ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Knowledge Organization as a domain: Criteria and decisions for the construction of a thesaurus-dictionary of the field

Mário Barité

Universidad de la República, Uruguay

Rita do Carmo Ferreira Laipelt

Universidade Federal do Rio Grande do Sul, Brasil.

Patricia Daniela Petroccelli Rodríguez

Universidad de la República, Uruguay

Rene Faustino Gabriel Junior

Universidade Federal do Rio Grande do Sul, Brasil.

Abstract

The KO field only has two KOS dedicated to the thematic representation of the area for classification and indexing purposes, for information and documentation retrieval: a thesaurus-glossary and a classification system. A critical review seems to indicate that KO needs new, improved and updated KOS that are specific to its field in terms of both scope and quality. The main objective of this work is to describe the methodology currently used to construct a thesaurus-dictionary specialized in KO, based on the domain's theoretical, historical, and structural characteristics. This project is under the responsibility of a binational research group. Seeking prior expert consultation from researchers worldwide, with the aim of achieving the most consensual final product possible, is one distinctive aspect expected to be established as a prerequisite for publication of the thesaurus-dictionary.

1.0 Introduction

A possible indicator of a field's maturity is the existence of reference works and Knowledge Organization Systems (KOS) specialized in it, or in some of its subfields, or in its areas of intersection with other domains.

As Hjørland (2002) observes, the Knowledge Organization (KO) literature on classification is limited in both quantity and methodological scope. Most studies focus on universal classification schemes, while relatively little attention has been given to special domains, creating difficulties for updating and revising even universal systems. Another issue is the contrast between the small number of studies on special classifications and the large number of existing systems, many of which—such as MEDLINE and PsycINFO—were developed ad hoc and independently of Information Science methodologies. As Hjørland

(2002, 425) notes, "thesauri are mainly domain specific vocabularies, and the methodology of designing them can also be seen as one (implicit) form of domain analysis". The literature on thesauri is somewhat more extensive in methodological terms, with faceted analysis predominating, although approaches based on natural language processing (NLP), artificial intelligence (AI), and citation analyses are also employed in their construction.

Currently, the Knowledge Organization (KO) field has only two Knowledge Organization Systems (KOS) specifically dedicated to the thematic representation of the domain for classification and indexing in information retrieval. The first is a thesaurus-glossary (Dégez and Ménillet 2001). Although pioneering, it has several limitations: it is little known, published only in French, lacks critical reviews, and is not available online, which limits its international recognition and use. The second is the Classification System for Knowledge Organization Literature (CSKOL), created by Ingetraut Dahlberg to organize the *Knowledge Organization Literature* section of the journal *International Classification*. This publication ran from 1974 to 1992 and has since been continued by *Knowledge Organization*. Three editions of CSKOL were published: the first as *Classification Literature Classification* (CLC) in volume 12, issue 3 (1985) of *International Classification*; the second in volume 20, issue 4 (1993) of *Knowledge Organization*; and the third in volume 26, issue 4 (1999). Subsequent editors introduced minor modifications that were not publicly documented and whose criteria remain unknown.

The weaknesses of CSKOL have been listed in a 2020 paper: a) "The classification criteria used by the different editors have not been adequately spread. b) There is no known protocol available establishing the guidelines for the revision of the schemes and their periodic update. c) It is necessary to add auxiliary tables (place, time, form, and others), or to expressly enable the combination with the corresponding auxiliary tables of other systems (DDC, UDC, etc.). d) The lack of a general index of the system makes specific searches by subject difficult. It is an important limitation for the analysis of the terminological coverage of the area and the detection of gaps. e) The warrant used to justify the selection of the terminology is unknown" (Barité and Rauch 2020, 462-463). Furthermore, given that the third and final edition of CSKOL dates from 1999, it is highly likely that the system is outdated, especially in areas of increasing importance in KO.

KO also has an encyclopedia (Hjørland and Gnoli 2017-); two dictionaries (Satija 2004; Barité et al. 2015); and several dozen glossaries containing hundreds of terms, which can eventually be used as descriptors.

