

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/346629952>

A Methodological Proposal Towards Domain Ontology Enrichment

Conference Paper · December 2020

DOI: 10.5771/9783956507762-23

CITATIONS

0

READS

107

2 authors:



Webert Araújo

Federal University of Minas Gerais

12 PUBLICATIONS 4 CITATIONS

[SEE PROFILE](#)



Gercina Angela de Lima

Federal University of Minas Gerais

142 PUBLICATIONS 560 CITATIONS

[SEE PROFILE](#)

Webert Júnior Araújo – Federal University of Minas Gerais, Brazil
Gercina Ângela de Lima – Federal University of Minas Gerais, Brazil

A Methodological Proposal Towards Domain Ontology Enrichment

Abstract:

Since the current methods for domain ontology enrichment present some gaps, due to knowledge dynamicity, this investigation aims to develop a methodology for domain ontology enrichment that overcomes the existing methods' gaps. To address the goal, four steps sustain the research methodology: 1) An exploratory study of Knowledge Organization Systems maintenance and updating; 2) Mapping and analysis of the methods for enriching ontologies, from the literature review; 3) Qualitative content analysis of documents selected in Phases 1 and 2; 4) Development of the methodology for domain ontology enrichment. The result is a novel methodology for domain ontology enrichment, called METHODOE.

1.0 Introduction

Since ontologies are a type of Knowledge Organization Systems and knowledge is dynamic, ontologies must be updated periodically. Unfortunately, most ontologies developers ignore the ontology maintenance and updating area. They focus only on the development of these KOS, ignoring the fact that knowledge can change, terms used to represent concepts become obsolete, new terms emerge, and new scientific discoveries are made. So, ontologies must pursue this knowledge evolution.

An approach to update ontologies is through the ontological enrichment process. The enrichment process aims to expand an already developed ontology with new components (e.g. concepts, relationships, properties, and axioms); in consequence, the domain representation increases its potential. Hence, this study aims to develop a domain-independent methodology for the ontology enrichment process.

The literature presents several proposals for ontology enrichment. These proposals have the following limitations: 1) enrichment only of some ontologies' components (e.g., Faatz 2001; Faatz and Steinmetz 2002; Valakaros *et al.* 2004); 2) enrichment based on a very particular data source (e.g., Navigli and Velardi 2006; Amar, Gargouri, and Hamadou 2013); 3) enrichment applied to a specific domain (e.g., Faatz and Steinmetz 2002; Valakaros *et al.* 2004; Navigli and Velardi 2006; Booshehri *et al.* 2013); 4) intuitive and non-systematic methods.

The question of this investigative research is: how to develop a methodology for domain ontology enrichment that overcomes the gaps in existing methods?

We assume that literature can provide indications of how we can get the knowledge to develop this kind of methodology. Especially in the literature regarding Knowledge Organization Systems maintenance and in the empirical studies about ontology enrichment.

The motivation behind this theoretical investigation is the concern in improving existing domain ontologies since these instruments are important in the communication, interpretation, and reasoning of knowledge. Besides, ontologies help in the Semantic Web context, favoring the semantic integration between different systems and vocabularies. Likewise, it supports the organization and retrieval of information. Thus, studies are necessary to contribute to the topic of ontology maintenance because ontology researchers still neglect this area.

2.0 The research methodology

The research methodology is sustained in the following four steps: 1) An exploratory study of Knowledge Organization Systems maintenance and updating; 2) Mapping and analysis of the methods for enriching ontologies, from the literature review; 3) Qualitative content analysis of documents selected in steps 1 and 2; 4) Development of the methodology for domain ontology enrichment.

1) An exploratory study of Knowledge Organization Systems maintenance and updating

In this research, we consider the ontology enrichment process as belonging to a more comprehensive area within ontological engineering, which is the ontology maintenance. Thus, we determine the main norms and methods for maintaining Knowledge Organization Systems as a knowledge source, from which we could extract inputs for the development of the methodology we intend to work within this study.

