International Journal of Digital Content Management
Allameh Tabataba'i University
Vol. 2, No. 2, Winter & Spring 2021

Research Paper

An Analytical Approach to Categorize Knowledge Organization Systems (KOSs) in Digital Content Engineering and Management

Maziar Amirhosseini*

Date Received: 2021/05/07 Date Accepted: 2021/05/26

Abstract
Purpose:The principal purpose of the present article is to analyze the technical mechanisms and the categorization procedure of thesaurus, ontology, and their types in storage and retrieval of digital and non-digital content. The types of knowledge organization systems (KOSs) which are addressed in this article are Roget's thesaurus, thesauri, Micro, Macro and Meta thesaurus, ontologies, and lower, middle, or upper level ontologies. The study attempts to demonstrate categorization procedure through determining the position of KOSs in the context of data, information, and knowledge (DIK) by explaining their engineering mechanisms such as data, information and knowledge engineering in content storage and retrieval, especially digital contents. Method:The research method relies on documented and historical methodology. Findings: As ontologies have taken the highest position amidst KOSs in making digital resources available in web-based environment, it is suggested that Iran and other developing countries use the capacities and capabilities of ontologies, especially in the development of national ontologies, in order to construct their knowledge-based infrastructure and system to achieve high performance in digital content engineering and management.

Keywords: Knowledge Organization Systems (KOSs); Thesauri; Ontologies; Engineering Mechanism; Categorization Procedure.

---

* Faculty Member & Assistant Professor, Academic Relations and International Affairs, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran. Email: mazi_lib@yahoo.com

Introduction

Knowledge Organization Systems (KOSs) is a generic term used for referring to a wide range of items (Zeng 2008) such as lists, authority files, gazetteers, dictionaries, encyclopædias, synonym rings, taxonomies, folksonomies, classification schemes, [subject headings] thesauri, and ontologies (Hjørland, 2008). In their historical evolution, KOSs have benefited from the developmental experiences of previous KOSs (Amirhosseini, 2008; 2021) and hence, their structures are not simply a repetition of the past (Zeng 2008). They focus on their specific purposes to model the underlying semantic structure of a domain for storage and retrieval (Tudhope and Nielsen, 2006). KOSs from the simplest to the most complex ones are different from each other in following their specific purposes and in operating their specific organization methods to make varied forms of data, information, and knowledge available. Therefore, they share similarities in terms of applying a kind of order to make information available (i.e., their goal). On the other hand, they differ in the method of creating order to access specific forms of information (i.e., their purpose) (Amirhosseini, 2021).

KOSs have used their technical procedure such as data engineering, information engineering, and knowledge engineering to organize data, information and knowledge (DIK) (Soergel, 2009). In analyzing the applications of KOSs' technical procedures, it can be said that these engineering mechanisms cover three main roles in achieving the goals and purposes of KOSs. Firstly, engineering mechanisms are the main factor in creating an internal order to develop the semantic relations network (Mazzocchi, 2018) in development of KOSs in specific domains or tasks based on standard techniques and methods (International Organization for Standardization, 2013). Secondly, the creation of compatibility and interoperability between different types of KOSs to develop various types of knowledge organizations is possible through relying on special techniques such as Mapping, Switching, Merging, and Integration (Aitchison, Gilchrist, and David Bawden, 2000). Lastly, technical mechanisms play a great role in organizing various types of content such as digital content by using KOSs as a tool (National Information Standards Organization, 2003) to achieve the goals and purposes in storage and retrieval. Therefore, KOSs' engineering mechanisms have played a key role in creating various types of KOSs as well as organizing various types of content, while KOSs have their specific position in DIK.

The main purpose of this article is to analyze the DIK, engineering mechanisms, and categorization procedure of thesaurus and ontology and their types in the storage and retrieval of digital and non-digital content. In this article, the following issues are discussed to achieve the above main purpose. The development framework of KOSs such as data, information, and knowledge engineering of the DIK are explained in thesauri, ontologies, and their various types. Moreover, the categorization procedure is discussed to represent the position of KOS according to DIK.