However, KO is an area that is directly related to the systematization and evolution of knowledge, as well as the conceptual and terminological structuring of all scientific, technological, social and cultural domains. Therefore, it should have new, improved and updated KOSs that are specific to its field in terms of both scope and quality. This prompted the research teams of two universities (one Brazilian and one Uruguayan) to create a new KOS specializing in KO that can incorporate the theoretical and methodological advances the field has seen this century.

Another justification for creating a new KOS for the area can be found in Dextre Clarke's (2019) text, in which the author observes confusion surrounding the term 'thesaurus' and its

definition in KO. According to Dextre Clarke, it has been possible since the 1990s to identify the indiscriminate use of the term to designate any type of KOS. He explains that national and international thesaurus standards are presented as guidelines, not mandatory instructions. Consequently, thesauri are often developed that adhere only to some of these guidelines. Another factor is the misuse of the term by people with little or no specific training in KO who, when developing a KOS, randomly name it according to current trends. Thus, two situations can be identified: KOS that are not thesauri being named as such, and KOS that could be considered thesauri receiving other names.

In a more general analysis based on our research corpus definitions, we identified another difficulty: terms derived from 'thesaurus' are dispersed across reference works, with varying occurrences, definitions, and scopes. We identified 3 databases (Circular thesaurus, Corporate thesaurus, Master thesaurus), 1 dictionary (Roget's Thesaurus), and 18 thesauri. These thesauri are named according to specific characteristics that authors wanted to highlight, such as: language (Monolingual thesaurus, Multilingual thesaurus); size/scope (Macro-thesaurus, Micro-thesaurus, Roof thesaurus, Sub-thesaurus, Meta-thesaurus, Satellite thesaurus, Specialized thesaurus); audience/purpose (End-user thesaurus, Indexer thesaurus, Search thesaurus); methodology (Faceted thesaurus, Thesaurofacet, Terminological thesaurus, Poly-hierarchical thesaurus); and presentation (Electronic thesaurus, Systematic display thesaurus).

The binational research team aims to create a thesaurus-dictionary for the KO area. This will not simply be a hybrid tool combining the two conceptual structures, but will enrich the final product with additional functionalities and content that surpass individual thesauri and dictionaries.

The project also aims to effectively address the semantic enrichment described in the (International Organization for Standardization 2011, 11) standard, which defines a thesaurus as "[...] a controlled and structured vocabulary in which concepts are represented by terms, organized in such a way that the relationships between the concepts are made explicit, and the preferred terms are accompanied by introductory entries for synonyms or near-synonyms."

For their part, dictionaries remain important reference works for resolving various issues in the interpretation and understanding of concepts. In libraries, dictionaries constitute a particularly carefully preserved collection, to the point that they are a substantial part of the so-called reference services. They have reinvented themselves as a genre, becoming a type of work that has adapted very well to technological changes.

There are other works that respond to the structural and data presentation forms of dictionaries: glossaries, vocabularies, or lexicons. The differences lie in the size of the works, the greater or lesser number of terms included in their definition, or, as is the case with many glossaries, the attribution to a larger work that contains expressions that need to be explained. Specialized dictionaries generally have two origins: either they are editorial projects of a certain magnitude, or they arise from individual endeavors by specialists with basic knowledge of lexicography and terminology rules. In both cases, they are works that

provide "quick access to information, or sources of information on a subject" (Cunha and Cavalcanti 2008, 266).

The advantages of adding definitions to thesauri have been noted by various authors, particularly Buchan (1989) and Hudon (1996, 1998). Buchan proposes two ways to address the overlap between thesauri and dictionaries: linking thesauri to widely recognized dictionaries—generally general-language ones—or combining definitions from different dictionaries, provided that terminology is standardized to suit the structure and requirements of a thesaurus. He also recommends using plural forms for many terms, since dictionaries define words in the singular while thesauri often adopt plural descriptors. In addition, thesauri may select a synonym different from that used in dictionaries, and synonym dictionaries should assign only one meaning per term (Buchan 1989, 174).

For the thesaurus-dictionary under development, the research team will create original definitions to ensure greater notional coherence within the KOS.