KOSs, such as thesauri, classification systems, and taxonomies, have similarities with ontologies in some aspects. Therefore, some strategies for maintaining and updating these instruments can also be reused in ontologies. So after exploratory research on the topic, we selected the work of these authors: Kim (1973), Soergel (1974), ANSI/NISO Z39.19 (2005), ISO 25964-1 (2011), ISO 25964-2 (2013).

2) Mapping and analysis of the methods for enriching ontologies from the literature review

In this stage, a narrative type literature review was carried out, where the objective was to map the works addressing the ontology enrichment thematic to verify which researches have already addressed this theme. The research was carried out in Information Science and Computer Science databases using the following expressions (in English and Portuguese): *Ontology Enrichment*, *Ontological Enrichment*, *Ontology Expansion*, *Ontology Extension*, *Ontology Specialization*, *Ontology Refinement*, *Ontology Enlarge*, *Ontology Completeness*, *Ontology Improvement*. The search strategy used truncators, boolean operators, advanced search, and specific filters for each database. The period determined in the research was from 1990 to 2018.

After applying the search strategy, duplications and works not related to the scope of this research (by analyzing the title and abstract) were eliminated, obtaining a result of 35 works in total. Then, these 35 works were read using the following exclusion criteria: (1) works dealing with another Knowledge Organization Systems or databases enrichment; (2) works addressing another process and not enrichment (such as ontology learning, evolution); (3) works focusing only in the technique of knowledge acquisition and did not result in the enrichment of ontologies. In the end, 15 studies were considered for analysis in this review, which are: Faatz et al. 2001; Faatz and Steinmetz 2002; Valakaros et al. 2004; Navigli and Velardi 2006; Bendaoud, Toussaint, and Napoli 2008; Carvalho et al. 2010; Barbur, Blaga, and Groza 2011; Petasis et al. 2011; Hashimi and Kulathuramaiyer 2013; Booshehri et al. 2013; Amar, Gargouri, and Hamadou 2013; Booshehri and Luksch 2015; Al-Yahya, Al-Malak, and Aldhubayi 2016; Gómez-Moreno; Mestre-Mestre, 2017; Guerram; Mellal, 2018.

3) Qualitative content analysis of documents selected in steps 1 and 2

Given the recovered and selected documents in the previous steps, 1 and 2, a qualitative analysis of these documents' content was performed. We selected the main points which deal with maintenance and updating in KOS and based on this, we created categories to organize the extracted information.

The analysis of the documents selected in step 1 demonstrates that when maintaining Knowledge Organization Systems there are a few points to consider: (1) define the person responsible for maintenance; (2) categorize the type of change (which may cover the inclusion of a term, replacement, exclusion); (3) control changes made (like the source of the extracted information, inclusion date); (4) identify the consequences of maintenance for other systems using the KOS; (5) define a periodicity for the maintenance. Therefore, despite being focused primarily on updating thesauri, the way this process works has well-founded information adaptable to the context of ontology enrichment.

The analysis of the 15 selected papers revealed that they employ methods focusing on the technique used for information extraction, the knowledge source, and the type of enrichment performed. Among the identified information extraction techniques, there are statistical analysis, similarity measures, machine learning algorithms, syntactic analysis (part-of-speech, named entity recognition, parsing, stemming, tokenization, lemmatization), formal concept analysis, cluster techniques. Concerning knowledge sources we identified: textual corpus, semantically annotated corpus, thesaurus, ontology, web page content, lexical bases (such as Wordnet), machine-readable dictionary, data in linked data. Regarding the types of enrichment, they highlight lexical enrichment, conceptual enrichment, enrichment of taxonomic relations, enrichment of non-taxonomic relations, enrichment of axioms.

We realize that most works have a narrowed perspective and they approach specific methods for a knowledge domain, consequently presenting the following shortcomings: (1) there is no planning for enrichment; (2) lack of details on how to perform some steps; (3) the methods are empirical, intuitive and not systematic. Notwithstanding these shortcomings, the analysis of these articles and papers still provided valid inputs for the development of the enrichment methodology.

