A Glossary for Knowledge Organization Systems Terminology†

Benildes Coura Moreira dos Santos Maculan*, Maristela Sanches Lima Mesquita** and Luander Cipriano de Jesus Falcão***

Universidade Federal de Minas Gerais (UFMG), Av. Antônio Carlos, 6.627,
Belo Horizonte, MG, 31270-901, Brazil
* <benildes@gmail.com>/** <maristela.mesquita@gmail.com>/
*** <luanderfalcao@yahoo.com.br>

Benildes Coura Moreira dos Santos Maculan is a professor at the Escola de Ciência da Informação (Information Science School) of Universidade Federal de Minas Gerais (UFMG, Federal University of Minas Gerais), Brazil. She holds a master's degree and a Ph.D. in Information Science from UFMG (Brazil). Her interests include knowledge organization theories, methods of construction of knowledge organization systems (KOS), terminology, and semantic interoperability.

Maristela Sanches Lima Mesquita is Bachelor in Librarianship and Master student in Information Science at UFMG. She is a specialist in Systems and Network Support Administration from UNI-BH, and also in Homeopathy Science from the Universidade Federal de Viçosa. She is a Coordinator of the Extension project Biblioterapia, and of the project Online Integrated System of Catalographic Sheets at UFOP.

Luander Cipriano de Jesus Falcão is a PhD student in Information Science at UFMG with focus on NLP studies. He holds a master's degree in business administration and graduated in Economics, and MBA in Finance and Controlling. He has been working for over ten years as a Data Analyst, translating business needs into report and dashboard assignments. Specialisms in business vision, value generation, teamwork, strategic planning, systemic vision, and good relationship skills complete the profile.

Maculan, Benildes Coura Moreira dos Santos, Maristela Sanches Lima Mesquita and Luander Cipriano de Jesus Falcão. 2023. "A Glossary for Knowledge Organization Systems Terminology." *Knowledge Organization* 50(1): 23-33. 32 references. DOI:10.5771/0943-7444-2023-1-23.

Abstract: This paper describes the creation of a terminological glossary for the subject domain of knowledge organization systems (KOS). The glossary addresses the need for unambiguous conceptual information about the subject matter for students majoring in information science. It aims to create a glossary of definitions for concepts of different types of KOS. The theoretical and methodological foundations are based on corpus linguistics for the construction of the glossary. The methodology is characterized as an exploratory, descriptive, and applied study, using the *ISKO Encyclopedia of Knowledge Organization* (IEKO) as a source for data collection and terminology. The cut-off was based on the KOSs listed in NISO Z39.19 (2005; R2010) and ISO 25964 (2011; 2013). It is hoped that the terminological glossary created will improve students' understanding of the different KOSs, by compiling definitions and giving them greater semantics.

Received: 18 April 2023; Accepted 23 April 2023

Keywords: terminology glossary, knowledge organization systems, knowledge representation.

† This paper was selected as one of the two best papers at the Fifth Spanish-Portuguese ISKO Conference Lisbon, Portugal, November 25-26, 2021.

1.0 Introduction

The term knowledge organization systems (KOS) is a generic term that names a range of tools such as glossaries, classifications, taxonomies, thesaurus, ontologies, and others. In the scope of information science, they are tools used for the representation of information of a given domain. Their purpose is to organize knowledge and encode it in a language readable not only by humans, but also by computer systems. KOS are characterized by different structures and specific functions, with the common purpose of supporting the organization of information and facilitating its management and retrieval (Mazzocchi 2018). They also provide support in a digital world of information that is increasingly diverse and numerous in data, with a tendency for vertiginous growth. This results from multiple resources that have been expanding in a decentralized way, becoming a huge repository of documents, a space that renders retrieval of relevant information almost impractical (Souza and Alvarenga 2004).

Communication without noise, resulting in a shared understanding of terminology and meaning, is key for the retrieval of relevant information for the user, especially for those who seek specialized knowledge. In this context, the glossary of types of KOS stands out because it elucidates definitions from the perspective of their concepts (Morais, Ramalho and Souza 2019) and facilitates semantic interoperability between systems and people, since it allows the standardization of interpretation of meanings of information in a context.

In this study, we created a terminological glossary supported by corpus linguistics to meet the need for information about different types of KOS for students of information science. It is noteworthy that there is a lack of compiled definitions of the different types of KOS that explain their characteristics and functions in a more complete way. This gap has caused ambiguities that sometimes hinder the understanding and learning of these instruments by students. Thus, this study seeks to answer the following question: how can we facilitate the understanding of terms used in information science education, specifically regarding concepts of different types of KOS?

2.0 About glossaries: what are they?

The term glossary originates from the Latin *glossarium*, which means *collection of glosses* (glosses are the annotations within a text for the purpose of elucidation). A glossary is a collection of definitions about a specific subject, which explains words and expressions in a language, a specific field of knowledge, or specific human activity, among others. A glossary often includes the current meanings of expressions or words that have fallen into disuse, but that may still be useful in defining certain concepts or situations in earlier generations. Thus, when the glossary is of high quality, it minimizes the occurrence of misinterpretation. Since the word is contained and explained in the glossary, the reader will know what it means in that context.

