Assessment of semantic similarity in entities under monitoring: A systematic literature mapping

Evaluación de similaridad semántica en entidades bajo monitoreo: Un mapeo sistemático de la literatura

María Laura Sánchez Reynoso [1*] Mario José Diván [1]

[1] Grupo de Investigación en Ciencia de Datos, Escuela de Economía, Universidad Nacional de La Pampa. Coronel Gil 353, 1er piso. C. P. 6300. Santa Rosa (La Pampa), Argentina.

CITE THIS ARTICLE AS:

M. L. Sánchez and M. J. Diván. "Assessment of semantic similarity in entities under monitoring: A systematic literature mapping", *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 99, pp. 21-31, Apr-Jun 2021. [Online]. Available: `https://www.doi.org/10.17533/udea.redin.20200476`

ARTICLE INFO:

Received: October 23, 2019
Accepted: May 07, 2020
Available online: May 07, 2020

KEYWORDS:

Survey analysis; monitoring; semantic analysis; measurement; data processing

Análisis de inspección; monitoreo; análisis semántico; medición; procesamiento de datos

ABSTRACT: The detection and evaluation of semantically similar entities in measurement projects is a key asset for real-time decision making because it allows reusing their knowledge and previous experiences. In this way, the objective of this work is to map the thematic area of data stream processing to identify the topics that have been investigated in the detection of semantically similar entities. From the methodological point of view, a systematic mapping study was conducted obtaining 2,122 articles. Thus, 111 were kept refining the search strategy, and 25 were considered once the filters were applied jointly with the inclusion/exclusion criteria. After reading the 25 documents, just 6 were pertinent and allowed answering the research questions aligned with the research objective. The semantic similarity applied to entities under monitoring in the measurement and evaluation projects is a challenge. Real-time decision making depends on the obtained measures, the monitored entity, and the context in which it is immersed.

RESUMEN: La detección y evaluación de entidades semánticamente similares en proyectos de medición es un activo clave para la toma de decisiones en tiempo real, ya que permite reutilizar el conocimiento y experiencias previas. De este modo, el objetivo de este trabajo es mapear el área temática del procesamiento de flujos de datos para identificar los temas que han sido investigados en la detección de entidades semánticamente similares. Desde el punto de vista metodológico, se realizó un estudio de mapeo sistemático obteniendo 2.122 artículos. De esta manera, 111 se mantuvieron afinando la estrategia de búsqueda, y 25 se consideraron una vez que los filtros se aplicaron conjuntamente con los criterios de inclusión/exclusión. Después de leer los 25 documentos, sólo 6 fueron pertinentes y permitieron responder a las preguntas de investigación alineadas con el objetivo de la investigación. La similitud semántica aplicada a las entidades bajo monitoreo en los proyectos de medición y evaluación es un desafío. La toma de decisiones en tiempo real depende de las medidas obtenidas, de la entidad vigilada y del contexto en el que se encuentre inmersa.

1. Introduction

Measurement and evaluation (M&E) activities are considered key in terms of knowledge of the status of an entity under monitoring [1]. The M&E are a mechanism through which it is possible to identify the behavior related to an entity under analysis, detecting early deviations and/or anomalies regarding its expected behavior [2].

* Corresponding author: María Laura Sánchez Reynoso
E-mail: mlsanchezreynoso@eco.unlpam.edu.ar
ISSN 0120-6230
e-ISSN 2422-2844

