Ontologies in Knowledge Organization

Luís Miguel Oliveira Machado

Department of Philosophy, Communication and Information, University of Coimbra, 3004-530 Coimbra, Portugal; luis.machado@student.fl.uc.pt

Definition: Within the knowledge organization systems (KOS) set, the term “ontology” is paradigmatic of the terminological ambiguity in different typologies. Contributing to this situation is the indiscriminate association of the term “ontology”, both as a specific type of KOS and as a process of categorization, due to the interdisciplinary use of the term with different meanings. We present a systematization of the perspectives of different authors of ontologies, as representational artifacts, seeking to contribute to terminological clarification. Focusing the analysis on the intention, semantics and modulation of ontologies, it was possible to notice two broad perspectives regarding ontologies as artifacts that coexist in the knowledge organization systems spectrum. We have ontologies viewed, on the one hand, as an evolution in terms of complexity of traditional conceptual systems, and on the other hand, as a system that organizes ontological rather than epistemological knowledge. The focus of ontological analysis is the item to model and not the intentions that motivate the construction of the system.

Keywords: ontology; knowledge organization systems; ontological knowledge

1. Introduction

In the knowledge organization (KO)/information science (IS) community several authors, such as [1–5], see the work on ontology coming from the computation field as a kind of reinvention of the wheel or an etymological issue, as it concerns classification and other well-known aspects of knowledge organization processes. This situation was reflected in the different views regarding the typology of ontologies as representational artifacts, also known as knowledge organization systems (KOS). Considering the different technical, structural and functional characteristics of KOS, Mazzocchi [6] presents as a common denominator the function for which these “semantic tools” were designed: supporting the organization of knowledge and information, in order to facilitate their management and recovery.

Within the KOS set, the term “ontology” is paradigmatic of the terminological ambiguity in different typologies. Pieterse and Kourie [7] (p. 227), e.g., state that the term “ontology” refers to “a KOS that can be classified as a relationship list in Hodge’s classification [and] which is classified as a thesaurus in our classification.” Unlike these authors, Biagetti [8] (Section 3.1) considers that “ontologies are a kind of KOS that present the highest degree of semantic richness, as they allow to establish a great number of relations between terms, and provide attributes for each class.” Hjørland [9] (Section 3.3) also sees ontologies as a different kind of KOS “more general and more abstract” than other “traditional” KOS but, for the author, these “may just be understood as being restricted kinds of ontologies.” Contributing to the latter position will be the indiscriminate association of the term “ontology”, both as a specific type of KOS and as a categorization process, the latter considered by Smiraglia [10] to be a pillar in the development of any KOS. Though the association of the term “ontology” with the artifact and the process is found, its indiscriminate use is not advised by Souza and others [11] (p. 187): “it might be asserted that all KOS are the products of some kind of ontological modeling, but using the term ‘ontologies’ arbitrarily can cause confusion.” In addition to the KO/IS and computer science
areas, another area, philosophy, is necessary to bring to the debate to understand why the term “ontology” is also used as a categorization process.

Given the interdisciplinary nature of this topic, terminology issues are of vital importance for proper communication between different communities. In this context, we aim to present a systematization of the perspectives of different authors of ontologies, as representational artifacts, seeking to contribute to terminological clarification. This paper, in addition to this introduction, contains three more sections. In the second section, we present a brief historical contextualization of the term “ontology”, for a better understanding of the interdisciplinary issue. Then, in the third section, we present the systematization referred to above, focusing on the intention, semantics and modulation of ontologies, based on the two major approaches detected in the different perspectives. Finally, in the fourth section, we synthesize the insights of the present work.

2. Brief History: From Philosophy to Information Systems

The term “ontology” appeared in the 17th century, being attributed, in parallel but without known connection, to Rudolf Göckel and Jacob Lorhard; although, only after the publication of the work *Philosophia prima sive Ontologia* by Christian Wolff in 1730 did the spread of the term truly start [12,13]. In that work, Wolff called “ontology” Aristotle’s “first philosophy” [14]. The object of study of this discipline, which the philosopher also called “first science,” “wisdom” or “theology,” is now also called “Aristotle’s Metaphysics” [15]. In short, this work of Aristotle consisted in the systematization and categorization of all the entities that exist in the world [16,17].

