Entre Ciencia e Ingeniería, vol. 19, no. 38, julio-diciembre de 2025, páginas 16-22 DOI: https://doi.org/10.31908/19098367.3258
ISSN 1909-8367 (Impreso), ISSN 2539-4169 (En línea)

A proposal of knowledge representation about ontology population from texts written in natural language[1]

Propuesta de representación del conocimiento para la población de ontologías desde textos escritos en lenguaje natural

J. C. Blandón and C.M. Zapata.

Recibido: octubre 9 de 2025 – Aceptado: noviembre 19 de 2025

***Abstract*—Ontologies are referred to shared conceptualizations with several elements—*e.g.*, domains, concepts, instances, relations, and attributes—to be used in computational inference. Ontology population is referred to both extraction and classification of instances belonging to classes and relations defined in the ontology from any information sources. State-of-the-art review exhibits some proposals for representing ontology population, but they are incomplete and exhibit some drawbacks. Ontology population process is hard, and it requires a good knowledge representation. In this paper we propose a pre-conceptual schema for representing ontology population. We evaluate knowledge representation by using five fundamental roles: i) as a surrogate of reality; ii) as a set of ontological commitments; iii) as a fragmentary theory of intelligent reasoning; iv) as a medium for efficient computation; and v) as a medium of human expression. This contributes to a better understanding of the ontology population process and demonstrates the benefits of using a pre-conceptual schema to enhance clarity, reduce ambiguity, and organize knowledge. Finally, it contributes to the main objective of ontologies, *i.e.*, sharing knowledge.**

***Keywords*— Requirements analysis, pre-conceptual schemas, information extraction, instances, ontology population, knowledge representation.**

[1]Producto derivado del proyecto de investigación "Extracción de instancias de una clase desde textos en lenguaje natural independientes del dominio de aplicación", apoyado por la Universidad Católica de Pereira y Universidad Nacional de Colombia Sede Medellín.

J. C. Blandón, Universidad Católica de Pereira, Pereira, Colombia, email: juanc.blandon@ucp.edu.co.

C.M. Zapata, Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia, email: cmzapata@unal.edu.co.

Como citar este artículo: Blandón, J.C., Zapata, C.M. A proposal of knowledge representation about ontology population from texts written in natural language, Entre Ciencia e Ingeniería, vol. 19, no. 38 pp. 16-22, julio-diciembre 2025. DOI: https://doi.org/10.31908/19098367.3258.

***Resumen*— Las ontologías se refieren a conceptualizaciones compartidas y cuentan con varios elementos, por ejemplo, dominios, conceptos, instancias, relaciones y atributos, que se utilizan para hacer inferencias computacionales. La población de ontologías se refiere a la extracción y clasificación de instancias pertenecientes a clases y relaciones definidas en la ontología desde cualquier fuente de información. La revisión de la literatura exhibe algunas propuestas para representar la población de ontologías, pero son incompletas y presentan algunas desventajas. El proceso de población de ontologías es complejo y requiere una buena representación del conocimiento. En este artículo se propone un esquema preconceptual para representar la población de ontologías. Se evalúa la representación del conocimiento mediante cinco roles fundamentales: i) como un sustituto de la realidad; ii) como un conjunto de compromisos ontológicos; iii) como una teoría fragmentaria del razonamiento inteligente; iv) como un medio para la computación eficiente; y v) como un medio de expresión humana. Esto contribuye a una mejor comprensión del proceso de población de ontologías y demuestra los beneficios de usar un esquema preconceptual para mejorar la claridad, reducir la ambigüedad y organizar el conocimiento. Finalmente, contribuye al objetivo principal de las ontologías, compartir conocimiento.**

***Palabras clave*— Análisis de requisitos, esquemas preconceptuales, extracción de información, instancias, población de ontologías, representación del conocimiento.**

I. INTRODUCCIÓN

ONTOLOGY population is a process for adding new instances of both concepts and relations into an existing ontology, typically by identifying objects/instances within a corpus. The ontology population process keeps the structure of the ontology [1], [2]. Knowledge representation refers to the way knowledge of a topic/domain is symbolically represented and automatically manipulated by programs in order to perform knowledge inference [3],[4]. Another perspective on knowledge representation considers it at an ontological level, as this logic can be used to describe the real world; accordingly, conceptual graphs are suggested for graphical representation [5]. Zapata *et al.* [6] propose pre-conceptual schemas to represent knowledge, defining them as intermediate stages between natural language and UML conceptual schemas, while also serving as a formal mechanism for knowledge representation.