In general, glossaries are composed of technical or domain-specific terms that are useful for understanding the subject matter. It should not be confused with a dictionary, which is a reference work or reference framework in which explanations of all words or expressions in each language are found.

According to Krieger and Finatto (2004), different types of glossaries exist for different uses, such as: i) in literary works: to clarify neologisms constructed by the author for which it would not be feasible to illuminate the meanings in the course of the text, since this would significantly affect the reading pace; ii) in academic or scientific works: to facilitate the understanding of the meaning and the identification of the concepts and terms used by the researcher; iii) in translation work: for the explanation and translation of words into another language; iv) in specialty domains: for the explanation of technical terms, commonly used by professionals in specific areas. In this last case, the focus of this study, it can be said that a computer glossary involves concepts related to computers, programs, the Internet, and the like; and that an environmental glossary will contain words pertinent to issues such as environment, sustainability, and recycling, among others.

Specialty domain glossaries are those vocabularies that refer to a set of terms with precise meanings. In general, they are necessary for understanding the knowledge accumulated in a domain. They differ from generic meanings in that specialty terms may have meanings in other contexts, because "to every specialized denomination there underlies a specific epistemological cut and theoretical conception" (Barbosa 1995, 2). Mastering this specialty vocabulary is essential for successful communication in different scientific and/or thematic areas.

The representation of knowledge in different domains is a fertile field of research in information science, and for these studies it seeks input on terminology. As Costa and Silva (2004, 1531) point out, "the learning of a specialty language goes through the command of two indissoluble types of knowledge: the conceptual one and the linguistic one. The relation between them is fundamental for the construction of speeches about different areas of knowledge". Language is the key element in communication between individuals, whether they are specialists or not, and it is essential to apply knowledge of terminology in the development of knowledge organization systems (KOS).

In terminology, a glossary is considered a "repertory of lexical units of a specialty with their respective definitions or other specifications about their meanings" (Krieger and

Finatto 2004, 51). Hodge (2000) and Zeng (2008) consider glossaries as a type of knowledge organization system (KOS) that has simple semantic degree, situated in the list of terms class. In turn, Guarino (2006) considers them a type of light ontology, composed only of classes and instances, without the establishment of functions. McGuinness (2003) defines glossaries as alphabetical lists of terms that carry their definitions with the purpose of eliminating ambiguities (Mazzocchi 2018).

According to Vaz, Oliveira and Pierozzi Junior (2017, 83), glossaries in the past represented a "gathering of notes, previously interlinear (glosses), on the meaning of old or obscure words found in texts", and, currently, they are presented as "an organized set of terms from an area of knowledge and their meanings and definitions". The authors add that, in an even more relevant and current connotation, glossaries are known as "tools for knowledge representation, along with term lists and controlled vocabularies" (83).

2.1 Definitions in specialty terminology glossaries

The etymological sense of the term "define" is to delimit, and in a logical sense, it is to restrict the understanding of a given object: it is to say what a thing is. The definition is the descriptive representation of the concept within a domain, normalizing the form of communication in that field of expertise.

We can start the discussion about definitions in glossaries by bringing the basic definitional precepts that Aristotle set in his work *Posterior Analytics*: a definition needs to provide the essence of what is being defined, a definition should not be circular, a definition should not be negative when it can be positive, and a definition should not be expressed in figurative or obscure language.

In Aristotelian essentialism, it is believed that it is possible to find that which shows the "what" of things, their essence. With this approach, the definition brings the distinction of a thing, providing it with an identity, characterizing it. This type of definition shows the identity and uniqueness of what is being defined, arriving at a real definition of a thing or entity that is already given. If we consider the nominalist conception, the essence is a universal and not a real entity, but rather names, terms, or sounds, among others. Thus, we can state that any kind of definition is an arbitrary element, and the one that best fits a certain purpose should be chosen. According to Engelmann (2006, 26), cited in Gomes and Campos (2019, 40):

> [D]efinitions play an epistemological role rather than an ontological one: they serve to indicate or classify beings and, therefore, posit their existence. In this way, "the act of defining is a construction of selected aspects of the referent performed by the 'look' of a group that acquiesces to its understanding in a given knowledge space.

A definition can be elaborated, for example, from the close genre and the specific difference, the close genre being "the portion of the definition that expresses the general category or class to which the defined entity belongs; [and] the specific difference is the indication of the particularity that distinguishes this entity in relation to others of the same class" (Krieger and Finatto 2004, 93). Applying this approach creates subdivisions that represent hierarchical relationships.

Dahlberg (1983, 20) proposes the elaboration of definitions based on the theoretical contributions of concept theory, which is based on the Aristotelian sense, stating that "a definition is the equivalence between a *definiendum* (the symbol that must be defined) and the *definiens* (how something must be defined) for the delimitation of the understanding of the *definiendum* in any communicative situation".

The definition delimits and fixes the meaning of a concept in a communicative situation and is an element that sets boundaries for the intent of a concept, which is determined by the characteristics listed in the concept definition (Dahlberg 1978).