While in the context of Philosophy the term “ontology” is related to a process (study or analysis), in the context of areas linked to knowledge organization/information systems the same term is associated with a product, an artifact. In this context, the term “ontology” can designate either a concrete system or, in a more abstract sense, a theory, and both can be in a formal (logic-based) or informal format [17].

It is in the area of artificial intelligence (AI) that the term “ontology” makes the transition from the field of Philosophy to the area linked to information systems. It was Mealy who, in 1967, first used the term “ontology” in this new context [18]. However, this was done in the philosophical sense of the term, referring to ontological analysis, because it would enable a better understanding of the structure of the world and thus facilitate its modeling, or part of it, in computational terms [19]. In this same context, another initial milestone occurred with the work of Hayes [20] in the following decades, seeking “an adequate theory of the world of common sense” [21] (p. 225) for application in robotics.

Despite those early uses of the term “ontology” to designate a “theory of a modeled world” [22] (p. 1964), the first formulation of a definition of ontology in the context of information systems only happened in a study by Neches and others published in 1991: “The ontology of a system consists of its vocabulary and a set of constraints on the way terms can be combined to model a domain” [23] (p. 40). Some authors [24] consider this study by Neches to be a pioneer in the field of IS despite the connection of its authors to the field of AI. Others, e.g., [1,25], point out an author declaredly associated with IS, B.C. Vickery, and his 1997 work where he claims that the issues faced by “ontological engineers” are not new to the IS community [2] (p. 285).

Among the group of researchers who participated in Neches’s study was T.R. Gruber, whose definition of ontology—an “explicit specification of a conceptualization” [26,27] (p. 199, 908)—would become paradigmatic, particularly in areas directly related to computing [28]. This definition is from a work first published online in 1992, where the author recognizes the philosophical origin of the term but explicitly departs from its meaning in this context: “this definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy” [29] (n/p.). In the formal publication of this definition, Gruber also points out the difference in his approach from that made in the original context of the term: “An ontology is an explicit specification of a conceptualization. The term is borrowed from
philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what ‘exists’ is exactly that which can be represented” [26,27] (p. 199, 908, italics in original). The key term of this definition is “conceptualization”, used by Gruber in the sense given by Genesereth and Nilsson [30] (p. 9, italics in original): “The formalization of knowledge in declarative form begins with a conceptualization. This includes the objects presumed or hypothesized to exist in the world and their interrelationships. The notion of an object used here is quite broad.”

Gruber’s definition created an almost synonymous association between the terms “ontology” and “conceptualization”. This association, potentially enhanced by the “recognized authoritative” role that Gruber appears to represent in his field [28] (pp. 206–207), contributed to the very broad use of the term “ontology” when applied to representational artifacts. Systems as simple as catalogs or slightly more complex as glossaries, till taxonomies, thesauri and the most expressive ones, using the axioms of full first-order, higher order or modal logic—all these types of information systems satisfy Gruber’s definition, and all are now common bedfellows under the rubric of ‘ontology’ [31] (p. vi).

In the words of Grenon [32] (p. 69): “the term ‘ontology’ applies to virtually any structure resembling, to some extent, a set of terms hierarchically organized which may be put in a machine-processable format.”

Despite this scenario it is possible to notice two broad perspectives regarding ontologies as artifacts. Definitions like the one proposed by Studer and others [33] (p. 184)—“an ontology is a formal, explicit specification of a shared conceptualization”—exemplify one of these positions. Exemplifying the other, we have the proposal by Arp and others [34] (p. 1) that “ontology = def. a representational artifact, comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations between them.” Several authors distinguish these two approaches in different ways. We can see these three criteria as complementary aspects of an analysis of these representational artifacts.

3. Two Types of Artifactual Ontologies

This section presents three common dichotomies: (i) reference vs. application ontologies; (ii) description logics vs. resource description framework (RDF)-based semantics; and (iii) ontological vs. conceptual models. These dichotomies can be understood as the result of a faceted analysis with the following criteria: (i) general objective of the ontology; (ii) formal language used; and (iii) applied modulation approach.

3.1. Reference vs. Application Ontologies

Smith [35] and Jansen [36] distinguish two types of ontologies: the “reference ontologies”, for use in scientific domains; and the “application ontologies”, for practical and specific purposes. While in the latter an ad hoc development can be applied and, in some cases, it is even unavoidable, this approach is totally discarded for the former [35,36]. These reference ontologies, designed to serve scientific purposes, should be developed according to the “principle of orthogonality” that would not only address the data silos problem but would also bring additional benefits, such as mutual consistency of ontologies, unnecessary mapping between ontologies, reduced redundancy, facilitated findability of specific ontological resources and optimized management of the ontological labor division [35].