Although this research field is still under development, concerns have emerged regarding the improvement of ontology population processes. Some studies highlight instance extraction as a critical step, as instances can be used to identify concepts/classes still absent from the ontology. Schlaf and Remus [7] present a framework for learning categories and their corresponding instances through the use of contextual features. Varges and Mellish [8] demonstrate how to generate instances using the IGEN architecture, employing search methods known as "search expectation-driven" and "dynamic grammar rule selection". Davis *et al.* [9] present an information extraction application for a personal semantic wiki, aimed at enabling users without experience in ontology tools or semi-automatic languages to annotate their own wiki pages. Pan and Shaw [10] develop a hybrid system that combines a case-based paradigm with rule-based adaptations; they employ both an annotated corpus and grammatical rules.

Although the systems in the literature adopt some form of knowledge representation, they lack a fully adequate model for forward ontology population and instance extraction, particularly when we evaluate it in relation to the five fundamental roles of knowledge representation defined by Davis *et al.* [3]— A surrogate for reality, a set of ontological commitments, a fragmentary theory of intelligent reasoning, a medium for efficient computation, and a vehicle for human expression. These approaches reveal critical shortcomings, highlighting the need for a more comprehensive solution.

In this paper, we aim to identify the main limitations of the previous approaches for acting ontology population and, so, we propose a novel form of knowledge representation. We introduce pre-conceptual schemas as a new framework for acting the ontology population process in order to address such shortcomings, using their ability to reduce ambiguity and ensure semantic consistency.

Pre-conceptual schemas are intermediate representations between textual specifications in natural language (“verbal models”) and the conceptual schemas commonly used for we model software applications [6]. Their expressive and structural capabilities make them well-suited for supporting the extraction and organization of instances from natural language texts into ontology structures.

Extracting class instances from natural language texts is a crucial step in ontology population because it ensures that relevant concepts and relationships are formally integrated into the ontology. This process is more systematic and less given to ambiguity because it uses pre-conceptual schemas as an intermediate representation, thereby its facilities the transformation of textual information into structured knowledge. In this way, the proposed approach reinforces the theoretical roles of knowledge representation defined by Davis *et al.* [3], and it provides practical benefits for knowledge organization, reuse, and sharing.

This paper is organized as follows: in Section II, we present the related work; in Section III, we present the problem; in Section IV, we present the proposed solution; in Section V, we discuss the validation of our proposal. Finally, in Section VI, we discuss the conclusions.

II. Related Work

A. *Theoretical Framework*

a) Information Extraction

Information extraction (IE) is related to collecting texts and altering them into structured information that can be more easily understood and analyzed. The relevant text fragments are identified, the pertinent information is extracted, and this information is organized into a coherent structure. Crucial information contained within documents is recognized and converted into a predefined format for process and retrieval [11].

The goal of IE research is to develop systems capable of identifying items of interest for human analysis from documents. Relevant information should be accurately extracted, while irrelevant/extraneous content is disregarded. IE systems typically operate on specific types of texts and have so far achieved partial results [12].

Information extraction (IE) is focused on structured information contained within texts that is relevant to the study of a particular domain, known as the extraction domain. The objective of an IE system is to identify and connect the relevant information, while disregarding any remaining irrelevant content [12].

b) Pre-conceptual Schemas

Pre-conceptual schemas are intermediate representations bridging textual specifications in natural language (“verbal models”) and the various conceptual schemas used for modeling software applications. Pre-conceptual schemas constitute the initial attempt to model the problem domain and are considered work products of the definition phase [6]. They are valuable for understanding the problem to be solved and are also employed by analysts to validate their comprehension of the problem [6]. The primary symbols of pre-conceptual schemas are illustrated in Fig.1.

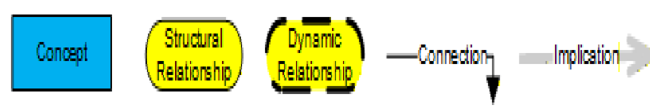

Fig. 1. Symbols of pre-conceptuals schemas adapted from [6].