We created an action plan from the inputs generated in steps 1 and 2, which brought up information about what should be considered when maintaining a KOS and also about the main components of the enrichment process (such as knowledge extraction technique, knowledge source, and type of enrichment). This action plan will help in the development of the enrichment methodology, which will appear in the next stage of the research methodology.

4) Development of the methodology for domain ontology enrichment

Information gathered in the previous phases helped to develop an action plan with key strategies for the development of the ontology enrichment methodology. As seen in Table 1, this plan has seven features and 11 strategies that will guide the methodology's development ensuring it is better managed.

Table 1. Action plan for the ontology enrichment methodology development

#	Functionality	Action / strategy
1	Assessment of the need for enrichment	1.1) Develop a step to analyze the objectives of the ontology, users, and competency questions, if any.
2	Ontology diagnosis	2.1) Describe the possible ways to make the diagnosis.
3	Knowledge acquisition	3.1) Describe the possible knowledge sources; 3.2) Describe the possible techniques for extracting knowledge. 3.3) Explain how to extract knowledge.
4	Knowledge processing	4.1) Explain how to handle the extracted information.

5	Enrichment of ontology components from extracted content	5.1) Analyze the content extracted in the previous step; 5.2) Enrichment of the ontology components.
6	Evaluation and validation of the enrichment content	6.1) Describe the possible forms of evaluation and validation. 6.2) Check out the ontology after the content inclusion.
7	Methodology Documentation	7.1) Describe what was executed in each step of the methodology.

The methodology developed according to the action plan above will be described in the next section, "Results."

3.0 Results

As a result, we present the three phases which unfold in seven steps of the proposed methodology, which we call METHODOE (Methodology for Domain Ontology Enrichment): 1) Pre-enrichment; 2) Enrichment; 3) Post-enrichment. Figure 1 presents a general outline of METHODOE.

This methodology maps the entire enrichment process and makes it structured and organized. This methodology does not intend to describe in detail each of the possible techniques for extracting knowledge to carry out enrichment. Rather, it will be flexible and allow the ontologist to choose the knowledge source and the extraction technique that best fits the domain represented by the ontology since each domain has its particularities.

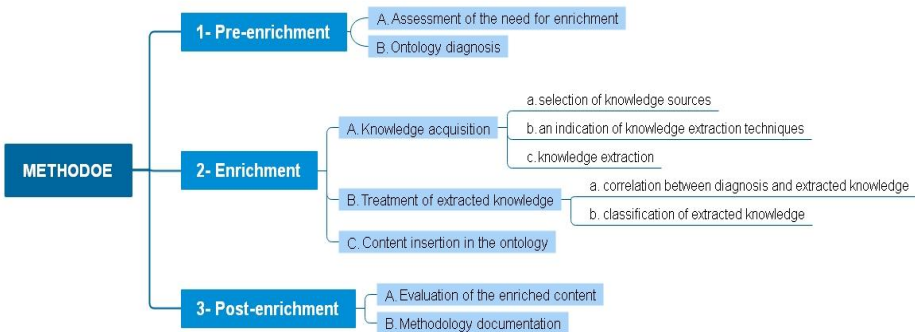


Figure 1. METHODOE general outline

3.1 Pre-enrichment

Pre-enrichment is the first phase proposed in METHODOE and has the following steps: (A) Assessment of the need for enrichment; (B) Ontology diagnosis.

In *Step A - Assessment of the need for enrichment* occurs the analysis to understand if the ontology responds to its the objectives, this can be done using various methods from the ontology evaluation area. A very common way to do it is through Competency Questions (CQs). Thus, if the ontology is unable to answer the competency questions elaborated in its development project, there will be indications showing that it needs enrichment. It is also possible to propose new CQs for the ontology. The assessment frequency of the need to enrich the ontology will depend on each domain, or how often

the terminology and the domain evolve. At this stage, a report should be generated, describing the need, or not to enrich the ontology with the necessary justifications.

