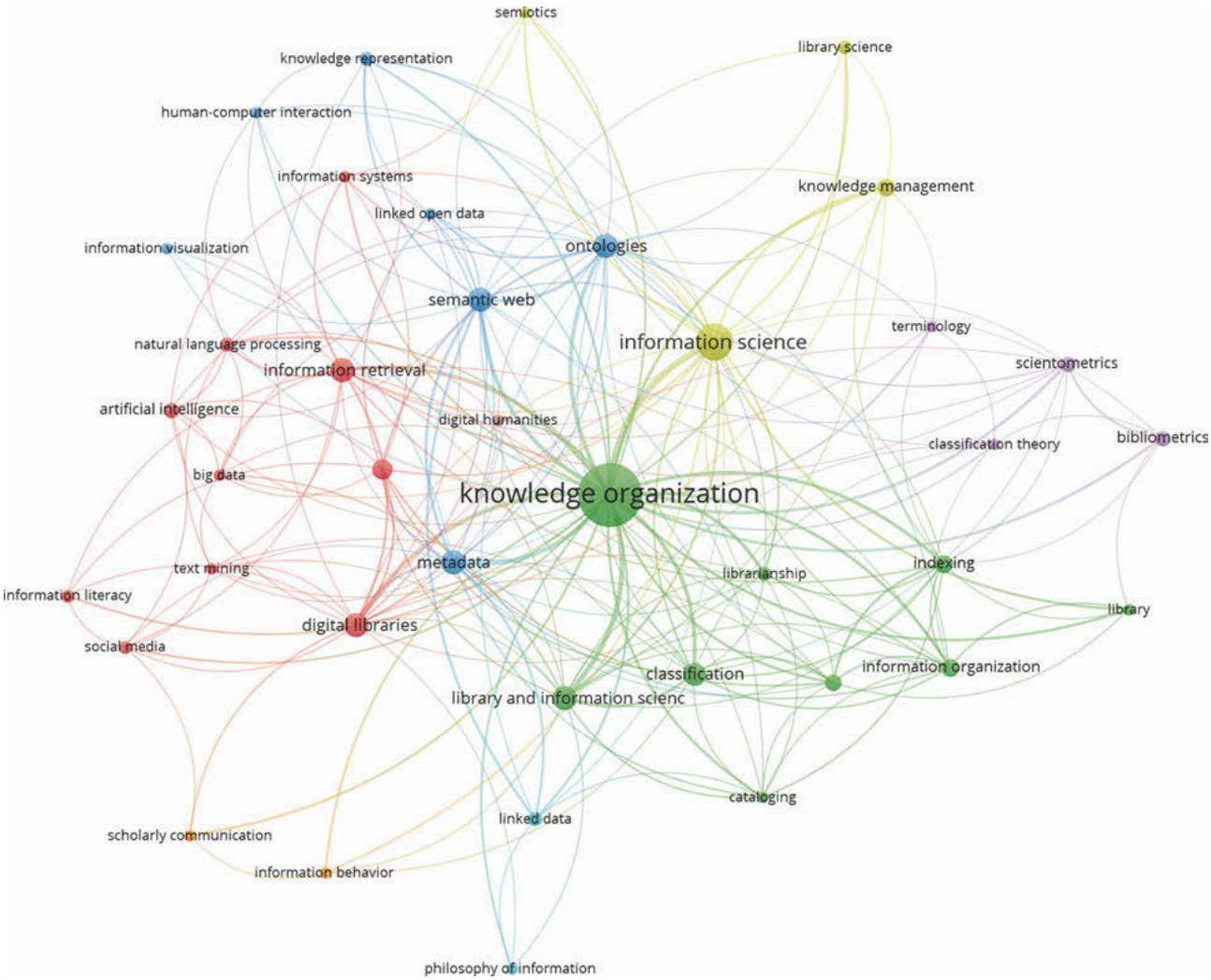

Figure 2. The network of the keywords co-occurrence.