Concepts are represented with rectangles, which can contain nouns phrases. Structural relationships model the static organization of the domain, showing the composition and permanent links between concepts. Dynamic relationships model the behavior and transformation of the system. A concept-to-relationship connection indicates that a concept performs the activity expressed by the relationship. Conversely, a relationship-to-concept connection indicates that a concept receives the activity expressed by the relationship. Thick arrows represent implication. These elements can be connected only from one relationship to another, meaning that the target activity is performed only if the source activity has been executed, functioning as a precondition in if-then constructs [6].

B. Background

Some studies have been conducted on natural language processing. Schlaf and Remus [7] identify instances by using contextual categories and propose a three-step framework. The first step involves extracting content by using learning rules. In the second step, high-quality rules are selected by using two consecutive filters: the first is based on the number of rule occurrences, and the second considers two interdependent features—a rule precision and the number of instances it extracts. In the third step, instances of a new category are identified by using the filtered rules from step two (see Fig.2).

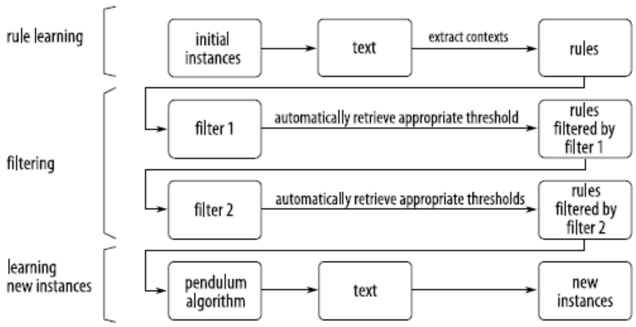

Fig. 2. Representation of the framework [7].

They report an average accuracy of 0.849, with four categories successfully learned: first name, last name, profession, and city. Additionally, as future work, they propose expanding the approach to learn additional categories and to incorporate other sources such as newspapers, articles, blogs, and similar textual materials.

Varges and Mellish [8] explore methods for generating instances, adopting an approach that differs from purely statistical techniques. Their method is hybrid, combining rule-based grammar with a corpus-based ranker. The fundamental idea relies on applying rules to create candidate instances and ranking such candidates based on corpus data in order to generate appropriate words. The authors note that the system performs well for short sentences (see Fig.3).

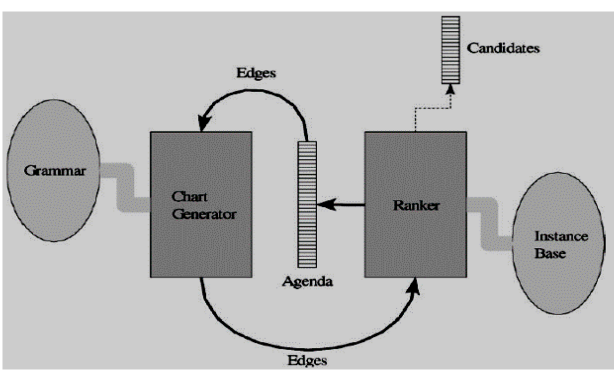

Fig. 3. Architecture of the instance-based generator [8].

Davis *et al.* [9] employ Controlled Language Information Extraction (CLIE). The following components are required for ontology generation: a tokenizer, a part-of-speech tagger, a morphological analyzer, finite-state transducers for identifying quoted strings and chunking noun phrases, and a gazetteer list for recognizing relevant key phrases. After the initial preprocessing, the parser searches for sentences containing the predefined types of key phrases. Sentences that mismatch any key phrases are then used as names for generating ontological objects (see Fig.4).

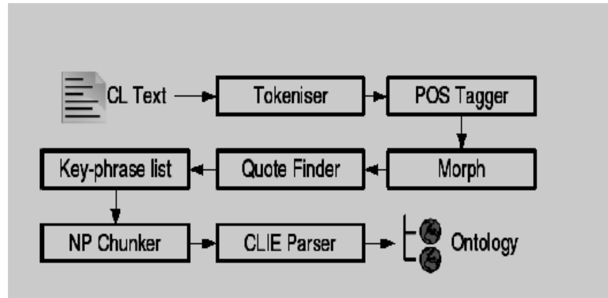

Fig. 4. Description of CLIE [9].