In *Step B - Ontology diagnosis*, an examination of the ontology is implemented to find out in which parts the enrichment will be necessary. Again, there are multiple ways to do this. The ontologist can check the entire ontology (each concept and relationship) with a domain specialist guidance and look for possible points where improvements might be necessary. Another possibility is the use of 41 pitfalls for diagnosing ontologies developed by Poveda-Villalón (2016), as they deal with common mistakes made when building ontologies, these errors are indications for carrying out the enrichment activity. Not all pitfalls apply to the enrichment process. In this step, a detailed report (identifying mainly the location of the problem in the ontology structure) should be generated from the diagnosis result.

3.2 Enrichment

After the pre-enrichment phase, the ontology enrichment process is carried out. The steps composing this phase are (A) Knowledge acquisition; (B) Processing of extracted knowledge; (C) Content insertion in the ontology.

Step A - Knowledge acquisition deals with access to the raw material (information) for enrichment, and can also be an iterative step that happens throughout the enrichment process. This stage has three activities: (a) selection of knowledge sources; (b) indication of knowledge extraction techniques; (c) knowledge extraction.

About the *Activity a - selection of knowledge sources*, there are various knowledge sources, such as domain experts, textual sources (such as articles, books, reports), and other Knowledge Organization Systems (thesauri, taxonomies, glossaries, ontologies). In short, all sources considered qualitatively able to acquire information on the domain represented by the ontology can be recognized as a potential knowledge source. The participation of a domain expert in the survey of these sources can be very useful and should be considered. Moreover, one should consider the diagnostic report done in the Pre-enrichment phase because it can provide inputs to find the knowledge sources. Furthermore, we suggest exploratory research in domain databases and repositories of KOS. After surveying all potential knowledge sources, a table with all sources, and the motive for choosing each should be generated.

Concerning *Activity b - an indication of knowledge extraction techniques*, it can range from interviews with domain experts, analysis of documents, and KOS. These techniques can be manual, semi-automatic, or automatic using linguistic techniques (Natural Language Processing mainly), statistics, and based on machine learning algorithms. This step must generate a table with the selected techniques accompanied by a motive for choice.

The knowledge source and the technique must have a strong relationship since the chosen source's nature will greatly influence the knowledge extraction technique that will be used. Having chosen the technique, the information is extracted from the selected knowledge sources. In METHODOE, we do not explain which source or technique should be used since this will depend heavily on which domain the enrichment process will be developed. We highlight that one of this methodology's characteristics is to be domain-independent. Thus, one must consider the existing sources in each specific domain.

Activity c - knowledge extraction, refers to the application of knowledge extraction techniques in the selected knowledge sources. Again, the report on the ontology

diagnostic stage is important because it will assist the specific search for knowledge to enrich the ontology. The goal here is not to extract all knowledge regarding the domain represented by the ontology, but to extract the knowledge the ontology has yet to cover. To this end, specific questions to the knowledge sources are made. These questions can be asked through interviews with domain experts, through text analysis (manually, automatically, or semi-automatically) of the domain in pursuit of answers. The manner used to extract knowledge must be associated with the type of knowledge extraction technique and the chosen knowledge source.

Stage B – Processing of extracted knowledge deals with the organization of knowledge acquired in the previous phase and comprises two activities: (a) correlation between diagnosis and extracted knowledge; (b) classification of extracted knowledge.

Activity a - correlation between diagnosis and extracted knowledge tries to relate the gaps identified in Step 1.B (Ontology diagnosis), and the possible solutions found through Activity 2.A.c (Knowledge Extraction). The objective is to facilitate the identification of possible answers to the questions and gaps the ontology presents.

Activity b - classification of extracted knowledge refers to the attempt to group the acquired knowledge into types of enrichment. The types of enrichment are: (1) lexical enrichment, which deals with the acquisition of terminological variations of a concept, synonyms, and definitions in natural language; (2) conceptual enrichment, refers to obtaining new concepts, they can be specific or general; (3) enrichment of taxonomic relations, that deals with the acquisition of 'gender-species' and 'part-of' relationships between the ontologies' concepts; (4) enrichment of non-taxonomic relations, which deals with all other types of associations that can happen between the ontologies' concepts; (5) enrichment of axioms, discusses the definition of rules to the concepts and relationships, its purpose is the formalization and consequent restriction on the interpretations of the represented knowledge. The results generated by *Activity a - correlation between diagnosis and extracted knowledge* will be relevant for the classification of knowledge into types of enrichment.