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
digital libraries	knowledge organization	Semantic Web	information science	scientometrics
information retrieval	library and information science	ontologies	knowledge management	bibliometrics
knowledge organization systems	classification	metadata	library science	classification theory
information systems	indexing	Linked Data	semiotics	terminology
natural language processing	information organization	knowledge representation		
digital humanities	thesauri	linked open data		
big data	librarianship	human-computer interaction		
social media	cataloging	philosophy of information		
artificial intelligence	library	information visualization		
text mining				
information behavior				
scholarly communication				
information literacy				

Table 8. Five clusters of keywords based on high level of association.

portant role in the cluster, they often co-occurred with ontologies and semantic web. This cluster may be referred to what Hjørland (2016, 481) called "KO on the internet" as one of the technological platforms for the application of knowledge organization systems.

Cluster four represented rather general research interests of researchers related to the knowledge organization domain on GSC. There were four keywords in this cluster. The most linked keyword here was information science and knowledge management second. The level of connectivity of the other two keywords was rather low. The topic of semiotics was also found to be related to the other two from Cluster three—semantic web and ontologies.

Finally, cluster five was the smallest and the weakest group of keywords in terms of their inner connectivity. There were two keywords that organized the network of relationship, and they represent a quantitative approach to studying scientific output—scientometrics and bibliometrics. However, researchers who declared their interests in both of them also pointed to information science as a field of study. A relationship was observed between the topics' terminology, classification theory, and ontologies. This cluster refers to a bibliometric approach identified by Hjørland (2016) as another research tradition developed outside KO.

7.0 Conclusion

The main goal of this study was to investigate the ontological and sociological dimension of the knowledge organization domain on Google Scholar Citations. Three hundred seventy nine publicly available Google Scholar Profiles were identified, out of which less than half of them contained keywords directly referring to the knowledge organization domain either by its name or by the general term "knowledge organization systems." The assumption was made that discipline is what is performed by those who cultivate it (Martín-Martín et al. 2018, 1252), and the decision was made to use a list of authors from the *Knowledge Organization* journal, a major platform for dissemination of research results in this domain, as a point of reference in this study. Surprisingly, only 39% of authors who have published their papers in this journal had their publicly available Google Scholar Profile, and only 9% of them used one of the general terms in the description of their profiles. A similar study was conducted by Martín-Martín et al. (2018); however, the subject of their investigation was bibliometrics. They were able to identify 811 publicly available Google Scholar Profiles. A research study with a broader scope on Google Scholar Citations was conducted by Ortega (2015). The results of his study showed that the disciplines with the highest number of labels are computer sciences, engineering and physics, and astronomy. Only 6% of labels used in Google Scholar Profiles were classified by him as social science. The total number of Google Scholar Profiles related to social science was 2,833, but this study was conducted in December 2012.

It is interesting to note that although the majority of Google Scholar Profiles are affiliated with European institutions, there is strong evidence to suggest the USA (19%) and Brazil (15%) as leading countries. It would require further research to determine the causes of this phenomenon, however, reference can be made to Ortega's (2014, 128) research on academic search engines. The results of his study showed that as for 2013, the USA, the UK, and Brazil were the most represented countries on Google Scholar Citations.

The ontological dimension of the knowledge organization domain on Google Scholar Citations consisted of concepts that represented research interests of its members. By taking into account the keyword frequency, a strong connection was found between knowledge organization and library and information science and metadata-related concepts. Using a clustering technique based on keyword co-occurrence in the profile, it was possible to create five clusters of associated keywords. These represent the main areas of research interest for this group of researchers and, at the same time, the structure of the ontological dimension of this domain on Google Scholar Citations. However, these results need to be interpreted with caution, because the keyword analysis applied was only based on their lexical forms, without semantic categorization.

Finally, although Google Scholar Citations is considered to be an important source of information about researchers, it can be argued that systematic manual curation of one's own profile is necessary. Although setting up the Google Scholar Profile is being done manually by the user, the list of publications is being imported from Google Scholar automatically. This may cause serious consequences when personal names are not properly recognized. During this study, several examples of such mismatching were found, which resulted in false citation metrics. This was the main reason why bibliometric indicators were not included in this study.

