
THE RELATION BETWEEN THE DOMAINS OF INFORMATION RETRIEVAL AND KNOWLEDGE ORGANIZATION IN INTERNATIONAL JOURNALS

A relação entre os domínios de recuperação da informação e organização do conhecimento em periódicos científicos internacionais

Paula Carina de Araújo (1), Edberto Fereda (2), José Augusto Chaves Guimarães (3)

(1) Universidade Estadual Paulista Julio de Mesquita Filho (UNESP), Rua Presidente Faria, 121 AP 508 – Centro – Curitiba, PR, paula.carina.a@gmail.com (2) Av. Hygino Muzzi Filho, 737 - 17525-900 - Marília - SP - Brazil, ferneda@marilia.unesp.br (3) Av. Hygino Muzzi Filho, 737 - 17525-900 - Marília - SP - Brazil, guima@marilia.unesp.br

Abstract

The objective of this research is to analyze the thematic of international scientific production published during the 2006 – 2015 period on Knowledge Organization and Information Retrieval indexed in Web of Science and Scopus databases. This study develops a qualitative and exploratory research of a corpus of 100 papers that establishes a relationship between Knowledge Organization and Information Retrieval. The study also analyses the papers that are part of the corpus and classifies them into three categories, using the dimensions of Knowledge Organization that are considered themes in the Brazilian Congress on Knowledge Organization and Representation organized by the International Society for Knowledge Organization (ISKO): epistemological dimension, applied dimension, social and political dimension. We found that most of the scientific production is concentrated in the applied dimension, which can be explained by the applied characteristic of Information Retrieval. The epistemological dimension is the second recurrent, and the social and political dimension has low incidence, which can be explained once the corpus is formed by the international scientific production and is not reflected in a dimension designed for a Brazilian Congress.

Keywords: Information retrieval. Knowledge organization. Scientific production. Dimensions of Knowledge Organization.

Analisa os temas da produção científica internacional, publicada entre 2006 e 2015, sobre Organização do Conhecimento (OC) e Recuperação de Informação (RI) indexada nas bases de dados Web of Science (WoS) e Scopus. Desenvolve uma pesquisa qualitativa e exploratória e tem como corpus 100 artigos que relacionam a Organização do Conhecimento e a Recuperação de Informação, publicados entre 2006 e 2015 em periódicos indexados nas bases de dados Web of Science e Scopus. Analisa os artigos que compõem o corpus e classifica-os utilizando como categorias as dimensões da organização do conhecimento, consideradas subtemas do Congresso Brasileiro de Organização do Conhecimento da International Society for Knowledge Organization (ISKO): dimensão epistemológica, dimensão aplicada, dimensão social e política. Reconhece que a maioria da produção científica está concentrada na dimensão aplicada, o que pode ser explicada pela característica aplicada da RI. A dimensão epistemológica aparece como segunda mais recorrente e a dimensão social e política tem baixa incidência o que pode estar relacionado ao fato do corpus ser formado pela produção científica internacional e não refletir em uma dimensão que foi pensada para um congresso brasileiro.

Keywords: Recuperação de Informação. Organização do conhecimento. Produção Científica. Dimensões da Organização do Conhecimento

Resumo

1 Introduction

The scientific production of a domain determines its degree of development and evolution. The analysis of a domain is essential to understand it, identify research trends and contribute to the advancement of production and scientific communication. For, "it is through knowledge of the scientific and academic productivity, [...], that one may know what is being researched and

how the production of these disseminated research may influence the scientific community." (Fujino et al., 2007, pp. 199-200).

It is understood that the publications in the field of Information Science (IS) - in this research especially on the theme "Knowledge Organization (KO) and Information Retrieval (IR) - "reflect the status of this science and make it possible to examine and assess the

content produced by scientists, as well as trends, theoretical methods and influences." (Arboit et al. 2010, p.19).