According to Krieger and Finatto (2004), in specialty terminology glossaries, there are three complementary types of definitions: lexicographic, encyclopedic, and terminological. We distinguish between these three types of definitions below.

1. Lexicographical definition: a lexicographical definition explains meanings to clarify the meaning and usage forms of a given term, listing only the essential semantic features. The lexicographical definition should establish a relationship between the general (genus) and the individual (species), giving rise to hierarchical relationships. This type of constructive definition uses the onomasiological process, which starts from the meaning (or concept) to arrive at the signifier, which is the linguistic expression for the concept it wants to express.

2. Encyclopedic definition: encyclopedic definitions provide broad and diverse information that carries both relevant and irrelevant features in characterizing a concept. This definition is mainly concerned with referential information and the description of things. It is the lexicographic definition and the terminological definition in terms of the detail of the information it provides since it carries detailed information about both words and knowledge of things. The encyclopedic definition may provide important elements for terminology research work.

3. Terminological definition: a terminological definition highlights information that allows a concept to be distinguished within a conceptual system by providing meanings appropriate to a communicative situation and conveying

concepts from a certain domain. It is the expression of a particular piece of specialized knowledge. It is recognized as the one that most deals with technical-scientific terms. It is now generally accepted that a terminology definition should include both lexicographic and encyclopedic elements. Such a definition may not always be expressed by the formula: genus proximate and specific difference. It should preferably include elements of definition by understanding, starting with the next generic concept, and indicating the semantic distinctions and the relationships (logical and ontological) between them. It should also include elements of definition by extension when listing all species (of objects and phenomena) that are at the same level of abstraction or all individual objects that belong to the defined concept.

3.0 Methodology

The methodology is characterized as an exploratory, descriptive, and applied study. The compilation of glossaries is based on corpus linguistics, which is the study of a language based on empirical evidence, obtained from the exploration and analysis of a given corpus (Berber Sardinha 2004), which is a homogeneous set of that language. Tagnin (2005, 21) points out that one can create glossaries employing corpus linguistics as a methodology (corpus-based terminology) or as an approach (corpus-driven terminology).

> For LC [corpus linguistics], a corpus is a collection of texts, necessarily in electronic format, compiled and organized according to criteria dictated by the intended research goal. The electronic format allows these texts to be investigated and analyzed automatically, with the use of specific computational tools.

The author emphasizes that the texts used to compose the corpus should be in natural language, i.e., they should not have been prepared with the intention of a linguistic analysis.

Specialty texts should be seen as content produced by a discourse community, because it is in them that we can verify specific linguistic elements of that domain, which are the result of discourse in professional life. Costa and Silva (2004) state that these texts are produced in an organized way by members of the specialty domain, or those individuals recognized by their peers. For the authors, it is through the texts that the specialists organize ideas, debate issues, and construct and deconstruct knowledge.

Taking this into consideration, the corpus was composed of texts available in the *ISKO Encyclopedia of Knowledge Organization* (*IEKO*), especially those that deal with KOS. The most representative terms and their definitions were extracted from this corpus (Finatto 2002). The identification of the terms and definitory excerpts was carried out by means of specific software, which assists the researcher and reduces execution time, costs, and specialized labor. When terms relevant to the context of this study were found in the *IEKO* texts, the definitory contexts were identified from the citation with the purpose of clarifying the concept and function, or also identifying equivalent terms, in another language or not (De Lucca 2006).

We chose to select a small corpus with fewer than 80,000 words (Berber Sardinha 2000) because this study is a glossary for a specific domain. The selection of texts provides a guide for the development of definitions of the types of knowledge organization systems.

The Freeware Concordance Program (AntConc) was used to identify the list of terms and the definitional excerpts. AntConc was developed by Laurence Anthony, Professor in the Faculty of Science and Engineering at Waseda University, Japan, and it performs computational text analysis, using different tools, automatically extracting all the noun phrases from a corpus. AntConc is a multiplatform tool executable in versions for Windows, Linux, and Macintosh. Its file size is about 4 MB, so it is considered light and does not require installation, which makes it easy to use without the need of an expert in the area.

The terms extracted from the corpus were then analyzed and defined, forming the structural basis of the terminology glossary. The KOS construction web application software TemaTres (available at https://www.vocabularyserver.com/) was used to manage the glossary.

The methodological procedures followed the steps suggested by Krieger and Finatto (2004, 1) work planning; 2) terminological recognition and initial preparation; 3) terms listing and selection; 4) data registration; and 5) final phase, detailed below.

3.1 Work planning

In the planning stage, issues pertinent to the entire glossary development process are raised and resolved, such as: i) issues around the corpus; ii) determination of the glossary's macrostructure and microstructure; iii) elaborating definitions; and iv) elaborating an initial introductory text for the glossary.

Thus, in planning, it is necessary to compose a corpus for terminological analysis, seeking first to determine whether the candidate term belongs to the vocabulary of the domain being studied or whether it is a sub-domain or aspect. ISO 1087 (2000, online) defines a domain as "a part of knowledge whose boundaries are defined according to a particular point of view".