It is necessary to consider the use of a formal measurement and evaluation (M&E) framework that allows the definition of concepts, terms, and the required relationships to understand the M&E process, which will provide a quantitative point of view about the monitored entity [3]. The entity to be monitored could be tangible (e.g. a person) or abstract (e.g. a system); for that reason, it is important to early describe it for the identification of properties or characteristics feasible of quantification. However, such description or characterization should be agreed by stakeholders to foster the measurement process automatization. The measurement and evaluation framework C-INCAMI (Context - Information Need, Concept Model, Attribute, Metric, and Indicator) [2, 4] is a conceptual framework with an underlying ontology, which defines the concepts, terms, jointly with the necessary relationships useful for specifying an M&E process. In this way and through its use, it is possible to get a common understanding related to which the entity represents (e.g. a person, a system, a vehicle, etc.), its characterization through attributes, the metrics responsible for the quantification of each attribute, and the indicators as a mechanism for interpreting each measure using the decision criteria. Thus, there is a project definition organized under a given framework, that allows defining the entity, describing it through a discrete perspective (i.e. characteristics or attributes) useful for quantifying a point of view. The quantification is carried forward using metrics which contains the association among the attribute to be analyzed (e.g. the value of the corporal temperature), the device to be used (e.g. a digital thermometer), and the definition about the range of expected values, units, scale, jointly with the indicator definition, responsible for interpreting each value following decision criteria.

Therefore, each measurement and evaluation project can be defined in terms of the C-INCAMI framework, using the GOCAME strategy as a guide (Goal-Oriented Context-Aware Measurement and Evaluation) [5]. The result is an eXtensible Markup Language (XML) file organized under the concepts defined in the framework, which is readable and understandable by humans and machines, to foster the measurement project interoperability and its implementation (i.e. automatization)[6]. In general, the object or subject to be monitored are referred to as an entity, because the meaning and attributes are unknown up to the project definition is read. In such a moment, for example, it is possible to know whether the entity is a person or a system. In other words, the measurement Project definition (i.e. a simple XML file) is a self-contained file that specifies the entity, their attributes, metrics to quantify each attribute, decision criteria used for interpreting the measures, among other aspects.

The data stream processing strategy (DSPS) is a data stream engine based on Apache Storm (`https://storm.apache.org`), which automatizes an M&E project defined in terms of the C-INCAMI framework. It allows integrating heterogeneous data sources using technologies such as IoT, and to make decisions in real-time. DSPS uses a case-based Organizational Memory (OM) which stores the previous experiences and knowledge related to the entities under monitoring. In this way, when some situation for an entity is detected, an alarm could be thrown jointly with a set of recommendations from the OM [7, 8].

It could be possible that an entity has no previous experience or specific knowledge in a situation (e.g. the M&E project is new). However, this does not mean that the OM is empty, because it keeps the knowledge and previous experiences from all the projects. If it is possible to look for a similar semantically entity, then it would be possible to reuse its specific knowledge by analogy (i.e. an approximated recommendation is better than nothing). Thus, it would be adequate to identify whether it is possible to detect entities semantically similar in an M&E project based on the C-INCAMI framework. In this way, the main contribution in this work is to carry out systematic mapping of the literature to identify and find the techniques, and methods that allow the detection of semantically similar entities through the use of the research method termed Systematic Mapping Studies (SMS) [9–13].

This paper is organized as follows. Section 2 presents the research work and the Systematic Mapping Studies methodology. Section 3 describes the execution of the protocol related to the SMS research method. Section 4 outlines the results related to the study. Section 6 proposes a comparative analysis between the obtained documents jointly with a citation analysis. Section **??** summarizes reached conclusions and future works.

2. Research method

As indicated, this research is designed to follow the guidelines for SMS as specified in [9–13] and it is applied to the Data Stream Processing and Data Science. A Systematic Mapping Study may be more appropriate than a Systematic Literature Review (SLR) when we realize that there is very little evidence or that the research topic is very broad. In this way, applying a SMS allows tracing the evidence in a domain to a high level of granularity. For example, in [14] the wide search related to requirements elicitation process model is preferred instead of a specific and deep search because the global perspective is prioritized.

The subject of semantic similarity of entities, feasible of monitoring under a real-time decision-making process, is not a common or frequent aspect. However, the application of systematic mapping in machine learning and data mining is not new, for example, the systematic mapping has been applied in process mining [15], missing values techniques in software engineering data [16], machine learning applied to software testing [17], data quality based on computational intelligence techniques [18], among others.