KO and IR are key areas in IS studies. The relationship of these two areas is evident, given that information is organized in order to be retrieved by users. Therefore, the objective of this research is to analyze the themes of the international scientific literature published between 2006 and 2015, about Organization of Knowledge and Information Retrieval indexed in Web of Science and Scopus databases.

This analysis is proposed due to the authors' interest on the scientific production in KO and interest in the study of articles relating KO with IR. This research question arose when attending the course Computational Models of Information Retrieval at the Information Science Graduate Program at Universidade Estadual Paulista Julio de Mesquita Filho (UNESP).

Considering that the analysis of the scientific production of a domain or the relationship between domains allows its knowledge, its construction and interdisciplinary, as Rendón Rojas (2008) stated, we can state that this is another motivating factor that justifies the conducting of this research.

Following this introduction, the theoretical reference on IR and KO is presented. Then the data collected and their analysis are shown, making it possible to recognize the scientific production on KO and IR. Finally, the conclusions are presented.

2 Information Retrieval and Knowledge Organization

The term Information Retrieval first appeared in 1951 and was presented by Calvin Mooers, an American computer scientist. In this classical definition, the author stated that "information retrieval is the name for the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him" (Mooers, 1951, p. 25).

The author also explain that "information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques, or machines that are employed to carry out the operation. Information retrieval is crucial to documentation and organization of knowledge".

From this definition, and especially the statement that "information retrieval is crucial for documentation and organization of knowledge" one can realize that from the beginning of the studies in this domain, there is the understanding of the close relationship between IR and KO, core domains of IS.

IR is studied by both Computer Science (CS), and IS. "In field of information science, information retrieval refers to the interaction of people with information retrieval systems for relevant judgments of the information retrieval results from the selection of the search strategy" (Wilson, 2000 cited by Khapre and Basha, 2012, p 232.). Another characteristic is that "its research focuses on the specific behavior of people search for location information" (Marchionini, 1995 cited by Khapre and Basha, 2012, p 232.).

Khapre and Basha (2012, p. 238) conducted a research to carry out a comparative analysis of IR in the field of IS and CS. In their research, the authors have shown that the two fields have different trends. Some theoretical developments turn to integration, others are more static. It is noteworthy that "the similarity of the two disciplines field of information retrieval is more active and cross the field of personalizes, adaptive and implicit evaluation of information use of this techniques has significantly improved the retrieval performance."

It is important to note that Information Retrieval systems, regardless of the area, are implanted to answer questions to users, this is why they exist and it is their basic objective, in addition to being their main characteristic, observed during their conception. (Saracevic, et al., 1988). From this observation, it is possible to set out a reflection about the relationship between IR and KO, as proposed in this research.

Guimarães (1990) studied the Thematic Information Retrieval, understood as the one which allows access to information resources through the subject. In this context, he explains that IR should be understood as the link of a chain in the information flow, considering this as its dynamic aspect. In this way, the author understands IR as a data processing procedure along with document analysis.

From another point of view, "the information retrieval process consists of identifying the set of documents (corpus) of a system, which meets the user's need for information" (Ferneda, 2003, p. 14). When representing the IR process, the author indicates the representation of documents as the second step involving the subject analysis of a document, and the translation into a linguistic expression that presupposes a documentary language to ensure the standardization of indexing and effectiveness in IR.

Dahlberg (1993) recognized that the need to organize knowledge in ancient times was always closely related to librarians and philosophers. And, over the years, other professionals have been engaging in this activity. Currently, technology professionals are very much present and interested in applying KO methodologies, for example. This involvement in the 90s was due to the accelerated development of information and communication technologies that began to provide new forms of representation and IR.

When Hjørland and Albrechtsen (1995) claim that it is necessary to incorporate knowledge on the cultures in which information systems are functioning, they are proposing a socio-cognitive approach to KO. Later, in explaining it, Hjørland (2013) states that one cannot create an operational, transferable and standardized definition of a domain that ignores the historical, social and political issues defining the field.