2.0 Objectives

The main objective of this work is to describe the methodology currently used to construct a thesaurus-dictionary specialized in KO, based on the domain's theoretical, historical, and structural characteristics. To this end, techniques and methods specific to the construction of controlled vocabularies and terminology are considered, with direct reference to the domain analysis postulates related to “Constructing special classifications and thesauri” (Hjørland 2002).

In addition, the aim is to provide an update on the project's progress, describing the methodological phases that have already been completed and those under development.

3.0 Methodology

The construction of a thesaurus-dictionary involves a predominantly qualitative research process, in which the relevance, specificity, and quality of the selected terminology, as well as the relationships between the terms, are assessed.

The following methodological phases were established:

Phase 1: Determination of the final product. The goal is to create a digital thesaurus-dictionary in English for international recognition. It will provide equivalent terms in Portuguese, Spanish, and other languages. The resource will cover traditional relationships between terms (synonyms, hierarchical, and associative) as well as broader associative semantic relationships. The thesaurus-dictionary will focus only on KO. It will exclude terms specific to related fields, such as Information Science, Information Retrieval, Terminology, Translation, and Natural Language Processing. Only the terms that name these related fields are included. To define the field of KO, we adopted Hjørland's strict operational definition: “a field of study concerned with the nature and quality of such knowledge

organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations, works and concepts" (Hjørland 2008, 86). This definition includes the foundations, theories, methodologies, institutions, and roles directly involved in the domain.

Phase 2: Formation of the corpus. The corpus was defined as consisting exclusively of specialized reference works on KO published worldwide, with no chronological or linguistic limitations: encyclopedias, dictionaries, glossaries, vocabularies, and similar works. As regards glossaries in particular, only autonomous glossaries (i.e., without reference to a larger work) are included, except when it comes from standards, technical norms, and well-known classification systems.

Works covering related areas, such as information retrieval or digital information, were included if KO terminology was prevalent (e.g., Bonotto et al. 2007). Works specializing in KO sub-branches, such as indexing or controlled vocabularies (e.g., Buchanan 1976), were also included.

At this stage, the following were excluded from the corpus: a) glossaries or vocabularies inserted within larger works such as monographs, manuals, these and similar texts, as they are numerous and secondary contributions, which in many cases have limited terminological coverage, whereby do not provide new information to the research, and, b) reference works whose content could not be accessed (particularly Riggs 1982).

The contextual information that makes up the initial corpus was published between 1958 and 2017, and in different languages (English, French, Portuguese, Italian and Spanish), which, on the one hand, demonstrates the regularity in the publication of this type of work, and on the other hand, anticipates the advantages and difficulties that definitions presented in five different languages can offer. On the other hand, the fact that the works comprising the corpus have been published in North and South America, Europe, and Asia also contributes to the internationalization and standardization of terminology in the area. The chronological delimitation (1958–2017) then establishes the publication dates of the first and last specialized reference work on KO to date, although the last one (2017) continues to expand its contents.

At a future stage, the corpus will be expanded to include works focused on a single term (e.g., "taxonomy"; Gilchrist 2003; Bhat and Shafi 2014). In this way, the thesaurus-dictionary will be guided entirely by a combination of literary and terminological warrants to validate the selection of terms, and will be able to offer guidance on scientific articles focused on specific terms, which are best regarded in the field. Literary warrant is the methodological guideline that justifies the selection of terminology based on the existence of a body of literature specific to a topic or subject (Bullard 2017). For its part, terminological warrant is based on the terminology of a subject field, as it is mostly recognized by specialists (Beghtol 1986, 113).

It should also be noted that each work was framed within its category (encyclopedia, dictionary, glossary, thesaurus-glossary), as indicated in the title by its authors or those

responsible. Additionally, some of these works, published prior to 1989, refer to the field as Classification, the field's name before that year.

Phase 3. Compilation of terminology. The selection of terms for the thesaurus-dictionary is carried out in three steps:

(i) Critically analyze the entry terms recorded in each work of the corpus, both with and without definitions. Confirm that each term actually belongs to KO, and not to a related area such as Information Retrieval, Linguistics, Terminology, or Informatics.