The development framework of KOSs

In the process of organizing data, information and knowledge management, knowledge organizations intend to organize data, information, and knowledge (DIK) or other various types of content to achieve semantic structures. The DIK can be arranged based on a hierarchical pyramid (Ackoff, 1989), where data is at the lowest level and information and knowledge occupy the highest levels, respectively. DIK have infrastructural roles in data, information (Chen et al., 2009), and knowledge management (Zeng et al., 2020). In fact, KOSs from the simplest to the most complex ones provide a framework or schema for storing and organizing data, information, and knowledge (Soergel, 2009) to participate in data, information, and knowledge management. Such framework resembles a tool (Hjørland, 2008) to display efficacy in organizing various types of content, particularly in organizing digital content (Golub, Schmiede and Tudhope, 2019), when developing digital libraries (Hodge, 2000). This process has been down through operating technical methods (International Organization for Standardization, 2011; 2013; British Standards Institution, 2005-2008; National Information Standards Organization, 2003) based on data, information and knowledge engineering. Therefore, KOSs are tools for describing resources based on DIK and the related engineering mechanism to increase the performance of content storage and retrieval regarding digital and non-digital contents, especially to develop digital libraries.

The development framework of Thesauri

The idea of a semantic order was proposed by Peter Mark Roget's thesaurus in 1852, which was not a simple alphabetical list of words (Foskett 1980) , but had a conceptually-based structure in making the relationship between related words in a specific semantic domain to help express the thoughts of writers (Arano 2005). As early as 1951, Hans Peter Luhn, influenced by Roget's semantic order, established semantic relations (hierarchical and associative relations) between related terms called thesaurus to operate in information storage and retrieval (Foskett 1980). Thesauri, while using the semantic order method in Roget's thesaurus are effective tools in organizing content, especially digital content, by relying on a semantic network system between descriptors to increase the performance of information storage and retrieval systems. The purpose of thesaurus development in specific fields of knowledge (Lancaster 1972) has resulted in the fast production growth of a large number of specialized thesauri in the late 1960s and early 1970s (Amirhosseini, 2021). Production of a large number of specialized thesauri caused the lack of coordination and compatibility between the available thesauri thus, interfering with information storage and retrieval system. The integration and compatibility between specialized thesauri were achieved by following certain theoretical foundations (Amirhosseini, 2021) through operation of the specific standard techniques (Aitchison, Gilchrist, and David Bawden, 2000). These standard methods and techniques, complemented with the support of information engineering mechanisms, have been used in creating thesauri and their various types in the 1980s are: Micro, Macro and Meta thesaurus (Zeng 2019) for their application in information storage and retrieval. Therefore, thesauri are placed in the role of information in fulfilling their purpose in increasing the performance of information storage and retrieval, while information engineering is considered the most important mechanism for the development of thesauri and their types.

The development framework of ontologies

The hierarchical and associative relations resulted in developing general or imprecise semantic relations. In this case, a novel KOS has been developed to prepare granular semantic relations between concepts (Cat 2017) on a particular topic in knowledge storage and retrieval (Dragoni et al. 2012), especially in developing digital libraries (Biagetti, 2021) in the 1980s (Gruber, 2009) which is called specialized ontologies. The large numbers of specialized ontologies have been constructed in various scientific fields until the beginning of the 21st century (Dombayci 2019). While the large number of specialized ontologies led to the formation of incoherence between them (Amirhosseini, 2021), the operational techniques in semantic interoperability were developed to make consistency and harmony between specialized ontologies (Jin 2018). In early 21st century, this process resulted in developing various types of ontologies such as lower, middle, or upper level ontologies (Mascardi et al., 2007). Semantic interoperability in ontologies can be defined as the ability of different agents, services, and applications to communicate data, information, and knowledge — while ensuring accuracy and preserving the meaning of that same data, information, and knowledge (Zeng, 2019). Therefore, since ontologies and their types use knowledge storage and retrieval techniques to create knowledge-based systems, they have the position of knowledge and the appropriate engineering mechanism for them is knowledge engineering.

Discussions on categorizing of KOSs

This section presents the categorization procedure based on "semantic staircase" used to determine the position of the aforementioned KOSs through clarifying the DIK that they intend to organize and analyzing their related engineering mechanism that support each KOS. The idea of a "semantic staircase" presented by Olensky (2010), which glossaries (or other less structured KOSs) are placed at the lower grade and ontologies at the top of the hierarchy of KOSs (Mazzocchi, 2018). The following figure shows the positions of the mentioned KOSs based on a semantic staircase.