References

Araújo, Paula Carina De, Edberto Ferneda and José Augusto Chaves Guimarães. 2017. "The Relation between the Domains of Information Retrieval and Knowledge Organization in International Journals." *Pesquisa brasileira em ciência da informação e biblioteconomia* 12, no. 1: 82-8.

Beak, Jihee, Jeanette Glover, Daniel Martínez-Ávila and Suellen Oliveira Milani. 2013. "International Comparative Domain Analysis in Knowledge Organization Research Topics in Four Countries - Brazil, South Korea, Spain and the United States." In *Proceedings from North American Symposium on Knowledge Organization*, Vol. 4.

University of Wisconsin-Milwaukee, ed. Richard P. Smiraglia, 30-9. doi:10.7152/nasko.v4i1.14643

Castanha, Renata Cristina Gutierres and Dietmar Wolfram. 2018. "The Domain of Knowledge Organization: A Bibliometric Analysis of Prolific Authors and Their Intellectual Space." *Knowledge Organization* 45: 13-22.

Cross, Mai'a K. Davis. 2013. "Rethinking Epistemic Communities Twenty Years Later." *Review of International Studies* 39: 137-60.

Delgado López-Cózar, Emilio, Nicolas Robinson-Garcia and Daniel Torres-Salinas. 2013. "Manipulating Google Scholar Citations and Google Scholar Metrics: Simple, Easy and Tempting." *EC3 Working Papers 6*. doi:10.3916/C41-2013-04 PDF

Van Eck, Nees Jan and Ludo Waltman. 2019. *VOSviewer Manual*. Leiden: Universiteit Leiden, http://www.vosviewer.com/download/f-13w2.pdf

Haas, Peter M. 1992. "Introduction: Epistemic Communities and International Policy Coordination." *International Organization* 46: 1-35.

Hjørland, Birger. 2002a. "Domain Analysis in Information Science: Eleven Approaches - Traditional as Well as Innovative." *Journal of Documentation* 58: 422-62.

Hjørland, Birger. 2002b. "Epistemology and the Socio-Cognitive Perspective in Information Science." *Journal of the American Society for Information Science and Technology* 53: 257-70.

Hjørland, Birger. 2003. "Fundamentals of Knowledge Organization." *Knowledge Organization* 30: 87-111.

Hjørland, Birger. 2008. "What Is Knowledge Organization?" *Knowledge Organization* 35: 86-101.

Hjørland, Birger. 2016. "Knowledge Organization (KO)." *Knowledge Organization* 43: 475-84.

Hjørland, Birger and Hanne Albrechtsen. 1995. "Toward a New Horizon in Information Science: Domain-analysis." *Journal of the American Society for Information Science* 46: 400-25.

Ibekwe-SanJuan, Fidelia and Eric Sanjuan. 2010. "Knowledge Organization Research in the Last Two Decades: 1988-2008." In *Paradigms and Conceptual Systems in Knowledge Organization. Proceedings of the Eleventh International ISKO Conference 23-26 February 2010 Rome, Italy*, ed. Claudio Gnoli and Fluvio Mazzocchi. Advances in Knowledge Organization 12. Würzburg: Ergon, 115-21.

Jacsó, Péter. 2012. "Google Scholar Author Citation Tracker: Is It Too Little, Too Late?" *Online Information Review* 36: 126-41.

Lande, Dimitry V. and Valentyna B. Andrushchenko. 2016. "Formation of Subject Area and the Co-Authors Network by Sounding of Google Scholar Citations Service." In *Summer InfoCom Advanced Solutions 2016 Conference Proceedings II Scientific and Practical Conference Київ, Україна 1-3 Червня 2016 Року,* ed. П.І. Бідюк et al. Київ: Вид-во Інжиніринг, 66-9. https://iconfs.net/uploads/infocom2016/proceeding.pdf

López-Huertas, María J. 2008. "Some Current Research Questions in the Field of Knowledge Organization." *Knowledge Organization* 35: 113-36.