(ii) identification of the number of different records of each term in the corpus (occurrence indicator);

(iii) Complementarily, confirmation in Google Scholar of enough documents that record the term as a main or significant topic and provide information about it. In this research, finding 50 different KO documents in which the term is mentioned, defined, or explained— regardless of whether it is the main topic— seemed a reasonable measure to ensure the existence, validity, and knowledge of the terms among KO specialists.

(iv) Manual extraction of terms, their definitions and their data, because many of the works in the corpus are only available in printed version. A preliminary analysis was conducted of possible relationships between terms, including synonyms, hierarchies, and associations. Ambiguities and inconsistencies were marked for further, more in-depth analysis.

Phase 4: Construction of the domain tree. A domain tree is a type of taxonomy used in terminological research that displays the main branches of a domain or area of knowledge, with the aim of assigning terms to these branches that represent the domain. It is also a useful tool for identifying semantic affinities, identifying terminological gaps, and establishing conceptual boundaries of the domain (Barité 2017, 91ff.). For the elaboration of the KO domain tree, a bottom-up approach (an inductive method) was followed, starting from the universe of collected terms and grouping them into branches based on their semantic affinities.

Phase 5: Identification of relationships between terms. In the first stage, the thesaurus-dictionary will include synonymy, antonymy, hierarchical, and associative relationships. In the second stage, additional semantic relationships will be identified and incorporated to expand associative relationships and enrich the domain's conceptual framework.

Phase 6: Development of original definitions in English.

Phase 7: Adaptation of Thesa Software. The Thesa system was designed according to the specific needs of the research. Thesa is open-source software developed by researchers in the field of Organizational Communication that enables the creation and management of dictionaries, glossaries, controlled vocabularies, and thesauri. Its design is based on the international standards ISO 25964 (parts 1 and 2) and ANSI/NISO Z39.50, using the PHP language and MySQL. This infrastructure enables the sharing and collaborative development of the tool in a web environment. Among its main differentiators is the ability to build semantically rich, multilingual thesauri based on the SKOS ontology, with

descriptions of specific relationships and the ability to reuse other vocabularies. These characteristics favor the sharing of structures and data in complex systems, such as Linked Data (Gabriel Junior and Laipelt 2017, 2023). The adaptations made to the software for the thesaurus-dictionary consist of improving the existing terminological record and creating new typologies of associative semantic relationships, according to the domain's needs.

Phase 8: Begin by inserting information into the Thesa, setting the foundation for subsequent development phases.

Phase 9: After insertion, release a preliminary version of the thesaurus-dictionary for closed expert evaluation, followed by open review for worldwide expert feedback.

Phase 10: Based on feedback received during Phase 9, make content adjustments informed by external evaluations.

Phase 11: Following revisions, make the final version of the dictionary-thesaurus available.

4.0 Results

At present, phases 1 and 2 have been completed, and we are working on phases 3, 4, and 7. Therefore, the results of these five phases will be described below.

Phase 2. Corpus. The literature review identified a corpus of twenty-five (25) sources, distributed by type of reference work:

- encyclopedias: 1 (Hjørland and Gnoli, Eds. 2017-);
- dictionaries: 2 (Satija 2004; Barité et al. 2015);
- autonomous glossaries: 8 (Ranganathan 1958; Buchanan 1976; Barité 1997; Wellisch 2000; Gnoli et al. 2006; Bonotto et al. 2007; Will 2008; Harpring 2010);
- glossaries included in standards, technical standards, and classification systems: 13 (ISI 1964; American National Standards Institution, 2005a, 2005b, 2010a, 2010b; British Standards Institution, 2005, 2007; Association for Library Collections and Technical Services, 2006; International Organization for Standardization, 2011, 2013; International Federation of Library Associations and Institutions, 2011; Bliss 2016). For the glossaries from the British Standards Institution, the International Organization for Standardization, and the American National Standards Institute, only the most recent editions have been included in the corpus at this stage; earlier editions will be reviewed in a subsequent phase.
- 1 thesaurus-glossary (Dégez and Ménillet 2001).

Phase 3. Compilation of terminology. If 402 terms have been registered, 357 of them are considered descriptors (which will contain definitions), and the other 45 are considered non-descriptors, which will not have definitions and will only contain a reference value. It is assumed that the record of new terms will not be of such magnitude that it exceeds a final number of 550.