Step C - Content insertion in ontology works with the ontology expansion with the acquired content. This step must be performed by the ontologist or together with the domain specialist. The purpose is to enrich the ontology according to *Step B - Processing of the extracted knowledge's* results and correctly insert the content in the ontology. If something remains unclear, it is possible to return to the knowledge acquisition stage. At the end of this step, one must generate a detailed report with all the content inserted in the ontology, the knowledge source, and the inclusion date in the ontology.

3.3 Post-enrichment

The last phase of METHODOE consists of the following steps: (A) Evaluation of the enriched content; (B) Documentation of the methodology.

Step A - Evaluation of the enriched content verifies if the content has been properly inserted in the ontology and if there are any necessary changes to the structure or ontology component since the inserted content can have an impact on its structure. This analysis is performed with the help of the domain specialist. Finally, a document specifying the evaluation result is generated.

Stage B - Methodology documentation refers to the registration of all procedures performed in each stage of the ontology enrichment. This step should happen throughout the entire process, not merely at the end. However, only at the end of the enrichment process, will it be possible to generate the final document of the methodology.

4.0 Conclusions

We presented a novel methodology for domain ontology enrichment based on indications found in the literature related to Knowledge Organization Systems' maintenance and updating, and in empirical studies regarding domain ontology enrichment. This methodology's fundamental goal is the attempt to organize the entire domain ontologies' enrichment process. It pursues to do so presenting the pre-enrichment phase steps to be executed before the enrichment itself, and steps to be done afterwards, differing from all methods presented in literature so far. Thus, it contributes so that ontologists can have a systemic and holistic notion of how to improve existing ontologies.

METHODOE is still in its first version, so it needs to be applied in domain ontologies to be diagnosed and validated, which is one of the main limitations of this research. However, we believe the methodology presents a relevant and systematic view on the process of domain ontologies enrichment, surpassing the methods previously developed in some aspects.

As a forthcoming proposal, we intend to validate this methodology in the enrichment of two ontologies from different domains, thus proving the METHODOE's domain-independent characteristic. Furthermore, we intend to develop a practical and exemplified manual for each stage of the methodology.

References

- Al-Yahya, Maha, Sawsan Al-Malak, and Luluh Aldhubayi. 2016. "Ontological Lexicon Enrichment: The Badea System For Semi-Automated Extraction Of Antonymy Relations From Arabic Language Corpora." *Malaysian Journal of Computer Science* 29, no. 1: 56–73. <https://doi.org/10.22452/mjcs.vol29no1.5>.
- Amar, Feten Baccar Ben, Bilel Gargouri, and Abdelmajid Ben Hamadou. 2013. "Domain Ontology Enrichment Based on the Semantic Component of LMF-Standardized Dictionaries." In *Knowledge Science, Engineering and Management. KSEM 2013*, edited by M. Wang. Lecture Notes in Computer Science 8041. Berlin: Springer, 404–19.
- ANSI/NISO. 2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, MD: National Information Standards Organization.
- Barbur, Gabriel, Bogdan Blaga, and Adrian Groza. 2011. "Ontorich - A Support Tool for Semi-Automatic Ontology Enrichment and Evaluation." In *2011 IEEE 7th International Conference on Intelligent Computer Communication and Processing*, 129-132. doi: 10.1109/ICCP.2011.6047855
- Bendaoud, Rokia, Yannick Toussaint, and Amedeo Napoli. 2008. "PACTOLE: A Methodology and a System for Semi-Automatically Enriching an Ontology from a Collection of Texts." *Conceptual Structures: Knowledge Visualization and Reasoning. ICCS 2008*, edited by P. Eklund and O. Haemmerlé. Lecture Notes in Computer Science 5113. Berlin: Springer, 203–16.
- Boosshetri, Meisam, Abbas Malekpour, and Peter Luksch. 2013. "Ontology Enrichment by Extracting Hidden Assertional Knowledge from Text." *International Journal of Computer Science and Information Security* 11, no. 5: 64–72.
- Boosshetri, Meisam and Peter Luksch. 2015. "An Ontology Enrichment Approach by Using DBpedia." *WIMS '15: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*. New York: Association for Computing Machinery, 1–11. <https://doi.org/10.1145/2797115.2797127>
- Carvalho, Miguel. G. P., Vanessa Braganholo, Maria L.M. Campos, and Maria L.A. Campos 2010. "Enriquecimento de Ontologias: uma Abordagem para Extração de Conhecimento do Campo Definição". Presented at *Ontobrás*. Florianópolis, Santa Catarina.
- Faatz, Andreas and Ralf Steinmetz. 2002. *Ontology Enrichment with Texts from the WWW*.
- Faatz, Andreas, Stefan Hermann, Cornelia Seeberg, and Ralf Steinmetz. 2001. *Conceptual Enrichment of Ontologies by Means of a Generic and Configurable Approach*.