López-Huertas, María. 2013. "Reflexions on Multidimensional Knowledge: Its Influence on the Foundation of Knowledge Organization." *Knowledge Organization* 40: 400-7.

Martín-Martín, Alberto, Enrique Orduna-Malea and Emilio Delgado López-Cózar. 2018. "A Novel Method for Depicting Academic Disciplines through Google Scholar Citations: The Case of Bibliometrics." *Scientometrics* 114: 1251-73.

Mazzocchi, Fluvio. 2018. "Knowledge Organization System (KOS): An Introductory Critical Account." *Knowledge Organization* 45: 54-78.

Mustafa El Hadi, Widad. 2008. "Discourse Community Analysis. Sense Construction Versus Non-Sense Construction." In *Culture and Identity in Knowledge Organization: Proceedings of the Tenth International ISKO Conference 5-8 August 2008 Montréal, Canada*, ed. Clément Arsenault and Joseph T. Tennis. Advances in Knowledge Organization 11. Würzburg: Ergon, 302-6.

Orduna-Malea, Enrique, Alberto Martín-Martín, Mike Thelwall and Emilio Delgado López-Cózar. 2017. "Do ResearchGate Scores Create Ghost Academic Reputations?" *Scientometrics* 112: 443-60.

Ortega, José Luis. 2014. *Academic Search Engines: a Quantitative Outlook*. Oxford: Chandos.

Ortega, José Luis. 2015. "How Is an Academic Social Site Populated? A Demographic Study of Google Scholar Citations Population." *Scientometrics* 104: 1-18.

Ortega, José Luis and Isidro F. Aguillo. 2012. "Science Is All in the Eye of the Beholder: Keyword Maps in Google Scholar Citations." *Journal of the American Society for Information Science and Technology* 63: 2370-7.

Ridenour, Laura and Richard P. Smiraglia. 2016. "How Interdisciplinary Is Knowledge Organization? An Epistemological View of Knowledge Organization as a Domain." *Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society. Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil,* ed. José Augusto Chaves Guimarães, Suellen Oliveira Milani and Vera Dodebei. Advances in Knowledge Organization 15. Würzburg: Ergon, 43-50.

Smiraglia, Richard P. 2012. "Epistemology of Domain Analysis." In *Cultural Frames of Knowledge,* ed. Richard P. Smiraglia and Hur-Li Lee. Würzburg: Ergon, 111-24.

Smiraglia, Richard P. 2013. "The Epistemological Dimension of Knowledge Organization." *IRIS: informação, memória e tecnologia* 2, no. 1: 2-11.

Smiraglia, Richard P. 2014. *The Elements of Knowledge Organization*. Cham: Springer International.

Smiraglia, Richard P. 2015. *Domain Analysis for Knowledge Organization: Tools for Ontology Extraction*. Amsterdam: Elsevier.

Szostak, Rick, Claudio Gnoli and María López-Huertas. 2016. *Interdisciplinary Knowledge Organization*. Cham: Springer International.

Tennis, Joseph T. 2003. "Two Axes of Domains for Domain Analysis." *Knowledge Organization* 30: 191-5.

Tetsworth, Kevin, Dave Fraser, Vaida Glatt and Erik Hohmann. 2017. "Use of Google Scholar Public Profiles in Orthopedics: Rate of Growth and Changing International Patterns." *Journal of Orthopaedic Surgery* 25, no. 1: 1-7.

Yu, Min-Chun, Yen-Chun Jim Wu, Wadee Alhalabi, Hao-Yun Kao and Wen-Hsiung Wu. 2016. "ResearchGate: An Effective Altmetric Indicator for Active Researchers?" *Computers in Human Behavior* 55, part B (February): 1001-6.

Zuccala, Alesia. 2006. "Modeling the Invisible College." *Journal of the American Society for Information Science and Technology* 57: 152-68.