Pan and Shaw [10] present a study on natural language generation called SEGUE. The authors use an annotated corpus as a knowledge base and generate new phrases through rule-based methods. They also create semantic representations from plain text, where each training sentence in the repository is associated with a corresponding sentence representation (SEMGRAPH). Subsequently, a tree structure (REATREE) is constructed to represent the syntax and lexical structure of the sentence. SEGUE operates in three phases: retrieval, adaptation, and learning. In the first phase, SEMGRAPH is used to identify a ranked list of candidates for adaptation. In the second phase, one or more adaptation operators are applied to the REATREE corresponding to the SEMGRAPH identified in the previous phase, resulting in an adapted REATREE. In the third phase, all information is verified and stored in the general repository, where the generated sentences are maintained (see Fig.5).

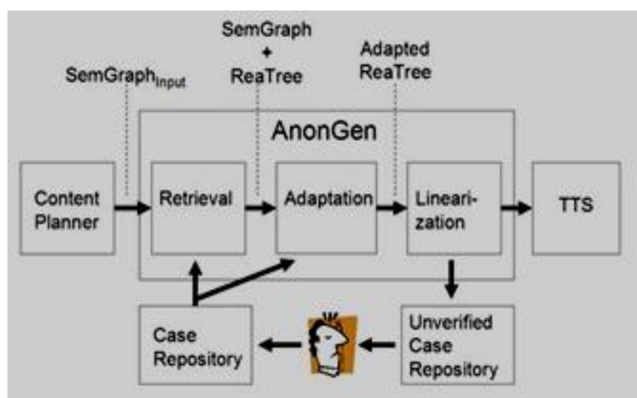

Fig. 5. Architecture of SEGUE [10].

Muscetti *et al.* [13] propose an approach for populating multimedia ontologies by employing a Convolutional Neural Network (CNN), specifically VGG16, to extract both global and local visual descriptors, which are subsequently combined with semantic analysis using WordNet and textual labels. The data are sourced from Google Images and the Common Objects in Context (COCO) dataset, and a hierarchical process is applied for first classifying images into general concepts and then into more specific categories. Finally, we incorporate the images as instances within the ontology nodes, allowing the evaluation of both accuracy and coverage. Although the method reaches promising results on controlled datasets, it exhibits limitations regarding scalability and its reliance on external resources for semantic alignment. The architecture is illustrated in Fig.6.

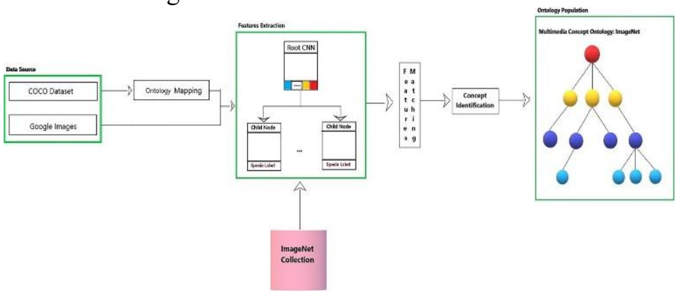

Fig. 6. Architecture of VGG16 [13].

III.PROBLEM

The main challenge in ontology population from natural language texts rests in information extraction and in the limitations of the knowledge representations employed by existing approaches. These representations often lack the formality and clarity, meaning to semantic ambiguity, inconsistencies, and difficulties in reuse [14], [15]. Besides, the literature has shown that the limited interpretability and expressiveness of such models constrain expert validation [16] and that the lack of mechanisms to adequately address the variability of natural language undercuts the robustness of the ontologies [17]. Even when extraction processes achieve satisfactory results, the weaknesses in representation compromise both the efficiency and scalability of the systems [18].

In this context, pre-conceptual schemas emerge as a promising alternative because they offer an intermediate graphical representation that eases semantic comprehension, enhances communication with users, and supports the subsequent formalization of ontologies [19], [20], [21]. In Table I, the different representations are shown in relation to the roles established by Davis *et al*. [3].

TABLA I
COMPARISON AMONG KNOWLEDGE REPRESENTATION PROPOSALS

Role / Author	Surrogate	Ontological commitment	Reasoning theory	Efficient computation	Human expression
Schlaf and Remus [7]	X	X		X	
Varges and Mellish [8]	X		X	X	
Davis *et al.* [9]	X	X			X
Pan and Shaw [10]			X	X	
Muscetti *et al.* [13]	X	X		X	

IV. PROPOSAL

A proposed solution for representing the knowledge about instance extraction from texts written in natural language is proposed in Fig.7. The extraction process is based on natural language (text, images, audio) [13].