After planning issues related to the corpus, we will move on to the characteristics of the glossary composition, describing characteristics of the macrostructure and microstructure. A glossary is originally a terminographic product.

Prescriptive terminography is the field that studies and creates specialized glossaries for a given domain. The purpose of a glossary is to standardize the terminology required for unambiguous communication. Sager (1990, 57) explains the difference between general language and special languages:

> Unlike in general language, where the arbitrariness of the sign is accepted, the special languages (of subject fields and domains) strive to systematize principles of designation and to name concepts according to prespecified rules or general (terminographical) principles. General language (and therefore general dictionaries) fully exploits polysemy, metaphor, and adjectival determination; genuine word creation is relatively rare. Where it occurs, it is based on the experience of every-day life and thus represents a pre-scientific approach of knowledge. The process of scientific observation and description includes designation of concepts and this in turn involves re-examining the meaning of words, changing designations, and coining new terms. ... Designation in special languages (of subject fields and domains) therefore aims at transparency and consistency (standardization).

The glossary compiles the terminological units of value in a certain domain and communicative situation, and establishes the characteristics of these lexical units in accordance with that situation. The importance of context is once again emphasized.

The development of specialized glossaries is based on the communicative theory of terminology. This theory was created by Cabré (1999) in the late 1990s because of his concern with communicative phenomena in specialty domains. This theory emphasizes context, since Cabré "does not accept the drastic distinction between a terminological unit (term) and a lexical unit of a general language (word). He considers terms to be linguistic units that express technical and scientific concepts, but they are still signs of a natural (general) language, with similar characteristics and properties" (Barros 2004, 57).

Among the basic principles developed for the communicative theory of terminology, which are relevant to consider when building terminology glossaries, we highlight the following:

- language is a system that includes grammar, semantics, and pragmatics;
- terms are lexical units whose meanings are activated by pragmatic situations (knowledge of the world) relative to a type of communication, constituted by form (or denomination) and meaning (or relative content);
- linguistic variation occurs both in specialized knowledge and in specialized texts;
- terminological units may occur at different levels of specialization and may be described at different levels of representation, bringing them closer to the specialty cutoffs;
- over time, the lexical units of specialized texts and discourses become part of natural language;
- specialized meaning units are at once linguistic, cognitive, and communicative;
- the term (or terminological unit) can be from different perspectives (social, linguistic, and cognitive);
- the socio-cultural and linguistic changes in a community influence concept.

As a prescriptive terminological product with a strong communicative character, the macrostructure of the glossary considers the reality of the target audience it serves and includes the following parts: preface, introduction, appendices, bibliography, and the specifications for the form of use and the order of entry. According to Barros (2004, 151), the macrostructure refers to "the internal organization of the work, composed of all the information pertinent to the entries and their organization".

In turn, the microstructure is the terminographic part of the glossary that encompasses the grammatical and lexical information of the terms, such as: grammatical category, definition of the entry, scope and historical notes, context of use, and other data that may be required. It shows the finished entry, whose information is arranged in its respective compositional fields that make up the terminological datasheets (Faulstich 1995, 23) as shown in the following example.

> **Entry** = + term + grammatical category (+ - noun, + - terminological syntagma, + - verb) + - gender + - synonym + - variants + - sources + - areas + - subareas + definition + source + - context + - source + - cross-references + - equivalents + - sources

As with the macro- and microstructure, it is important to establish cross-references in the organization of the entries in the glossary since they will take the user from one place to another in the KOS. You can choose to present the cross-references in a different typeface than the one used in the entry. Cabré (1999, 142) "classifies the references into two types: informative and prescriptive", as shown in Table 1.

Cross-references allow the creation of links between concepts, expanding the semantic level of the representation, reconstructing the semantic field of a concept, and thus preserving semantic coherence.

Informative	Prescriptive
Terms are related for the purpose of expanding their names or conceptualizations. They show relationships within the same semantic field.	A term refers to another term to indicate priority usage or avoidance, or to show alternatives.
They are placed within a context of semantic equivalence or contrast.	These are inserted by virtue of a terminology policy.
Equivalence (synonymy): variants, acronyms and their complete forms and respective abbreviations, term and its scientific name, term and the symbol that represents it.	Synonyms are classified as either primary or secondary.
Contrast or inclusion: antonyms, hyponyms, hyperonyms, and cohyponyms.	

Table 1. Informative *versus* prescriptive remissives (adapted from Cabré 1999).

3.2 Terminological recognition and initial preparation

In this step, the basic principles for glossary construction are established, based on the compilation of the corpus to be worked on. According to Krieger and Finatto (2004, 129), this stage refers to "the recognition of technical or scientific texts and the identification of textual types, whether they are more or less 'specialized' or more or less terminologically dense".

In this step, it is necessary to identify the needs of the target audience and ensure the use of reliable data sources. Thus, the texts used in the corpus should be analyzed to verify whether they deal with the subject matter within the specified context.

In general, the focus of specialty glossaries is on nominal expressions, which have the basic function of naming objects in the world (a section of reality). This is the case of the terminological product developed in this study.