Phase 4: Construction of the domain tree. The tree proved to be hospitable, absorbing all the terms registered so far into its branches, although it was also necessary to establish

sub-branches, which are currently under study. Each branch of the tree received a note indicating whether the term was included or excluded. For example, 3.1. Principles of KOS. Inclusion note: includes terms related to the theoretical and methodological principles applied in the construction of KOS.

The tree now has its final form and has 6 main branches:

1. Knowledge Organization in general (includes fundamentals and theories; basic concepts; related fields). Terms currently included: 39 (33 descriptors + 6 non-descriptors).

2. KO Processes (includes concepts, techniques, and methods; classification; indexing; abstracting; coding; ordering). Terms currently included: 76 (66 descriptors + 10 non-descriptors).

3. Knowledge organization systems (includes principles; typologies; structure, parts, and elements of KOS; representation). Terms currently included: 257 (229 descriptors + 28 non-descriptors).

4. Methods. Terms currently included: 17 (16 descriptors + 1 non-descriptor).

5. Persons related to KOP and KOS. Terms currently included: 10 (all descriptors).

6. Institutions related to KO. Terms currently included: 3 (all descriptors).

The first three branches have been subdivided into more specific branches.

The main findings that can be established in this first stage of development of the thesaurus-dictionary can be summarized in four points: (i) it is confirmed that the most productive sub-areas in terms of concepts and terms correspond to processes and systems of knowledge organization, in accordance with the axes anticipated by Hjørland (2008); (ii) it can be anticipated that the specific terminological core of the field of KO comprises between 500 and 550 terms; (iii) this core comprises, for the most part, terms that have been consolidated over several decades, generally associated with theoretical bodies such as faceted theory or vocabulary control; (iv) the acceptance of new terminology is gradual and somewhat slow, and is concentrated on the incorporation of new approaches (domain analysis) or new types of systems (ontologies, gazetteers).

5.0 Final considerations

From a documentary point of view, every moderately developed and consolidated area of knowledge needs—in addition to specialized journals, conference proceedings, manuals, working papers, codes of ethics, educational and outreach texts—reference works such as dictionaries, glossaries, and vocabularies, as well as specialized KOS that allow for the organization of its disciplinary production. A possible indicator of a field's maturity lies in identifying these reference works and their thematic representation.

KO appears to be at an interesting point of evolution and maturity, given the significant number of reference works that have exclusively and predominantly covered the field since

the late 1950s. However, the need remains to have at least one classification tool and one indexing tool specialized in KO. These tools, in addition to receiving regular updates to ensure their validity and continuity, should be built following the most solid theoretical foundations, principles, and methodologies specific to the area, with corresponding explanations.

In this sense, building a specialized thesaurus-dictionary in KO represents both a challenge and an opportunity for users in the field. It can contribute a hybrid reference and indexing work with three significant advantages for promoting its use and dissemination internationally: it will be published in English, be digital, and have a binational university team to manage and update both the work's content and the original software that supports it.

Furthermore, the aspiration to develop a tool that includes semantic and relational enrichment features allows us to move beyond the idea of a work that merely combines the elements of a thesaurus with those of a dictionary, without any interrelationships. Seeking prior expert consultation from researchers worldwide in the field of knowledge sharing to achieve the most consensual final product possible is another distinctive aspect expected to be established as a prerequisite for publication.

The complete bibliography used to construct the corpus is available in the replication dataset deposited in CedapDados (Barité et al. 2026).

We did not use AI to create this work.

References

Barité, Mario. 2017. “Los árboles de dominio.” In Teoría y praxis en Terminología, edited by Sara Álvarez Catalá and Mario Barité, 91–102. Montevideo: Comisión Sectorial de Investigación Científica, Universidad de la República.

Barité, Mário, Rita do Carmo Ferreira Laipelt, Patricia Daniela Petroccelli Rodríguez, and Rene Faustino Gabriel Junior. 2026. Dados de replicação para: Knowledge organization as a domain. Version 1. Dataset. CedapDados. doi:10.48472/CEDAP/RQWGPA.