Hjørland is one of the prominent authors on epistemological and socio-cognitive issues of KO. He explains that KO studies the nature and quality of processes and KO systems. For the author, "Knowledge Organization is about activities such as document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of "Memory Institutions"" (2008, p. 86).

Following the socio-cognitive approach, Esteban Navarro and Garcia Marco (1995, p. 149) had a complete definition and, as stated by Guimarães (2008, p. 86), that strengthens the social dimension, materialized and cyclical of knowledge. The authors claim that KO is a

discipline devoted to the study and development of fundamentals and techniques of planning, construction, management, use and evaluation of description systems of, cataloging, ordering, classifying, storing, communicating and retrieving of documents created by men to testify, preserve and transmit their knowledge and their actions, from their content, in order to ensure their conversion into information capable of generating new knowledge. (translated by the author).

Barite (2001, p.39-40) discusses about access to knowledge and states that KO seeks to provide conceptual content for various practices and social activities related to them. In addition, "it intends to function as an instrument for processing, management and use of information, comprehensive and inclusive of phenomena and applications related to the structure, layout, access and dissemination of socialized knowledge."

Hjørland (2002) explains that the central point in his approach is that instruments, concepts, meanings, information structures, information needs and relevance criteria are established in discourse communities where the communication process is established. It is noted that in this view, there is a change of focus of Information Science, from individuals and/or computers to the socio-cultural and scientific world.

In treating the organization of documents, Jaenecke (1994, p. 8) states that "it's main objective has so far been the ordering and supply of knowledge." The author makes this statement in the article in which he questions the purposes of KO. The provision of knowledge, pointed out by Jaenecke (1994) relates to IR, once information is organized in order to be found by those who need it. And just as Guimarães and Sales (2010), it is understood that the objectives of documentary analysis, one of the areas of interest of KO, are

directly linked to the representation of document content and IR.

Smiraglia (2002) clearly establishes the relationship between IR and KO by stating that the latter has been the domain of construction of storage and retrieval instruments of documentary entities. He also explains that "catalogs, indexes, and databases have been constructed to allow the rapid manipulation of and retrieval from large collections of surrogate records that represent documents, which in turn represent recorded knowledge" (Smiraglia, 2002, p. 331).

The concepts and the theoretical opinions presented here reinforce the explicit relationship between IR and KO. Chapter three presents the methodological trajectory and next the results related to the literature that identifies IR and KO.

3 Methodological Trajectory

In order to address the themes of international scientific production on KO and IR, published in the last ten years, a qualitative exploratory study was conducted. It was decided that scientific articles published between 2006 and 2015 and indexed in database Web of Science (WoS) of Thomson Reuters and Scopus from Elsevier databases would be collected. These databases were chosen as they are the most extensive today (Aghaei Chadegani et al., 2013) and as they allow, consequently, greater visibility and certify the quality of articles, because of their strict criteria for the journal selection that they index.

Data collection was carried out on July 30, 2015. The search was conducted in the fields title, abstract and keywords; the period was limited to the years 2006-2015, and the search strategy used was: "information retrieval" AND ("knowledge organization" OR "information organization"). In WoS database, 57 documents were retrieved, and 92 in Scopus. The retrieved articles were stored and organized using the reference management software Zotero.

Once organized, Zotero was used to identify repetitions. This first analysis of data enabled the identification of 41 repeated articles, that is, articles that were indexed in both databases, therefore the repetitions were excluded. Although the investigation was limited to scientific papers, the search retrieved: one paper from proceedings, two book chapters, one interview, one book review. Initially, 46 documents were excluded from the research corpus.

The second stage of the research consisted in analyzing the content of the articles by reading the titles, abstracts and keywords. The method of organization and analysis was the categorization, and the categories used were the three sub-themes named dimensions of KO proposed by ISKO in the Brazilian chapter congresses: epistemological dimension, applied dimension, social

and political dimension. At this stage, three other articles that were retrieved, but were not part of the scope were excluded. Therefore, the corpus of this research is now composed of 100 scientific articles.