Articulation with the ontological study deriving from the disciplinary area of origin is essential for the approach to reference ontologies: “information-systems ontology is itself an enormous new field of practical application that is crying out to be explored by the methods of rigorous philosophy” [13]. Articulation is seen as necessary by several researchers, e.g., [17,37,38], for the development of artifact ontologies that seek ontological rigor in the representation of reality. In contrast, in Gruber’s understanding, rigor seems less important than the ontology’s usefulness: “if ontologies are engineered things, then we don’t have to worry so much about whether they are right and get on with the business of building them to do something useful” [39] (p. 1). For Smith [35] (p. 33), “it is as
if all ontologies, both inside and outside science, are assigned by default the status of application ontologies.”

Linked to the “application” approach is the conception that all ontologies are the result of the common agreement of a community over a portion of the world. As Gruber [39] (p. 5) states: “I find it critical to remember that every ontology is a treaty—a social agreement—among people with some common motive in sharing.” This conception is questioned by Poli and Obrst since this result is usually obtained by the lowest common denominator whose utility will be quite doubtful “because it is inconsistent, has uneven and wrong levels of granularity, and doesn’t capture real semantic variances that are crucial for adoption by members of a community” [37] (p. 10).

Regarding interoperability, these application ontologies also have their limitations derived from the ad hoc mode with which they are usually constructed [36,40]. If the system is developed in this ad hoc way, where its quality is essentially measured by the extent to which the needs of the various stakeholders are met [41], the ontological aspects lose their relevance in the modulation process. Alternatively, several application ontologies can be mapped to each other if they are developed “through a choice or combination of types from the reference ontology that are appropriate to the respective aim” [36] (p. 171). In this situation the reference ontology will serve as a common benchmark for the application ontologies.

3.2. Description Logic vs. RDF-Based Semantics

Kless and others [42] make a distinction between ontologies associated with the description logic (DL) semantics and ontologies associated with RDF-based semantics. Commonly, RDF semantics is associated with what is called “lightweight ontologies”, a term that, according to Zhu and Madnick [43] (p. 9), is used in the literature in a very loose way: “data dictionaries, product catalogs, and topic maps are often considered to be lightweight ontologies.” The authors add that, “generally speaking, a lightweight ontology refers to a set of concepts organized in a hierarchy with is a relationships [and in opposition] are formal ontologies, which often use formal logic to specify constraints, relationships, and other rules that apply to the concepts” [43] (p. 9).

The proliferation of the “lightweight ontologies” in the 1990s was due to the need to provide the Web with systems capable of automatic inferences and interoperability as a means to achieve the so-called Semantic Web (SW). The ability to generate inferences is, however, restricted to the semantic expressiveness of the language used, which in the case of RDF is quite limited; even with the increase provided by the RDFS (RDF-Schema) extension, it is still a Web-oriented language and not an SW language [44]. RDF language treats indifferently individuals (instances), classes (types) and properties (attributes or relations), viewing them all as “resources”. In DL, the abstract description of domain knowledge, that is, the structural and intensional component (the terminology, known as TBox), is kept separate from the description of facts about objects/individuals (the assertions, called ABox). TBox represents the scheme or taxonomy of the domain of knowledge and ABox describes the assertions (attributes, roles, etc.) about instances regarding their class membership with TBox [45]. RDF-based semantics ontologies do not make a clear metaphysical distinction between the different elements; that is why ontologies that “are the result of ontologically driven design processes and aim at reality representation” are likely published using DL semantics [42] (Section 2). Without this ontological rigor there is no basis for avoiding false inferences when two ontologies are combined [42]. This potential lack of interoperability keeps the problem of “data silos”, for which ontologies should be a solution and not part of the problem.