Barité, Mario, and Mirtha Rauch. 2020. “Classification System for Knowledge Organization Literature (CSKOL): Its Update, a Pending Task?” In Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference, edited by Marianne Lykke, Tanja Svarre, Mette Skov, and Daniel Martínez-Ávila, 460–464. Advances in Knowledge Organization 17. Würzburg: Ergon Verlag. doi:10.5771/9783956507762-460.

Beghtol, Clare. 1986. “Semantic Validity: Concepts of Warrant in Bibliographic Classification Systems.” Library Resources & Technical Services 30: 109–125.

Bhat, Mohammad Hanief, and Sheikh Mohammad Shafi. 2014. “Taxonomies in Knowledge Organisation: Need, Description and Benefits.” Annals of Library and Information Studies 61: 102–111.

Buchan, Ronald L. 1989. “Intertwining Thesauri and Dictionaries.” Information Services & Use 9: 171–175. doi:10.3233/ISU-1989-9307.

Bullard, Julia. 2017. “Warrant as a Means to Study Classification System Design.” Journal of Documentation 73: 75–90. doi:10.1108/JD-06-2016-0074.

Cunha, Murilo Bastos da, and Cordélia Robalinho de Oliveira Cavalcanti. 2008. Dicionário de Biblioteconomia e Arquivologia. Brasília: Briquet de Lemos/Livros.

Dextre Clarke, Stella G. 2019. “The Information Retrieval Thesaurus.” Knowledge Organization 46: 439–459. doi:10.5771/0943-7444-2019-6-438.

Gabriel Junior, Rene Faustino, and Rita do Carmo Ferreira Laipelt. 2017. “THESA: Ferramenta para construção de tesauro semântico aplicado interoperável.” P2P e Inovação 3, no. 2: 124–145. doi:10.21728/p2p.2017v3n2e-71424.

Gabriel Junior, Rene Faustino, and Rita do Carmo Ferreira Laipelt. 2023. “Thesa 2.0: a interação dos tesauros em tempos de dados lincados.” In Organização e representação do conhecimento em diferentes contextos: desafios e perspectivas na era da datificação, edited by Natália Bolfarini Tognoli, Ana Cristina de Albuquerque, and Brígida Maria Nogueira Cervantes, 555–562. Londrina: ISKO-Brasil; PPGCI-UEL.

Gilchrist, Alan. 2003. “Thesauri, Taxonomies and Ontologies – An Etymological Note.” Journal of Documentation 59: 7–18. doi:10.1108/00220410310457984.

Hjørland, Birger. 2002. “Domain Analysis in Information Science: Eleven Approaches – Traditional as Well as Innovative.” Journal of Documentation 58: 422–462. doi:10.1108/00220410210431136.

Hjørland, Birger. 2008. “What Is Knowledge Organization (KO)?” Knowledge Organization 35: 86–101. doi:10.5771/0943-7444-2008-2-3-86.

Hudon, Michèle. 1996. “Preparing Terminological Definitions for Indexing and Retrieval Thesauri: A Model.” In Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, Washington, DC, 15–18 July 1996, edited by Rebecca Green, 363–369. Frankfurt/Main: Indeks Verlag.

Hudon, Michèle. 1998. “The Usefulness of Standardized Definitions in Thesauri: An Assessment through Interindexer Consistency Measurements.” In Proceedings of the 26th Annual Conference of the Canadian Association for Information Science, edited by E. G. Toms, D. G. Campbell, and J. Dunn, 221–230. Toronto: Canadian Association for Information Science.

International Organization for Standardization. 2011. ISO 25964-1:2011. Information and Documentation — Thesauri and Interoperability with Other Vocabularies — Part 1: Thesauri for Information Retrieval. Geneva: ISO.

International Organization for Standardization. 2013. ISO 25964-2:2013. Information and Documentation — Thesauri and Interoperability with Other Vocabularies — Part 2: Interoperability with Other Vocabularies. Geneva: ISO.

Riggs, Fred Warren. 1982. Indexing Glossary: A New Rationale. FID/CR Report no. 20. FID Publication 405. Frankfurt/Main: Indeks Verlag.