4 International Scientific Production on Knowledge Organization and Information Retrieval

Articles collected from WoS and Scopus on KO and IR were categorized and analyzed, as according to Sales, Guimarães, Oliveira and Bufrem (2011, p. 1), it is believed that "looking at the intellectual space of a certain area of knowledge in order to understand the movement of the elements that compose it [...] enables greater understanding of the characteristics and behavior of this very area".

We identified 65 (sixty-five) scientific journals in which the 100 (one hundred) articles that comprise the corpus of this research were published. Figure 1 shows journals with 2 (two) or more articles, considered the most representative, totaling 13 (thirteen) titles. The journals with more articles are *Knowledge Organization* with 11 (eleven) articles and, secondly *Journal of the American Society for Information Science and Technology* with 9 (nine) articles.

Analyzing the scope of the two journals and considering they are prestigious international publications, one can understand the reason of this result. The journal *Knowledge Organization* is an international journal directed to the theory, classification, indexing and knowledge representation, as noted in the journal description. *Journal of the American Society for Information Science and Technology* points out as major areas of research in the journal scope: knowledge production, knowledge organization, design and evaluation of information systems, access and use of information and information policy.

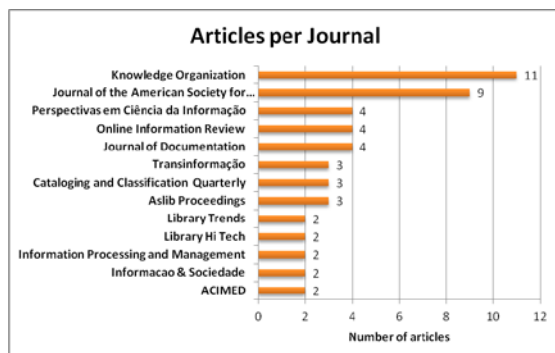


Figure 1: Articles per journal

It is also important to highlight the significant presence of Brazilian journals: *Perspectivas em Ciência da Informação* with four articles, *Transinformação* with three articles and *Informação e Sociedade* with two articles. These three journals are some of the most im-

portant Brazilian journals of Information Science and they cover IR and KO areas. Moreover, they are classified in Qualis stratum A1 of CAPES (Coordination for the Improvement of Higher Education Personnel), one of the factors justifying their indexing in both databases, because as WoS and Scopus, CAPES Qualis also uses strict quality criteria for journals.

Another discussed item was the type of co-authorship that comprise the corpus of this research. Most articles (sixty nine) were published in multiple authorship and 31 in single authorship, as shown in Figure 2. It is possible to note that co-authored publications prevail, and this scenario can be understood from the observation made by Beaver (2004) that co-authorship reflects the generation and exchange of new and current knowledge with which greater authority from the epistemological point of view is achieved, highlighting the solution of common problems.

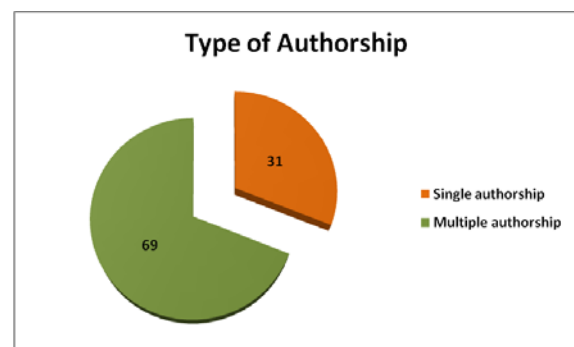


Figure 2: Type of authorship

The articles that make up the corpus of the investigation, consisting of 100 articles, were classified into three categories, the three dimensions of KO used by ISKO-Brazil as sub-themes of the Brazilian chapter congresses. After reading and analyzing the content of the titles, abstracts and keywords, most of the articles were categorized in the Applied Dimension of Knowledge Organization, represented by 68 articles (68%). We found 27 articles (27%) related to the Epistemological dimension and only 5 articles (5%) were categorized in Social and Political Dimension. (Figure 3).