Despite the limitations described above, it is these RDFS-based “light ontologies” that underlie what might be called the “web of linked data” which, although far behind the intended Semantic Web, is a non-negligible achievement [46]. It is also possible to see in this development of the Web a “democratization” of knowledge representation in this digital environment, meeting the original vision of Berners-Lee for it [47].
3.3. Ontological vs. Conceptual Model

Yet another distinction can be made between the result of two different approaches in knowledge representation that Grenon [32] calls “realist representationalism” and “pragmatist conceptualism”. While in the former the purpose is to capture the categories of the actual world, in what can be called an “ontological model”, in the latter the intention is to represent our conceptualization of a real or imaginary world, resulting in what may be called a “conceptual model”. Although there are similarities, the processes differ: “while conceptual modeling seeks to establish relationships between the abstract concepts of a domain, ontological modeling aims to identify objects and understand their nature through the description of their properties” [48] (p. 243, original in Portuguese).

An ontological model will result from the application of a philosophically well-founded ontological theory in an information system following rigorous ontological principles, such as whole-part theory, types and instantiation, identity and unity [17]. This endeavor is admittedly complex; however, it brings advantages in terms of stability and coherence of the developed model [13]. The ontological principles can also be applied in the development of a conceptual model, although more focused on the logical aspects that ensure the internal consistency of the model. The process of ontological analysis, in this case, does not necessarily seek an adaptation to reality in the same way as it is carried out in philosophical ontology.

Conceptual models are understood as explicit descriptions of mental models, which are considered to be “partial accounts of the external reality, filtered through the lens of a conceptualization, that people use to interact with the world around them” [49] (p. 4). The interaction between the individual and the surroundings involves a system of concepts; the conceptualization, “in terms of which the corresponding universe of discourse is divided up into objects, processes, and relations in different sorts of ways,” can be specified “to render explicit the underlying taxonomy” [13] (pp. 161–162). This can be a strictly pragmatic or epistemological enterprise when conceived of as consisting only in that of representing others’ conceptualization, so that reality falls out of the picture almost entirely [13,32].

However, the concept model can be built up from a rigorous ontological analysis: “ontological modeling can constitute a basis for conceptual modeling in order to provide the developer, clearly and unambiguously, with the necessary knowledge about the domain to be modeled” [48] (p. 243, original in Portuguese). This approach can be seen as a way to provide what Guarino and others [49] (p. 8) call a “grounding requirement” for conceptual models, which can be understood as a “sort of completeness requirement” for them. Ultimately, as Poli and Obrst [37] (p.6) say: “without ontology, there is no firm basis for epistemology.”

4. Final Remarks

As we could see the two approaches to ontologies differ in each of the three facets described. In the first facet, where the artifacts are analyzed according to the main intention for which it was developed, we have the contrast between reference and application ontologies that reflects what appears to be the main dichotomy in relation to the intended functionality for these artifacts as they are currently developed [32]. In the second facet, it is the semantic expressiveness associated with the two languages (description logics and RDF-Schema) that is in confrontation. These languages are two possible implementations of formal coding that make automatic processing by computers possible [44]. It is, therefore, a procedural stage in the development of these artifacts and should therefore not be seen as an intrinsic characteristic of ontologies [8,37]. Finally, in the third facet, the question of the difference between a truly ontological reading and another approach essentially epistemological is addressed. Poli and Obrst [37] (p. 3) describe the two approaches as follows: “Ontology is primarily about the entities, relations, and properties of the world, the categories of things. Epistemology is about the perceived and belief-attributed entities, relations, and properties of the world, i.e., ways of knowing or ascertaining things.”
To summarize, we have ontologies viewed, on the one hand, as a more complex conceptual system, and on the other hand, as a system that intends to organize ontological rather than epistemological knowledge. While the latter perspective maintains a relationship with the meaning of the term “ontology” coming from philosophy, the former presents a sense that deviates from the original. Although the appropriation of a term from another area of knowledge is common practice in the scientific community, especially when it involves new technologies [50], the case of changing the meaning of the term “ontology” shows traces of the process of “metaphorizing” which can be understood as “the result of encoding at the concept level. The resulting name or term for the concept can be understood in its new meaning without understanding the basis for the naming” [51] (p. 125).

The focus of ontological analysis is the item to model and not the intentions that motivate the construction of the ontology. This process involves an analytical complexity that makes the development of such systems quite onerous. However, it is the quality of this analysis that determines its true usefulness.

Funding: This research was funded by the Portuguese Foundation for Science and Technology, under the Ph.D. research grant (SFRH/BD/145937/2019).

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

Entry Link on the Encyclopedia Platform: https://encyclopedia.pub/3744.