- Gómez-Moreno, Pedro Ureña and Eva M. Mestre-Mestre. 2017. "Automatic Domain-specific Learning: Towards a Methodology for Ontology Enrichment." *Revista de Lenguas para Fines Específicos* 23, no. 2: 63-85.
- Guerram, Tahar and Nacima Mellal. 2018. "A Domain Independent Approach for Ontology Semantic Enrichment." In *Computer Science & Information Technology*, edited by Natarajan Meghanathan et al., 13-19 <https://doi.org/10.5121/csit.2018.80202>
- Hashimy, Amaal Saleh Hassan Al, and Narayanan Kulathuramaiyer. 2013. "Ontology Enrichment with Causation Relations." In *2013 IEEE Conference on Systems, Process & Control (ICSPC)*. Kuala Lumpur, 186-192. doi: 10.1109/SPC.2013.6735129.
- International Organization for Standardization (ISO). 2011. *ISO 25964 -1: Thesauri for information retrieval*. Geneva: International Standard Organization.
- International Organization for Standardization (ISO). 2013. *ISO 25964 -2. Interoperability with other vocabularies. International Standard Organization, Geneva*.
- Kim, Chai. 1973. "Theoretical Foundations of Thesaurus-Construction and Some Methodological Considerations for Thesaurus-Updating." *Journal of the American Society for Information Science* 24, no. 2: 148–56.
- Navigli, Roberto and Paola Velardi. 2006- "Ontology Enrichment Through Automatic Semantic Annotation of On-Line Glossaries." In *Managing Knowledge in a World of Networks*, edited by S. Staab and V. Svátek. Berlin: Springer, 126–140.
- Petasis, Georgios, Vangelis Karkaletsis, Georgios Paliouras, Anastasia Krithara, and Elias Zavitsanos. 2011. "Ontology Population and Enrichment: State of the Art." In *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, edited by G. Paliouras, C.D. Spyropoulos, and G. Tsatsaronis. Lecture Notes in Computer Science 6050. Berlin: Springer, 134–66. https://doi.org/10.1007/978-3-642-20795-2_6.
- Poveda-Villalón, María. 2016. "Ontology Evaluation: A Pitfall-Based Approach to Ontology Diagnosis." Ph.D. dissertation. Madrid: Universidad Politécnica de Madrid. http://oa.upm.es/39448/1/MARIA_POVEDA_VILLALON.pdf.
- Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. U.S.A.: Melville Pub. Co.
- Valarakos, Alexandros G., Georgios Paliouras, Vangelis Karkaletsis, and George Vouros. 2004. "A Name-Matching Algorithm for Supporting Ontology Enrichment." In *Methods and Applications of Artificial Intelligence. SETN 2004*, edited by G.A. Vouros and T. Panayiotopoulos T. Lecture Notes in Computer Science, vol 3025. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24674-9_40