In this section, the review protocol will be detailed to carry out the systematic review. The following steps will

Table 1 Motivations associated with research questions

Research questions	Motivations
RQ1: What methods exist to detect semantically similar entities in an M&E project?	To determine if there are different methods that allow the detection of similar entities and, in this way, evaluating which could be the most convenient application to the M&E project used.
RQ2: What kind of entities are compared to see their similarity?	To evaluate what types of entities are, and then, contrast them to determine whether there is a similarity between them or not.
RQ3: What kind of similarities can be found between entities?	To analyze the types of similarities that may exist to evaluate those applicable to the topic addressed.
RQ4: What type of publications have treated the topic of semantic similarity over time?	To consult all those publications where the topic has been presented, considering the type of publication and its evolution over time.

be considered: a) Identification of a need for revision, b) Specification of research questions, c) Determination of the search strategy, d) Specification of the data extraction process, e) Specification of the synthesis process and f) Application of the selection criteria of the articles found. Next, the research questions are described.

2.1 Specification of research questions

In this work, the objective is related to carrying out a systematic mapping of the literature in order to find techniques and methods that allow us to detect semantically similar entities in measurement processes. Following the methodology used, the Research Questions (RQ) which guide the work are defined. The main research question is defined, as follows **RQ1:** *What methods exist to detect semantically similar entities in an M&E project?* From the main question, the following questions arise, **RQ2:** *What kind of entities are compared to see their similarity?* **RQ3:** *What types of similarities can be found between entities?* and finally **RQ4: What type of publications have treated the topic of semantic similarity over time?** Table 1 introduces each research question, describing its definition jointly with the associated motivations.

2.2 Search strategy

Once the research questions have been defined, we describe the search strategy established to carry out the research work. Table 2 shows the defined search chain, according to the main objective established.

Once the search chain is defined, it is necessary to identify and define its associated strategy. Table 3 summarizes the search strategy, establishing the parameters considered to perform it.

Table 2 Description of the search string

Majors terms	Alternative terms
Semantic Similarity	((“similarity semantic”) OR (“coefficient similarity”) OR (“measure similarity”) OR (“similarity between text”) OR (“semantic measurement text”))

Table 3 Summary of the search strategy

Chosen Database	Scopus
Target items	Article Journal papers Conference papers Book chapter
Search applied to	Abstract – when this was not possible, the search is developed in the full text
Language	Papers are written in English
Publication period	From all to present

The selected database for carrying forward the search strategy was Scopus, focusing on the articles, journal papers, conference papers, and book chapters.

2.3 The selection process of articles

Once the search strategy has been established, the summary of the selection strategy of articles is presented in Table 4. It indicates the inclusion and exclusion criteria aligned with the selected methodology on both the abstracts as full articles.

At the time of selecting the articles, the steps performed

Table 4 Summary of the selection strategy

Inclusion criteria	**1.** Terms that meet the search string. **2.** Written publications in English. **3.** Articles, Journals, Conference Documents, Book Chapter. **4.** Date of publication: all articles until now.
Exclusion criteria for the title and abstract	**1.** Documents that do not focus on Computer Science. **2.** Articles that do not comply with the search string. **3.** Personal blogs.
Exclusion criteria for the full text	**1.** Publications that present a summary of some keynote, invited talk, or lecture. **2.** Use of semantic similarity in other areas than Computer Science.

Table 5 Data extraction form

Research questions	Dimensions	Categories
RQ1	Methods	Entities, Similar, Semantically
RQ2	Type of entities	Entities, Similar
RQ3	Type of similarity	Entities, Similarity
RQ4	Publications	Similarity, Semantics

were the following: a) All duplicate items were removed; b) The articles were filtered using the results from the previous step, applying the inclusion and exclusion criteria established in Table 4; c) Using the filtered results, an analysis was carried forward to select the definitive corpus.

2.4 Data extraction process

Using the obtained articles from the selection process (i.e. the definitive list of papers for the research), the data extraction form is prepared, indicating the classification scheme, detailing the corresponding dimensions, and categories (See Table 5).