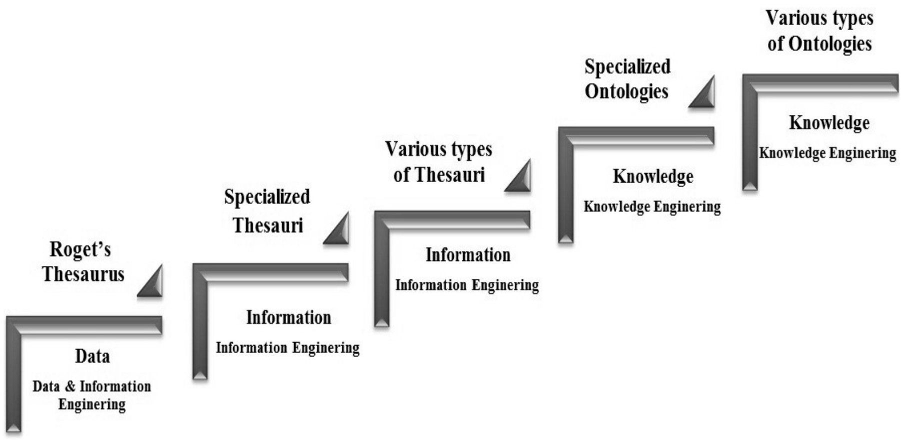

Figure 1: The Categorization of the KOSs based on "semantic staircase" (Olensky, 2010)

The above figure demonstrates that there is a specific position for each of the mentioned KOSs based on DIK. In this figure, the first position belongs to Roget's thesaurus as the simplest KOS, which establishes the idea of a semantic system between words or data through relying on data and information engineering mechanisms to help writers with finding the appropriate words to express their idea. The specialized thesauri are placed in the position of information, to play a role in establishing a hierarchical and dependent semantic relationship between terms in improving the performance of the information storage and retrieval system based on the information engineering mechanism. The third position belongs to various types of thesauri, which include micro, macro, and meta-thesauri which is the information position. The various types of thesauri are in a higher position than specialized thesauri, because they create compatibility between specialized thesauri through using the specific techniques based on information engineering. Specialized ontologies take the fourth position or knowledge position in terms of storing and retrieving knowledge by establishing precise and granular semantic relations between concepts through knowledge engineering. The various types of ontologies, such as lower, middle and upper level ontologies are placed in the final position by establishing interoperability between specialized ontologies in order to create compatibility between them by relying on specific knowledge engineering techniques.

Conclusion

KOS follows its own purpose to store and retrieve the content It aims to organize. Moreover, each of the KOSs has a specific position in the hierarchical relations of data, information, and knowledge. Although all KOSs have used the pre-existing experiences and techniques belonging to their previous generations, each KOS also uses its own engineering method and techniques including data, information, and knowledge engineering to organize content. Ontologies belong to the position of knowledge, which is the highest position in semantic staircase model, while encompassing the lower levels of information and data. Furthermore, while ontologies have also used information and data engineering methods, the appropriate engineering method for ontologies is the knowledge engineering mechanism for creating the ontology knowledge organization and increasing the performance of the content storage and retrieval. The use of ontologies as tools for knowledge organization provides integrated access to the use of digital objects based on semantic interoperability in making relations between the meanings attributed to documents managed by different repositories (Biagetti, 2021). Therefore, ontologies are located in the position of knowledge and are specifically designed to store, retrieve, and share knowledge, especially digital contents, through knowledge-based systems, especially in web-based environments, based on the principles of knowledge management systems. In conclusion, digital content management and engineering in the present age must use the latest human achievement, ontology, in the organization and dissemination of digital content. Thus, it is suggested that Iran and other developing countries use the capacities and capabilities of ontologies, especially in the development of national ontologies, in order to construct their knowledge-based infrastructure and system to achieve high performance in digital content engineering and management.

References

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9

Aitchison, Jean, Alan Gilchrist and David Bawden. (2000). *Thesaurus Construction and Use: A Practical Manual*. 4 ed., London: Aslib.

Amirhosseini, Maziar. (2008). Dialectic Scheme in Thesaurus Creation. *Library Philosophy and Practice.*

Amirhosseini, Maziar (2021) A dialectic perspective of the evolution of thesauri and ontologies based on historical progress of the related ideas. *Knowledge Organization* (in press)

Arano, S. (2005). Thesauruses and ontologies. Hipertext.net, num. 3, 2005. Retrieved from http://www.hipertext.net

Biagetti, M. T. (2021) Ontologies (as knowledge organization systems). In Edited by Birger Hjørland and Claudio Gnoli, *Encyclopedia of Knowledge Organization*. International society for organization of knowledge (ISKO) https://www.isko.org/cyclo/ontologies

British Standards Institution. (2005-8). *BS 8723: 2005-8. Structured Vocabularies for Information Retrieval - Guide*. (Published in five separate parts between 2005 and 2008), London: British Standards Institution.

Cat, Jordi. (2017). The unity of science. In Edward N. Zalta (ed.). *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/scientific-unity/#ConcWhyUnitWhatDiffDoesItRealMake.

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Liere, R. v., Ma, K.-L., Ribarsky, W., Scheuermann, G. & Silver, D. (2009). Data, Information, and Knowledge in Visualization. *IEEE Computer Graphics and Applications*, 29, 12-19.