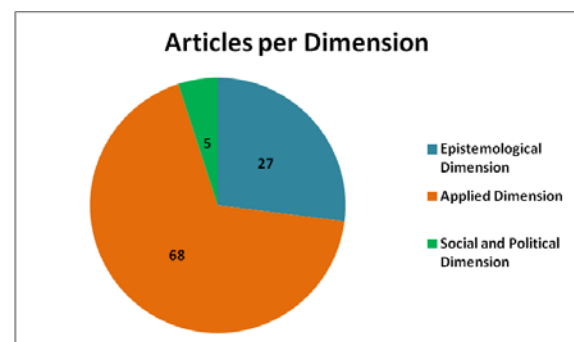


Figure 3: Dimensions of Knowledge Organization

The Epistemological Dimension of KO represents the studies on: conceptual bases; historical bases; methodological bases and interdisciplinary dialogues of knowledge organization. In this dimension, most studies are on conceptual bases of knowledge organization with 18 articles (18%) and secondly on interdisciplinary dialogues of KO (5 articles), as demonstrated in Figure 4. Studies related to this dimension are key to discuss theory that, consequently, shall underpin the practice represented in the applied dimension.

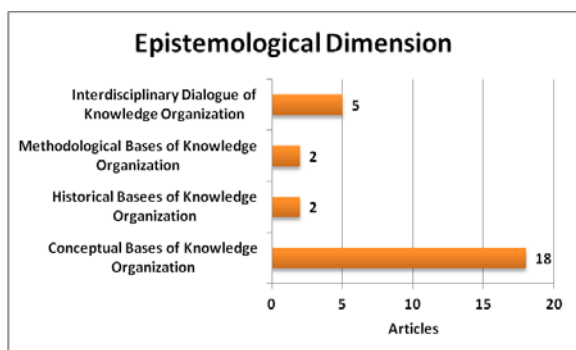


Figure 4: *Epistemological Dimension*

As stated before, the applied dimension stands out among the other categories. This dimension discusses studies of: KO models and formats; KO instruments; KO products and; structures in KO. The research related to knowledge organization instruments were the majority with 26 articles (26%) categorized in the Applied Dimension. The studies on structures in KO have also been highlighted with 21 articles (21%). (Figure 5).

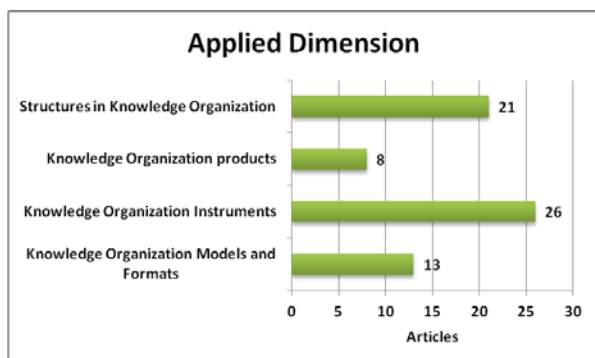


Figure 5: *Applied Dimension*

The predominance of research in the applied dimension can be explained by the characteristics of the corpus, considering that IR is a process involving application, tests, etc. This does not mean that studies in the epistemological dimension are not necessary, as recognizing and discussing concepts, addressing the history and methods of KO and IR are critical issues.

The Social and Political Dimension includes research on: training and professional practice in KO; ethics in KO; culture and identity in knowledge organization; ad KO and sustainable development. Among the five arti-

cles classified in this dimension, four are related to training and professional practice in KO and one article regarding culture and identity in KO.