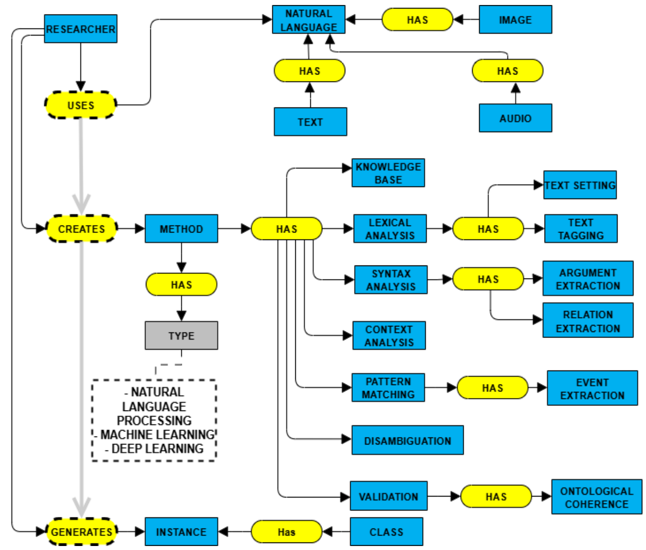

Fig. 7. Knowledge representation for the instance extraction from texts.

The method integrates one/more types of techniques, including Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL). Within the method, a knowledge base is then constructed to serve as a repository of conceptual fragments that can be reused to build new models [8]. During lexical analysis, the features of the text are identified, and words and phrases are tagged as nouns, verbs, and adjectives by using morphological analysis and dictionary lookup. For instance, the names “Fernando” and “Pablo” would be labeled as “proper names”.

Syntactic analysis involves identifying noun groups, verb groups, and, when it is required, head structures. This process enables the extraction of noun phrase arguments and their functional grammatical relationships. Contextual analysis represents the semantic environment that provides meaning to entities and relationships—*i.e.*, semantic consistency [7]. Afterwards, we apply task-specific patterns to identify the facts of interest [12]. Extraction pattern matching uses the roles determined by these patterns to extract events or relevant relations within the scene.

The disambiguation process is performed either during or after extraction. Information from multiple occurrences of the same object can be consolidated, and different objects sharing attribute values are disambiguated [9]. The validation method functions as a mechanism for assessing the coherence of the extracted knowledge [10]. As a result, the processed language contains the instances of the various classes present in the original document.

Finally, the knowledge representation based on pre-conceptual schemas is an approach to support ontology population. It specifies the essential concepts required to enable generic instance extraction and incorporates mechanisms to account for potential ambiguities.

V. VALIDATION

Our proposed approach to knowledge representation is grounded in pre-conceptual schemas [6], which constitute a formal graphical language designed to model concepts, relationships, and semantic constraints. Unlike other approaches that present limitations in terms of clarity, consistency, and interpretability, the validation demonstrates that the resulting pre-conceptual schema satisfactorily fulfills the five fundamental roles of knowledge representation defined by Davis *et al*. [3]. Consequently, it stands as a robust and semantically coherent alternative for representing knowledge.

A. Surrogate

The pre-conceptual schema precisely represents the entities, properties, and relationships of the domain, thereby ensuring semantic correspondence between the knowledge expressed in natural language texts and its graphical formalization. This approach minimizes information loss and addresses the semantic misalignment commonly found in textual or statistical representations [14]. Moreover, the validation confirms that the graphical constructs of the schema effectively capture not only the static concepts of the domain but also the dynamic processes and dependencies described in the texts, thus ensuring broader knowledge coverage. Consequently, the schema functions not merely as a reflection of the domain but as a verifiable bridge between natural language and formal ontology.

B. Ontological commitment

The domain commitments are explicitly established within the pre-conceptual schema, as it clearly defines which concepts and relationships are relevant, thus avoiding terminological ambiguities. This clarity in the commitments ensures the coherence of the model and aligns with the ontological quality criteria recognized as essential for reusability and maintainability [15]. Therefore, each pre-conceptual schema can be evaluated based on the robustness of its categories, the consistency of its relationships, and the absence of semantic contradictions. Making these commitments explicit enables the comparison of schemas across different domains or projects, thereby enhancing their interoperability potential and their value as reusable knowledge representations.