As a supporting material, an Excel file was prepared to incorporate all the articles found and indicating the reasons for including or excluding each one based on the reading of the Abstract/full Article.

2.5 Synthesis process

The synthesis process is an integral part of the used methodology. It is responsible for evaluating each found article, eliminating duplicates and selecting only those that comply with the search string, the inclusion and exclusion criteria, the indicated keywords as well as the types of filtered documents (i.e. journal articles, conference papers, among others). Next, each abstract of the downloaded articles was read, selecting those that gave a partial or total response to the research questions (see Table 1). Finally, the articles were selected as long as they could give answers to the research questions previously defined.

3. Execution of SMS

In this section, the execution of the defined protocol is detailed. The search is performed according to the search string established in Table 2, considering as well as the search strategy specified in Table 3. Next, the results were filtered using the inclusion and exclusion criteria (See Table 4) to refine the search for articles of the selected theme. Figure 1 shows a screenshot related to the first application of the search string in the Scopus (see Table 2), which ran at 14-Oct-2019 18:30hs (Argentina).

Applying the original search string defined in Table 2, 2,122 records were found. However, incorporating ("entity" OR ("entities")) AND at the beginning of the search string, allows refining the search string establishing that the terms *entity* or *entities* must be present in the listed documents as a result. The refined search string answered with 111 documents, then filtered using the following conditions i) Subject Area: it was limited to the "Computer Science" and "Engineering" areas, ii) Document Type: limited to conference papers, articles, and book chapters, iii) Keywords: the words "Semantics",

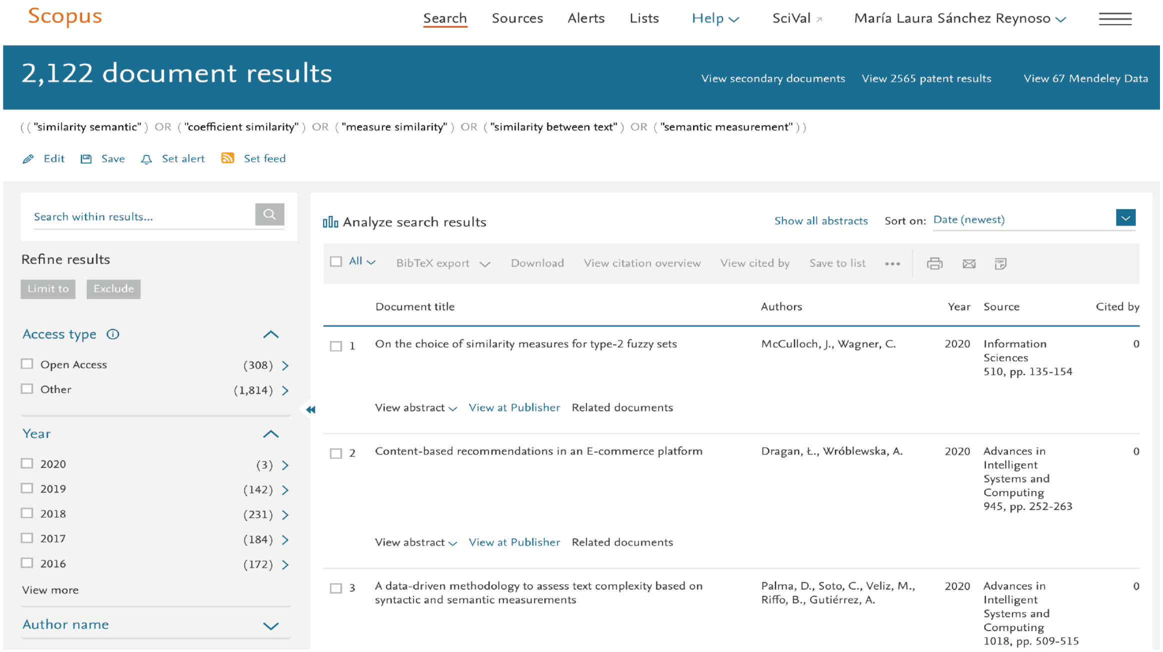

Figure 1 Screenshot related to the number of documents found 14-Oct-2019 18:30hs

"Semantic Similarity", and "Similarity" were chosen, iv) Language: English. Once the filters were simultaneously applied, just 25 documents were obtained.