3.3 Term listing and selection

In this step, the set of terms that will compose the glossary are determined. To list the terms, criteria such as frequency of occurrence in the corpus and suggestions from specialists, among others, can be used from statistical (frequency of occurrences), linguistic (morphology, syntactic, semantic and morphosyntactic analyses) or hybrid methods.

In the hybrid method, both the statistical and the linguistic methods are used, so that one complements the other, minimizing the problems arising from the isolated use of each method. The statistical method is criticized for often disregarding terms simply because they do not occur frequently. The linguistic method is criticized for being dependent on linguistic knowledge and labeling tools that can generate errors, since manual work would be costly and slow.

The terms selected from the list of candidate terms should take into consideration thematic relevance and pragmatic pertinence. Thematic relevance refers to the fact that every subject area has a stable core that carries the distinctive features that characterize the individuality of the domain, which is linked to concepts of the cognitive field of that given domain (Krieger and Finatto 2004).

Pragmatic pertinence concerns the need to include terms whose concepts support the understanding of the thematic domain works, playing an important communicational role. It involves the combination of different factors: the purpose of the glossary, the context of use, and the profile of the target audience, among others.

3.4 Data entry

The terminological datasheets that will be used to document the glossary will be prepared in a specific way by the modeler.

It is essential to generate a glossary whose entries can be "defined as a complete and organized record of information about a given term" and that "also includes operational information, such as the name of the person responsible for collection" (Krieger and Finatto 2004, 136).

The glossary should be presented in the form of a list of entries with the data that are essential to the target audience's needs. Compilers should use a terminological datasheets model that allows the recording of all the information in the work as completely as possible, and that will be part of the project documentation. Each project may require a different type of terminological datasheets, depending on the specifics of the project.

3.5 Final phase

This step includes the completion of the glossary's development. The applied part of the glossary, selecting the relevant data for the target audience and purpose, and presenting the final glossary proposal is finalized. The glossary is published by cutting information from the terminology sheets and se-

lecting only the information that is relevant to the target audience.

4.0 Results

The results are presented from the five steps of the procedures.

4.1 Work planning

To solve the problems associated with the corpus formation, the corpus was composed for terminological analysis, to observe morphological (grammatical class), syntactic (function), and discursive (context) aspects. The corpus was composed of the texts from the *IEKO*. To analyze this corpus, we used AntConc 3.5.8, which was the most current version for Windows available on its homepage in November 2021.

The AntConc interface was considered simple and intuitive with different analysis options that could be opened in the same window. In these analyses, it was possible to find the occurrence of a word, in which context it is cited, and how many times it appeared in the corpus (incidence). This allowed the main techniques of corpus linguistics to be performed, such as word frequencies, collocation, concordance, n-gram extraction, and comparison of corpora to any kind of text. Corpus analysis allows one to see patterns of grammatical and other usage and is useful for testing intuitions about texts and/or triangulating results from other digital methods (Barreiros 2017).

To determine the macrostructure and microstructure of the glossary, the macrostructure was organized alphabetically, without subentries, with each term as a new entry. This was done to facilitate the search for information, since the target audience are students in training, and thus are not experts in the field.

The microstructure, which is the internal structure of the entries, was formed with information relevant to the students (target audience), such as the definition of the term, foreign language equivalents, other equivalents (if any), and the acronym of the term (if any). References were also inserted: i) to entry terms, which direct the reader to other entries, which refer to the terms that appear in the definition of the entry term, representing an associative relationship; ii) to superordinate and subordinate terms, representing a hierarchical relationship; and iii) to equivalent terms, representing an equivalence relationship, which "represent an option to expand the pragmatic use of the instrument, since it helps the reader to quickly and objectively retrieve other information on the topic treated" (Marini 2013, 81). It is considered that this takes the glossary with a higher semantic level, characterizing it as a type of knowledge organization systems (KOS).

To develop the definitions for each glossary term, the elements that would compose the definition of the entries were planned, considering that the term must be observed in the context in which it is used (Almeida 2006). Thus, the definitions were prepared based on the excerpts extracted from the corpus, and when necessary, other works were consulted to make the definition more clear, precise, and complete.

In the end, the definition was elaborated encompassing three basic types of elements: i) lexicographic definition, which makes the meanings and their uses explicit, without redundancy; ii) encyclopedic definition, which provides a set of information sufficient to understand the term in the context of use; and iii) terminological definition, which delimits the concept of specialty within the notional system. Figure 1 shows an example.

In this step, an initial introductory text was written and will be updated as changes are made to the glossary. This introductory text presents information that allows the target audience to understand the approach taken in building the glossary, as well as the limits of its representation.

4.2 Terminology recognition and initial preparation

The hybrid method was used, where the statistical method was combined with the linguistic method to make the extraction of terms as complete as possible.

Two basic principles for glossary construction were established:

- meeting the needs of a target audience: in this case, students majoring in information science, and
- using reliable data from the IEKO encyclopedia and ensuring that the texts in the corpus are truly representative of the subject of knowledge organization systems (KOS) types from an information science perspective.