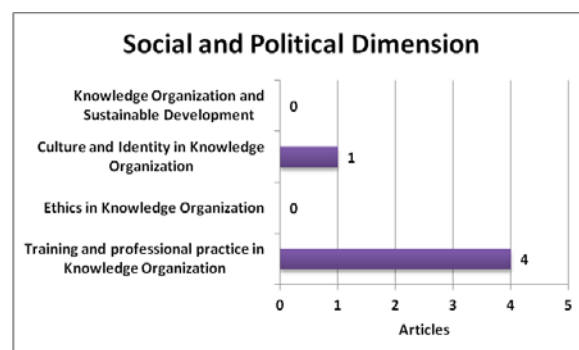


Figure 6: *Social and Political Dimension*

Studies in Social and Political Dimension of KO are also fundamental. However, the low incidence of articles with this theme can be explained because the corpus is formed by international articles, and, the Social and Political Dimension of KO was established as a sub-theme of the ISKO-Brazil congress because of a national yearning, since the first edition of the congress.

5 Conclusions

These results confirm the predominance of research related to the applied dimension of KO. This same result was obtained by Bufrem (2015) when she analyzed KO thematic representation in the literature of IS journals present in BRAPCI (Reference Database of Articles in Information Science Journals) and published by CNPq (National Council for Scientific and Technological Development) researchers named PQ1.

Many studies classified in the applied dimension in this investigation are related to Simple Knowledge Organization Systems (SKOS), to automatic indexing, semantic web and ontologies. These themes can be identified as some research trends regarding the relationship of the areas of IR and KO, once they stood out in the corpus referring to articles published in the past decade in the international literature.

It is suggested, for further research, the analysis of the social network formed by the authors of this corpus, as well as citation analysis, co-citation and bibliographic coupling, which provide a deeper look in these domains and their relationships. This type of investigation enables the identification of strengths, weaknesses, epistemic communities, research trends, methodological options and the comparison with the theoretical references used.

It is also highlighted that research related to social and political dimensions need to be encouraged, since this dimension was poorly representation in this research. This dimension involves important issues such as pro-

professional training, ethics, culture, identity and sustainability. In the research conducted by Bufrem (2015), professional training was the most significant issue in the national literature, followed by the theme on ethics in KO.

Therefore, as suggestion for further research, an analysis in national journal literature related to IR and KO domains could also be explored. Regarding IR and KO in this research, the analysis of the corpus allowed us to infer that there is a close relationship between these two domains in the international journal literature indexed in WoS and Scopus databases and published in the last ten years.