C. Theory of intelligent reasoning

The pre-conceptual schema incorporates structural patterns that enable the derivation of inferences, such as hierarchies, dependencies, and roles. This reasoning capability distinguishes it from other representations that are merely descriptive, thereby meeting the interpretability criterion identified as essential in critical domains [22]. The validation shows that, starting from a pre-conceptual schema, it is possible to identify inference paths that support ontology population and allow the verification of logical consistency among the elements. Consequently, the schema becomes a valuable resource for both knowledge storage and explanatory and predictive analysis within a domain.

D. Efficient computation

The nature of the pre-conceptual schema guarantees its direct translation into computable ontological languages such as OWL and RDF, optimizing the automation of instance population. Validation demonstrates that it reduces computational complexity and improves scalability compared to other models [18]. Furthermore, the use of a formal notation allows for the standardization of conversion processes and minimizes integration errors when interacting with automated reasoning tools. This means that the schema functions both as an independent representation and as a catalyst for efficiency throughout the entire ontology population and exploitation chain, ensuring reproducibility and robustness in large-scale environments.

E. Human expression

The graphical notation of pre-conceptual schemas is clear, readable, and easily validated by domain experts. Unlike more technical or less intuitive representations, this model promotes interdisciplinary collaboration and collective validation, thereby fulfilling the role of human expressivity described in the literature [17]. The results show that participants without formal training can interpret the schema, which speeds up the review and acceptance processes of the resulting ontologies. Therefore, pre-conceptual schemas establish a balance between formal rigor and comprehensibility, demonstrating that they play an important role as a reliable and collaborative validation tool.

F. *Correspondence of conceptual components of the pre-conceptual schema with other approaches.*

According to the validation, the pre-conceptual schema confirms that it integrates within its structure the main conceptual components identified in other knowledge representation models. Each part of the proposed schema addresses aspects previously explored by different authors; however, in our approach, these elements are coherently articulated within a single, formal, and consistent representation. Table II presents the correspondence between the elements of the pre-conceptual schema and other models.

TABLA II
CORRESPONDENCE AMONG THE ELEMENTS OF THE PRE-CONCEPTUAL SCHEMA AND OTHER PROPOSALS

Pre-conceptual Schema Element	Proposal Authors	Equivalence
TEXT HAS NATURAL LANGUAGE	Schlaf and Remus [7]	TEXT
KNOWLEDGE BASE	Schlaf and Remus [7]	RULES
CLASS HAS INSTANCE	Schlaf and Remus [7]	NEW INSTANCES
SYNTAX ANALYSIS, LEXICAL ANALYSIS	Varges and Mellish [8]	GRAMMAR
CLASS HAS INSTANCE	Varges and Mellish [8]	INSTANCE BASE
PATTERN MATCHING	Varges and Mellish [8]	RANKER
TEXT HAS NATURAL LANGUAGE	Davis *et al.* [9]	CL TEXT
SYNTAX ANALYSIS, LEXICAL ANALYSIS	Davis *et al.* [9]	TOKENISER, POS TAGGER, MORPH, QUOTE FINDER
ONTOLOGICAL COHERENCE	Davis *et al.* [9]	ONTOLOGY

TABLA II
CORRESPONDENCE AMONG THE ELEMENTS OF THE PRE-CONCEPTUAL SCHEMA AND OTHER PROPOSALS

Pre-conceptual Schema Element	Proposal Authors	Equivalence
CONTEXT ANALYSIS	Pan and Shaw [10]	SEMGRAPH + REATREE
KNOWLEDGE BASE	Pan and Shaw [10]	CASE REPOSITORY
METHOD	Pan and Shaw [10]	CONTENT PLANNER
IMAGE HAS NATURAL LANGUAGE	Muscetti *et al.* [13]	COCO DATASET, GOOGLE IMAGE
ONTOLOGICAL COHERENCE	Muscetti *et al.* [13]	ONTOLOGY MAPPING
CLASS HAS INSTANCE	Muscetti *et al.* [13]	ONTOLOGY POPULATION

The validation results confirm that pre-conceptual schemas fully satisfy the five roles proposed by *Davis et al.* [3] and overcome the limitations observed in other representations. They therefore emerge as an effective means of ensuring semantic fidelity, ontological clarity, explainable reasoning, computational efficiency, and human expressivity in ontology population from natural language texts. When evaluated from different perspectives, such schemas demonstrate comprehensive robustness, positioning them as a viable alternative and a coherent and effective form of knowledge representation.