To analyze the corpus using AntConc, it was necessary to convert the texts to .txt. Then, the selected excerpts were compiled, which had a total of less than 80,000 words, which characterizes it as a small corpus.

4.3 Listing and selecting terms

The listing and selection of terms was based on i) their occurrence in the corpus; ii) that the terms were considered a type of knowledge organization systems (KOS) under an information science approach (i.e., an information representation tool); and iii) one of the tools cited in NISO Z39.19 (2005; R2010), which are: term lists, ring of synonyms, taxonomy, and thesaurus; iv) one of the instruments cited in ISO 25964 (2011; 2013), namely: thesaurus, classification system, classification system for records management, tax-

Macrostructure (entry term).
CLASSIFICATION SYSTEMS
Microstructure (term definition)
A type of knowledge organization systems (KOS) of a conceptual nature that can be created in hierarchical or faceted arrangements, with numerical, alphabetical, or mixed notations; they are generally designed as general systems (covering all fields of human knowledge) to represent elements of any domain, whether these elements are a set of objects, subjects, or concepts. They are composed of a set of classes into which a given universe of things is subdivided and grouped in such a way that things that share certain characteristics, considered similar, are gathered into a class, and distinguished from things that do not share these same characteristics. Classes are always formed according to some criterion: order, principle, design, purpose or specific interest, or a combination of these elements, obeying a given ordering in a series, whose sequence is determined according to some logical principle of organization.
Equivalence relationship (cross-reference)
used for: classification schemes
used for: classification schemes
Hierarchical relationship (cross-reference)
generic term (superordinate): knowledge organization systems
specific term (subordinate): bibliographic classification systems; classification systems for records management
Associative relationship (cross-referenced)
associative terms: classification (process); hierarchical arrangements; faceted arrangements; numeric notations; alphabetic notations; mixed notations

Figure 1. Example entry in the glossary.

onomy, subject heading list, ontology, terminology, authority list, and ring of synonyms.

We obtained 2,904 types (number of words counted only once) and 19,913 tokens (total number of words in the corpus). In the concordance tab (which shows the results in keyword-in-context (KWIC) format), it was possible to analyze the words linked to the text in the corpus itself, thus checking their context.

The candidate terms were searched in the corpus, and those that had definitional excerpts in the corpus were selected to compose the glossary, representing the cut-off established for this study.

The candidate terms without definitory excerpts in the corpus itself were, at first, discarded. The equivalences in Portuguese of the terms were obtained by searching the Brazilian literature in information science outside the corpus.

4.4 Data entry

A glossary entry was created for each term in the study sample. For each entry, a terminology datasheet was produced in which the information needed to compose the entries was recorded and stored as documentation in the glossary.

The terminology datasheet was created with the following fields: term identification (unique 1D), input term, collection date, term source, grammatical category (masculine, feminine or neuter, represented by the letters M, F, or N, respectively), subject area, definition, source of the definition, context of the term in the corpus (definitional excerpts; example of a real application of the term in its context taken from the corpus), source of the context (in the corpus), cross-references to equivalent terms (with source), cross-references to superordinate terms, cross-references to subordinate terms, cross-references to associated terms, notes for additional information, name of the person responsible for completion.

A sample terminology datasheet is shown in Table 2.

When the information in the corpus was insufficient to constitute a complete definition, information from NISO Z39.10 (2005; R2010) and ISO 25964 (2011; 2013) was consulted.

4.5 Final stage

In this stage, the most relevant information for the target audience was extracted from the terminology datasheets. Elements such as the grammatical category, the source from which the definition was taken, and an example taken from the corpus were extracted.

1D of the term:	Individual 1D
Term and Grammatical Category Preferred Descriptor Non-preferred Descriptor	Cat. Grammatical: use M, F, or N Preferred term: Portuguese entry Non-preferred term: English and other equivalents
Equivalence Relationship Use (USE)	Indicate preferred term
Used For (UF)	Indicate non-preferred terms in Portuguese
Used For (UF in English)	Indicate non-preferred terms in English
Hierarchical Relationship Generic Term (GT)	Indicate cross-reference to superordination term
Specific Term (ST)	Indicate cross-reference to subordinate terms
Associative Relationship (AR)	Indicate cross-references to association terms
Definition (Def.)	Definition of the preferred term: if possible, formulated from the corpus excerpts
Scope Note (SN)	Explanatory notes related to the scope of the preferred term
Historical Note (HN)	Historical note concerning the preferred term
Primary data source	ISKO Encyclopedia of Knowledge Organization (IEKO) corpus
Secondary data source	NISO Z39.10 (2005, R2010); ISO 25964 (2011, 2013)
Collector name	Responsible for the record
Entry date	Collection date (dd/mm/yyyy)

Table 2. Glossary terminology datasheet.

After the information was included, the entries in the glossary entries were arranged in alphabetical order. At this point, the glossary introduction, which presents all the relevant information for understanding the instrument, is also finalized (see Figure 2).