Dombayci, Canan. (2019). *Conceptual Modelling for Integrated Decision-Making in Process Systems*. Ph.D Dissertation. Department d'Enginyeria Quimica – DEQ, Universitat Politecnica de Catalunya – UPC. Barcelona, Spain.

Dragoni, Mauro, Célia da Costa Pereira, and Andrea GB Tettamanzi. (2012). A Conceptual Representation of Documents and Queries for Information Retrieval Systems by Using Light Ontologies. *Expert Systems with applications*, 39 (12), 10376–88.

Foskett, D.J. (1980). Thesaurus. In A. Kent and H. Lancour (Eds.), *Encyclopedia of Library and Information Science* vol. 30, pp. 416-463.

Golub, K., Schmiede, R. & Tudhope, D. (2019) Recent applications of Knowledge Organization Systems: introduction to a special issue.

International Journal on Digital Libraries, 20, 205–207. https://doi.org/10.1007/s00799-018-0264-8

Gruber, Thomas R, L. Liu, O¨ zsu, and M.M. Tamer. (2009). "Ontology." In *the Encyclopedia of Database Systems*. Berlin Heidelberg: Springer-Verlag.

Hjørland, Birger. (2008). What Is Knowledge Organization (KO)?. *Knowledge Organization* , 35 (2/3), 86–102.

Hodge, Gail. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington: The Digital Library Federation, http://www.clir.org/pubs/abstract/pub91abst.html.

International Organization for Standardization (ISO) (2011). *ISO/FDIS 25964-1: Information and Documentation -Thesauri and Interoperability with other Vocabularies - Part 1: Thesauri for Information Retrieval.* Geneva: International Organization for Standardization.

International Organization for Standardization (ISO). (2013) *ISO/FDIS 25964-2: Information and Documentation -Thesauri and Interoperability with other Vocabularies - Part 2: Part 2: Interoperability with other vocabularies*. Geneva: International Organization for Standardization.

Jin, Z (2018) Ontology-Oriented Interactive Environment Modeling in: Chapter 4. In *Environment Modeling-Based Requirements Engineering for Software Intensive Systems*. San Francisco: Morgan Kaufmann, Pp 45-67.

Lancaster, F. W. (1972). *Vocabulary Control for Information Retrieval*. Washington: Information Resources Press.

Mascardi, V., V. Cord, and P. Rosso. (2007). A comparison of upper ontologies. In Matteo Baldoni, Antonio Boccalatte, Flavio De Paoli, Maurizio Martelli, and Viviana Mascardi, editors, *WOA*, 55–64. Seneca Edizioni Torino.

Mazzocchi, Fulvio. (2018). Knowledge organization system (KOS). *Knowledge Organization* 45 (1), 54-78. Also available in ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli. Retrieved from http://www.isko.org/cyclo/kos

National Information Standards Organization. (2003). *ANSI/NISO Z39.19 (2003) Guidelines for the Construction, Format, and Management of Monolingual Thesauri*. Bethesda: NISO.

Olensky, Marlies. (2010). Semantic interoperability in Europeana: An Examination of CIDOC CRM in Digital Cultural Heritage Documentation. *Bulletin of IEEE Technical Committee on Digital Libraries*, 6 (2). Retrieved from https://web.archive.org/web/20130620181231/https://www.ieee-tcdl.org/Bulletin/v6n2/Olensky/olensky.html.

Soergel, Dagobert (2009) Knowledge Organization Systems: Overview. Retrieved from http://www.dsoergel.com/SoergelKOSOverview.pdf

Tudhope, D. & Nielsen, M. L. (2006) Introduction to Knowledge Organization Systems and Services. *New Review of Hypermedia and Multimedia*, 12 (1), 3-9.

Zeng, Marcia Lei. (2008). Knowledge Organization Systems (KOS). *Knowledge Organization* 35 (2-3), 160-82.

Zeng, Marcia Lei. (2019). Interoperability. *Knowledge Organization*, 46 (2), 122-146. Also available in Hjørland, Birger and Gnoli, Claudio eds. ISKO Encyclopedia of Knowledge Organization. http://www.isko.org/cyclo/interoperability

Zeng, Marcia Lei; Hong, Yi; Clunis, Julaine; He, Shaoyi; Coladangelo, L. P. (2020). Implications of Knowledge Organization Systems for Health Information Exchange and Communication during the COVID-19 Pandemic. *Data and Information Management* , 4(3), 148-170.

Recommended Citation

Maziar Amirhosseini, Maziar(2021). "An analytical approach to categorize knowledge organization systems (KOSs) in digital content engineering and management". *International Journal of Digital Content Mangement, 1(2).*