VI. CONCLUSIONS

We proposed in this paper a representation approach to enrich the ontology population process from natural language texts using an intermediate representation model known as pre-conceptual schema. Such schemas ease semantic understanding, enable early validation of extracted information, and provide a starting point for users of ontological formalisms. The results indicate that they constitute an alternative for reducing the complexity of traditional approaches and improving the quality of populated ontologies.

Some limitations are related to the model dependence on the linguistic quality of source texts, which may lead to ambiguities/inconsistencies in instance extraction. Moreover, the proposal requires advanced automation mechanisms and scalability to operate effectively with large volumes of heterogeneous data.

Despite these limitations, our model demonstrates potential in domains that require explicit knowledge representation and validation, such as education, healthcare, document management, and decision support systems. Collaboration between technical and domain experts is enhanced using pre-conceptual schemas, which accelerate ontology construction and improve its relevance in real-world environments.

As future work, the integration of advanced Natural Language Processing techniques and large-scale language models is proposed to increase the accuracy and scalability of instance extraction. Additionally, exploring the interoperability of pre-conceptual schemas with formal ontological representation languages such as OWL or RDF would enable more expressive and standardized implementations. Finally, it is recommended to validate the model across different application domains to consolidate its practical utility.

REFERENCES

[1] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, y E. Zavitsanos, «Ontology Population and Enrichment: State of the Art», en *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, G. Paliouras, C. D. Spyropoulos, y G. Tsatsaronis, Eds., en Lecture Notes in Computer Science, no. 6050. , Springer Berlin Heidelberg, 2011, pp. 134-166.

[2] J. C. Blandón Andrade y C. M. Zapata Jaramillo, «Gate-Based Rules for Extracting Attribute Values», *Computación y Sistemas*, vol. 25, n.º 4, Art. n.º 4, dic. 2021, doi: 10.13053/cys-25-4-3493.

[3] R. Davis, H. Shrobe, y P. Szolovits «What is a knowledge representation?», *AI magazine*, vol. 14, n.º 1, p. 17, 1993.

[4] M. Gelfond y Y. Kahl, *Knowledge Representation, Reasoning, and the Design of Intelligent Agents: The Answer-Set Programming Approach*. Cambridge University Press, 2014.

[5] J. F. Sowa, *Knowledge representation: logical, philosophical, and computational foundations*. Brooks Cole Publishing Co, 2000. Accedido: 26 de febrero de 2017. [En línea]. Disponible en: http://www.mitpressjournals.org/doi/pdf/10.1162/089120101750300544

[6] Carlos. M. Zapata, A. Gelbuk, y F. Arango, «Pre-conceptual schema: a UML isomorphism for automatically obtaining UML conceptual schemas», *Research in Computing Science: Advances in Computer Science and Engineering*, vol. 19, n.º 2006, pp. 3--13, 2006, doi: 10.1.1.152.9690.

[7] A. Schlaf y R. Remus, «Learning Categories and their Instances by Contextual Features», en *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC*, Istanbul, 2012, pp. 1235-1239.

[8] S. Varges y C. Mellish, «Instance-based natural language generation», *Natural Language Engineering*, vol. 16, n.º 03, pp. 309-346, 2010, doi: 10.1017/S1351324910000069.

[9] B. Davis, S. Handschuh, H. Cunningham, y V. Tablan, «Further use of controlled natural language for semantic annotation of wikis», en *Proceedings of the 1st Semantic Authoring and Annotation Workshop at ISWC2006, Athens, Georgia, USA*, 2006. Accedido: 16 de mayo de 2013. [En línea]. Disponible en: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-209/saaw06-short06-davis.pdf

[10] S. Pan y J. Shaw, «SEGUE: A Hybrid Case-Based Surface Natural Language Generator», en *Natural Language Generation*, Springer Berlin Heidelberg, 2004, pp. 130-140. doi: 10.1007/978-3-540-27823-8_14.