The table of abbreviations used in the glossary was also written. Finally, the glossary presentation form was selected as follows: i) each entry appears only with the first letter capitalized on the left and in boldface; ii) then comes the definition, in which the terms in external references are highlighted in boldface type to refer to other entries; and iii) at the end appears the equivalence in another language, also in boldface type.

5.0 Conclusion

This study has created a terminological definitional glossary for knowledge organization systems (KOS) to support understanding of the concepts related to these instruments. The definitional excerpts were obtained from a corpus composed of texts from the *ISKO Encyclopedia of Knowledge Organization* (*IEKO*). All the terms that make up the glossary are accompanied by their definition to limit the meaning of the concept, thus avoiding ambiguities.

The glossary's development path highlights the relevance of outlining the definition. Its importance goes beyond the simple explanation of a concept, as it represents the conveyance of this concept and the consolidation of its meaning within a domain.

The choice of the type of definition used must take into consideration its suitability for the glossary's purpose, going beyond its accommodation to pre-defined templates. The choice involves knowing the specifics of the specialty area, the type of term defined, and the target audience, among other less obvious factors.

It also highlights the importance of the indication of all the equivalences for the same entry term, since some of the terms used are part of the specialty terminology of other fields of knowledge, in addition to information science.

Acknowledgments

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; National Council for Scientific and Technological Development) for the Research Productivity Grant (PQ-2), Process number 303 650/2019-2.

References

Almeida, G. M. B. A. 2006. *Teoria Comunicativa da Terminologia e a sua Prática*. São Paulo: Alfa. https://periodicos.fclar.unesp.br/alfa/article/view/1413/1114

GLOSSARY INTRODUCTION

The Terminological Glossary for Knowledge Organization System (KOS) types was created to meet the need of students for information about the concept of these tools in the context of Information Science education. The glossary presents the definitions of the different types of knowledge organization systems (KOS) considered as information representation instruments, based on the KOS listed in NISO Z39.10 (2005, R2010) and ISO 25964 (2011, 2013), which are: synonym ring, authority list, subject heading list, term list, ontologies, classification systems, classification system for records management, taxonomy, terminologies, and thesaurus. The definitions were developed from definitional excerpts contained in a corpus composed of texts obtained from the ISKO Encyclopedia of Knowledge Organization (IEKO), and they explain the characteristics and functions of KOS more fully, facilitating students' understanding and learning of these instruments.

References and abbreviations used in the glossary

To allow greater clarity and efficiency in the use of the glossary, the following abbreviations have been used:

Symbology	Meaning
Related term (RT)	Indicates a type of relationship (non-hierarchical)
General term (GT)	Indicates a superordinate term
Specific term (ST)	Indicates a subordinate term
Use (USE)	Indicates a preferred term
Use For (UF)	Indicates a non-preferred term referring to terminological equivalence: it should not be used in thematic representation
Scope Note (SN)	Explains the context of a term's use, or gives some significance to the adoption of the term
Historical Note (HN)	Information referring to the preferred term
English (EN)	For the English language equivalent term
Definition (Def.)	Definition of the preferred term
Data source (DS)	Source of the information

This glossary is not a final, closed product; it is open to suggestions and additions of new terms and information. It can be accessed on your host (URL: https://appcode.com.br/socs/vocab/).

Figure 2. Glossary introduction.

ANSI/NISO (American National Standards Institute / National Information Standards Organization). 2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.* ANSI/NISO Z39.19-2005 (R2010). Baltimore: American National Standards Institute. https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf

Aristotle. 1960. *Posterior Analytics; Topica*. Greek text with translation by Hugh Tredennick and E. S. Forster. Loeb Classical Library 391. Cambridge, MA: Harvard University Press, 1960.

Barbosa, M. A. 1995. "Contribuição ao Estudo de Aspectos da Tipologia de Obras Lexicográficas." *Ciência da Informação* 24, no. 3: 1-9. https://doi.org/10.18225/ci.inf..v24i3.572

Barros, L. A. 2004. *Curso Básico de Terminologia*. São Paulo: Editora da Universidade de São Paulo.

Barreiros, L. L. S. 2017. "O Uso de Ferramentas Computacionais na Elaboração do Vocabulário de Eulálio Motta: AntConc e FLEx." *A Cor das Letras* 18: 216-41. https://doi.org/10.13102/cl.v18i2.1835

Berber Sardinha, T. B. 2000. "Linguística de Corpus: Histórico e problemática." *Documentação de Estudos em Linguística Teórica e Aplicada* 16: 323-367. https://doi.org/10.1590/S0102-44502000000200005

Berber Sardinha, T. B. 2004. *Linguística de Corpus*. São Paulo, SP: Editora Manole.

Cabré, M. T. 1999. *La Terminología: Representación y Comunicación.* Barcelona: Institut Universitari de Linguística Aplicada.