References

- Aghaei Chadegani, Arezoo; Salehi, Hadi; Md Yunus, Melor; Farhadi, Hadi; Fooladi, Masood; Farhadi, Maryam; Ale Ebrahim, Nader (2013). A comparison between two main academic literature collections: Web of Science and Scopus Databases. // *Asian Social Science* 9:5 (2013) 18-26.
- Arboit, Aline E.; Bufrem, Leilah S.; Freitas, Juliana L. (2010). Configuração epistemológica da Ciência da Informação na literatura periódica brasileira por meio de análise de citações (1972-2008). // *Perspectivas em Ciência da Informação*. ISSN 1981-5344. 15:1 (jan./abr. 2010) 18-43.
- Barité, Mario. (2001). Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. // Carrara, K. (Org.). (2001). Educação, universidade e pesquisa. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. 35-60.
- Beaver, Donald D. (2004). Does collaborative research greater epistemic authority? // *Scientometrics* 60:3 (August 2004) 399-408.
- Bufrem, Leilah Santiago (2015). Perspectivas da pesquisa sobre organização do conhecimento no Brasil. // Guimarães, José Augusto Chaves; Dodebei, Vera (2015). Organização do conhecimento e diversidade cultural. Marília: ISKO-Brasil; FUNDEPE, 2015. 709-724. <http://isko-brasil.org.br/wp-content/uploads/2015/09/Organiza%C3%A7%C3%A3o-do-Conhecimento-e-Diversidade-Cultural-ISKO-BRASIL-2015.pdf> (2015-09-20).
- Dahlberg, Ingertraud (1993). Knowledge organization: its scope and possibilities. // *Knowledge Organization* 20:4 (1993) 211-222.
- Esteban Navarro, M. A.; García Marco, F. J. (1995). Las primeras jornadas sobre organización del conocimiento: organización del conocimiento e información científica. // *Scire* 1:1 (jan./jun. 1995) 149-157.
- Ferneda, Edberto. (2013). Recuperação de Informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. São Paulo: USP, 2003. Doctoral thesis.
- Fujino, Asa; Noronha, Daisy Pires; Aguiar Población, Dinah; Silva, José Fernando Modesto da (2007). Comunicação e produção científica: avaliação e perspectivas. // Lara, Marilda Lopes Ginez de; Fujino, Asa; Noronha, Daisy Pires. Informação e contemporaneidade: perspectiva. Recife: Néctar, 2007. 205-231.
- Guimarães, José Augusto Chaves (2008). A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). // *Revista Ibero-americana de Ciência da Informação (RICI)* 1:1 (2008) 77-99.
- Guimarães, José Augusto Chaves (1990). Recuperação temática da informação. // *Revista Brasileira de Biblioteconomia e Documentação* 23:1/4 (jan./dez. 1990) 112-130.
- Guimarães, José Augusto Chaves; Sales, Rodrigo de (2010). Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. // *DataGramaZero: Revista de Ciência da Informação* 11:1 (Fev. 2010).
- Hjørland, Birger (2002). Domain analysis in information science: eleven approaches traditional as well as innovative. // *Journal of Documentation*, 58:4 (January 2002) 422-462.
- Hjørland, Birger. (2008). What is Knowledge Organization (KO)? // *Knowledge Organization* 35: 2-3 (January 2008) 86-101.
- Hjørland, Birger. (2013). Theories of knowledge organization: theories of knowledge. // *Knowledge Organization* 40:3 (2013) 169-198.
- Hjørland, Birger; Albrechtsen, Hanne. (1995). Toward a New Horizon in Information Science: domain-analysis. // *Journal of the American Society for Information Science* 46: 6 (July 1995) 400-425.
- Jaenecke, Peter. (1994). To what end knowledge organization? // *Knowledge Organization* 21:1 (1994) 3-11.
- Khapre, Shailesh; Basha, M. S. Saleem. (2012). Theoretical paradigm of information retrieval in Information Science and Computer Science. *International Journal of Computer Science Issues* 9:5 (September 2012) 232-240.
- Marchionini, Gary. (1995). *Information Seeking in Electronic Environments*, Cambridge, UK: Cambridge University Press, 1995.
- Mooers, C. (1951) Zatoncoding applied to mechanical organization of knowledge. // *American Documentation* 2:1 (1951)20-32.
- Rendón Rojas Miguel Ángel. (2008). La ciencia de la información en el contexto de las ciencias sociales y humanas: ontología, epistemología, metodología e interdisciplina. // *DataGramaZero: revista de ciência da informação*, 9:4 (August 2008). <http://www.datagramazero.org.br/ago08/F_I_art.htm>. (2015-09-10)
- Sales, Rodrigo de; Guimarães, José Augusto Chaves; Oliveira, Ely Francina Tannuri; Bufrem, Leilah Santiago. (2011). Redes sociais em linguagens documentais: uma análise de coautoria a partir da realidade brasileira. //

Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação 16:31 (2011) 1–24.

Saracevic, Tefko; Kantor, Paul; Chamis, Alice Y.; Trivison, Donna. (1988). A study of information seeking and retrieving: background and methodology. // *Journal of the American Society for Information Science* 39:3 (1988) 161-176.

Smiraglia, Richard P. (2002). The progress of theory in knowledge organization. // *Library Trends* 50:3 (Winter 2002) 330-349.

Wilson, T.D. (2000), *Human Information Behavior*. // *Information Science* 3:2 (2000) 49-55.

Copyright: © 2016 Araújo (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: 2016-05-29. Accepted: 2016-05-31