References
1. Gilchrist, A. Thesauri, Taxonomies and Ontologies—An Etymological Note. *J. Doc.* 2003, 59, 7–18. [CrossRef]
2. Vickery, B.C. Ontologies. *J. Inf. Sci.* 1997, 23, 277–286. [CrossRef]
3. Soergel, D. The Rise of Ontologies or the Reinvention of Classification. *J. Am. Soc. Inf. Sci. Technol.* 1999, 50, 1119–1120. [CrossRef]
4. Currás, E. Ontologies, Taxonomies and Thesauri in Systems Science and Systematics; Chandos Publishing: Oxford, UK, 2010; ISBN 978-1-84334-612-8.
5. Dahlberg, I. What Is Knowledge Organization? *Knowl. Organ.* 2014, 41, 85–91. [CrossRef]
6. Mazzocchi, F. Knowledge Organization System (KOS). *Encycl. Knowl. Organ.* 2019. Available online: http://www.isko.org/cyclo/kos (accessed on 29 November 2020).
7. Pieterse, V.; Kourie, D. Lists, Taxonomies, Lattices, Thesauri and Ontologies: Paving a Pathway Through a Terminological Jungle. *Knowl. Organ.* 2014, 41, 217–229. [CrossRef]
8. Biagetti, M.T. Ontologies (as Knowledge Organization Systems). *Encycl. Knowl. Organ.* 2020. Available online: https://www.isko.org/cyclo/ontologies (accessed on 29 November 2020).
9. Hjørland, B. Knowledge Organization (KO). *Encycl. Knowl. Organ.* 2019. Available online: http://www.isko.org/cyclo/knowledge_organization (accessed on 29 November 2020).
10. Smiraglia, F.P. The Elements of Knowledge Organization; Springer International Publishing: Cham, Switzerland, 2014; ISBN 978-3-319-09356-7.
11. Souza, R.R.; Tudhope, D.; Almeida, M.B. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. *Knowl. Organ.* 2012, 39, 179–192. [CrossRef]
12. Guizzardi, G. Ontological Foundations for Structural Conceptual Models. Doctoral Thesis, Telematica Instituut/CTIT, Enschede, The Netherlands, 2005.
13. Smith, B. Ontology. In *Blackwell Guide to the Philosophy of Computing and Information*; Floridi, L., Ed.; Blackwell: Oxford, UK, 2003; pp. 155–166.
14. Henning, B. What is Formal Ontology? In *Applied Ontology: An Introduction*; Munn, K., Smith, B., Eds.; Ontos Verlag: Heusenstamm, Germany, 2008; pp. 39–56.
15. van Inwagen, P.; Sullivan, M. Metaphysics. *Stanf. Encycl. Philos.* 2018. Available online: https://plato.stanford.edu/archives/spr2018/entries/metaphysics/ (accessed on 29 November 2020).
16. Thomasson, A. Categories. *Stanf. Encycl. Philos.* 2018. Available online: https://plato.stanford.edu/archives/fall2017/entries/categories/ (accessed on 29 November 2020).
17. Almeida, M.B. Revisiting Ontologies: A Necessary Clarification. *J. Am. Soc. Inf. Sci. Technol.* 2013, 64, 1532–2890. [CrossRef]
18. Smith, B. Ontology and Information Systems. 2004. Available online: http://ontology.buffalo.edu/ontology_long.pdf (accessed on 29 November 2020).
19. Mealy, G.H. Another Look at Data. In *Proceedings of the Fall Joint Computer Conference, November 14–16, Anaheim, California (AFIPS Conference Proceedings, Volume 31)*; Thompson Books: Washington, DC, USA; Academic Press: London, UK, 1967; pp. 525–534.
49. Guarino, N.; Guizzardi, G.; Mylopoulos, J. On the philosophical foundations of conceptual models. In *Information Modelling and Knowledge Bases XXXI*; Dahanayake, A., Huiskonen, J., Kiyoki, Y., Thalheim, B., Jaakkola, H., Yoshida, N., Eds.; IOS Press: Amsterdam, Netherlands, 2020; Volume 321, pp. 1–15. ISBN 978-1-64368-044-6.

50. English, K. Understanding Science: When Metaphors Become Terms. *ASP* 1998, 151–163. [CrossRef]

51. Temmerman, R. The Process of Revitalisation of Old Words: “Splicing”, a Case Study in the Extension of Reference. *Terminology* 1995, 2, 107–128. [CrossRef]