Costa, R. and Raquel Silva. 2004. "The Verb in the Terminological Collocations: Contribution to the Development of a Morphological Analyser: MorphoCom." In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. Lisboa: LREC. http://lrec-conf.org/proceedings/lrec2004/pdf/795.pdf

Dahlberg, I. 1978. "A Referent-Oriented Analytical Concept Theory of INTERCONCEPT." *International Classification* 5: 142-50. https://www.nomos-elibrary.de/10.5771/0943-7444-1978-3-142.pdf

Dahlberg, I. 1982. "Terminological Definitions: Characteristics and Demands." In *Problèmes de la Définition et de la Synonymie en Terminologie. Actes du Colloque International de Terminologie,* [1983]. [Quebec]: Association Internationale De Terminologie, 15-34.

De Lucca, J. L. 2006. "Identificação de Padrões Recorrentes no Discurso Técnico e Científico Para a Extração Auto-

mática de Candidatos a Contextos Definitórios em Língua Portuguesa." *Revista Intercâmbio* 15. https://revistas.pucsp.br/intercambio/article/view/3683

Engelmann, J. 2006. *Teoria da Definição: das Definições Reais às Definições Predicativas.* Tese de Doutoramento, Departamento de Filosofia, Pontifícia Universidade Católica do Rio de Janeiro. https://doi.org/10.17771/PUCRio.acad.9106

Faulstich, E. L. de J. 1995. "Socioterminologia: mais que um Método de Pesquisa, uma Disciplina." *Ciência da Informação* 24, *no* 3. http://revista.ibict.br/ciinf/article/view/566/567

Finatto, M. J. B. 2002. "O Papel da Definição de Termos Técnico-Científicos." *Revista da ABRALIN* 1, no 1 (julho): 73-97. https://lume.ufrgs.br/handle/10183/184253

Gomes, H. E. and M. L. D. A. Campos. 2019. *A Organização do Conhecimento na Web*: *Contribuições de Shiyali Ramamrita Ranganathan e Ingetraut Dahlberg*. Grupo de Pesquisa Estudos Ônticos e Ontológicos em Contextos Informacionais: Representação, Recuperação e Métricas - Cadernos Acadêmicos 1. Niterói: IACS/UFF. http://eooci.uff.br/wp-content/uploads/sites/319/2020/09/eooci_ca-1.pdf

Guarino, N., and G. Guizzardi. 2006. "The Defense of Ontological Foundations for Conceptual Modelling." *Scandinavian Journal of Information Systems* 18, no 1: 115-26. http://aisel.aisnet.org/sjis/vol18/iss1/1

Hodge, G. 2000. *Systems Of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington, DC: The Digital Library Federation, Council on Library lnformation Resources. https://files.eric.ed.gov/fulltext/ED440657.pdf

ISKO Encyclopedia of Knowledge Organization (IEKO), edited by Birger Hjorland and Claudio Gnoli. https://www.isko.org/cyclo/

ISO 1087-1. 2000. *Terminology Work - Vocabulary, Part 1: Theory and Application / Travaux Terminologiques - Vocabulaire - Partie 1: Théorie et Application*. Genève: ISO. https://www.iso.org/standard/20057

ISO 25964-1. 2011. *Thesauri and Interoperability with Other Vocabularies - Part 1: Thesauri for Information Retrieval.* Geneve: ISO. https://www.iso.org/standard/53657.html

ISO 25964-2. 2013. *Thesauri and Interoperability with Other Vocabularies - Part 2: Interoperability with Other Vocabularies*. Geneve: ISO. https://www.iso.org/standard/53658.html

Krieger, M. D. G. and M. J. B Finatto. 2004. *Introdução à Terminologia: Teoria e Prática*. São Paulo: Contexto.

Marini, S. 2013. *Da Tradução Terminológica em Glossário Temático na Área de Saúde Suplementar.* Brasília: Departamento de Línguas Estrangeiras e Tradução. https://repositorio.unb.br/bitstream/10482/14645/1/2013_SatiaMarini.pdf

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

McGuinness, D. L., R. Fikes, L. A. Stein and J. A. Hendler. 2002. "DAML-ONT: An Ontology Language for the Semantic Web. *IEEE Intelligent Systems* 17, no. 5: 72-80. https://ieeexplore.ieee.org/https://ieeexplore.ieee.org/document/1039835

Morais, M. W., J. L. Souza and R. A. S. Ramalho, 2019. *Representação de umMicrotesauro em SKOS: Um Estudo Metodológico.* http://eprints.rclis.org/38636/

Sager, J. C. 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.

Souza, R. R., and L. Alvarenga. 2004. "A Web Semântica e suas Contribuições para a Ciência da Informação." *Ciência da Informação* 33, no. 1: 132-41. https://doi.org/10.1590/S0l00--19652004000100016

Tagnin, S. E. O. 2005. *O Jeito que a Gente Diz: Expressões Idiomáticas e Convencionais - Inglês E Português*. São Paulo: Disal.

Vaz, G., L. M. Oliveira and I. Pierozzi Jr. 2017. "Visualização de Glossário em Sistemas de Recuperação de Informação." In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*. Porto Alegre: Sociedade Brasileira de Computação, 83-92. https://aclanthology.org/W17-6611.pdf

Zeng, M. L. 2008. "Knowledge Organization Systems (KOS)." *Knowledge Organization* 35: 160-82. https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-160.pdf