[11] J. C. Blandón A., «Extracción de instancias de una clase desde textos en lenguaje natural independientes del dominio de aplicación», PH.D. Thesis, Universidad Nacional de Colombia, Medellín, Colombia, 2017. [En línea]. Disponible en: https://repositorio.unal.edu.co/handle/unal/58998#

[12] J. Cowie y W. Lehnert, «Information extraction», *Communications of the ACM*, vol. 39, n.º 1, pp. 80-91, 1996.

[13] M. Muscetti, A. M. Rinaldi, C. Russo, y C. Tommasino, «Multimedia ontology population through semantic analysis and hierarchical deep features extraction techniques», *Knowledge and Information Systems*, vol. 64, n.º 5, pp. 1283-1303, may 2022, doi: 10.1007/s10115-022-01669-6.

[14] W. Wong, W. Liu, y M. Bennamoun, «Ontology learning from text: A look back and into the future», *ACM Computing Surveys*, vol. 44, n.º 4, p. 20:1-20:36, 2012, doi: 10.1145/2333112.2333115.

[15] R. S. I. Wilson, J. S. Goonetillake, W. A. Indika, A. Ginige, y A. Gangemi, «A conceptual model for ontology quality assessment: A systematic review», *Semantic Web*, vol. 14, n.º 6, pp. 1109-1135, 2023, doi: 10.3233/SW-233393.

[16] B. Msheik, M. Adda, H. Mcheick, y M. Dbouk, «Survey on Knowledge Representation Models in Healthcare», *Information*, vol. 15, n.º 8, p. 435, 2024, doi: 10.3390/info15080435.

[17] S. Kim, Y. Y. Liau, y K. Ryu, «Ontology Modeling Using Fractal and Fuzzy Concepts to Optimize Metadata Management», *Applied Sciences*, vol. 15, n.º 13, p. 7193, 2025, doi: 10.3390/app15137193.

[18] A. C. Faria, G. Guizzardi, y J. P. A. Almeida, «Evaluating Ontology Representation and Reasoning in Conceptual Modeling Approaches»,

Data & Knowledge Engineering, vol. 146, p. 102175, 2023, doi: 10.1016/j.datak.2022.102175.

[19] C. M. Zapata-Jaramillo, A. Gelbukh, y F. A. Isaza, «Pre-conceptual Schema: A Conceptual-Graph-Like Knowledge Representation for Requirements Elicitation», en *Mexican International Conference on Artificial Intelligence (MICAI)*, en Lecture Notes in Computer Science, vol. 4293. Springer, 2006, pp. 27-37. doi: 10.1007/11925231_4.

[20] C. M. Zapata, G. L. Giraldo, y S. Londoño, «Esquemas preconceptuales ejecutables», *Revista Avances en Sistemas e Informática*, vol. 8, n.º 1, pp. 15-24, 2011.

[21] C. M. Zapata-Jaramillo y D. M. Torres-Ricaurte, «Test effort: A pre-conceptual-schema-based representation», *DYNA*, vol. 81, n.º 186, pp. 132-137, 2014.

[22] B. Msheik, M. Adda, H. Mcheick, y M. Dbouk, «Survey on Knowledge Representation Models in Healthcare», *Information*, vol. 15, n.º 8, p. 435, 2024, doi: 10.3390/info15080435.

Juan Carlos Blandón Andrade. System Engineer. M.Sc. in Engineering focused on systems and computation, Javeriana University Cali Colombia. Ph.D. in Engineering—systems and computing at Universidad Nacional de Colombia Medellín. Associate Professor at Universidad Católica de Pereira, Colombia. His research interests are focused in Software Engineering, Artificial Intelligence focused on Natural Language Processing and Pedagogy in Engineering.
ORCID: https://orcid.org/0000-0003-1566-1832

Carlos Mario Zapata Jaramillo. Civil Engineer, Information System Management Specialist, M.Sc. in System Engineering, Ph.D. in Engineering focused on Systems, all the titles from the Universidad Nacional de Colombia. Full Professor at Computer and Decision Science Department, Faculty of Mines, Universidad Nacional de Colombia, Medellín, Colombia. He is, moreover, President of the Executive Committee of the Latin American Chapter of Semat and one of the official translators of the book “The Essence of Software Engineering: applying the Semat kernel.” His research interests are focused on Software Engineering, Requirements Engineering, Computational Linguistics and Didactical Strategies for Teaching Engineering.
ORCID: https://orcid.org/0000-0002-0628-4097