The selection of the articles was carried forward through the reading of the abstract, considering the best and alternative terms, jointly with the research questions. Thus, after reading them, only 6 documents were retained from the original 25 documents, corresponding all of them to journals between 2010 and 2018.

4. Summary of the results obtained

From the final list related to the selected documents, four articles simultaneously satisfy the requirements related to filters, relevance, and research questions.

In the first work [19], the authors describe a process organized in three stages for measuring and comparing ontologies. In this sense, the concept of stable semantic measure is introduced. Even when the semantic similarity is applied to the ontologies, the underlying idea could be related to the entity under monitoring as a way for determining the semantic similarity between them.

In the second work [20], the authors propose an unsupervised approach for measuring the semantic similarity in short texts. It is very interesting because the text description is an alternative to incorporate at the moment in which an entity is defined.

In the third work [21], the author proposes ColorSim, a way in which the semantic similarity could be quantified considering the embedded semantic in OWL 2 (Web Ontology Language) annotations.

In the fourth work [22], the authors introduce a new technique for computing the semantic similarity between two entities. It is very interesting because the technique is based on structured knowledge coming from an ontology or taxonomy. The technique proposes the use of the multi-tree concepts for determining its similarity based on the underlying ontology or taxonomy.

In the fifth work [23], the authors propose to develop a thematic model to calculate the similarity between graphical entities. This article is interesting for our work because it proposes the use of a thematic model to measure similarity, allowing us to make a comparison with the previous articles where they also mention techniques and methods to measure semantic similarity. The underlying idea is to compute the similarity based on a topic model related to the document content jointly with the hierarchical content organization obtained from the knowledge graph.

In the sixth work [24], the authors propose a semantic similarity measurement model to determine the semantic similarity in geographic information. The semantic proposal addresses the geographical query strategies, improving the precision of keyword-oriented queries. The underlying idea falls in using metadata coming from the OGC geographic information services for contrasting them

Table 6 Comparison of articles

Articles	Entities	Similar	Semantically
"A graph derivation-based approach for measuring and comparing structural semantics of ontologies" [19]	X	X	X
"A resource-light method for cross-lingual semantic textual similarity" [20]		X	X
"Exploiting semantics from ontologies to enhance accuracy of similarity measures" [21]	X	X	X
"Measuring semantic similarity using a multi-tree model" [22]		X	
"Topic Model Based Knowledge Graph for Entity Similarity Measuring" [23]	X	X	
"Using semantic similarity model to improve OGC web services matching accuracy" [24]		X	X

with the desired service by users. Thus, the matching between the expected services from the users and those offered are prioritized through the similarity computation using metadata and query, making easy the searching and reaching the desired service.

The selected articles allow simultaneously answering each research question aligned with the objective of this research. The number is not high, but the results are pertinent to the research line and useful from the conceptual viewpoint.

5. Comparative analysis of results obtained

Considering the selected articles, four articles have satisfied simultaneously the requirements related to filters, relevance and research questions. Thus, it is possible to compare them to evaluate their content from the perspective of the authors.

In Table 6 below, it is possible to appreciate in detail the articles selected for the SMS along with the comparison between them.

In [19], the authors mention the use of a graphical derivation representation (GDR) approach to semantic measurement, which captures the structural semantics of ontologies. They focus on the experimental comparison based on new ontological measurement entities and distance metrics;that is to say that, in this article the concepts of entities and metrics are taken into account, allowing it to be selected in order to evaluate in our work the technique proposed by the authors and how they carry out the experimental process. The approach is based on a graph derivation of ontologies for reaching a stable semantic measurement. The authors outline different issues in relation to the expression of concepts in ontologies, for example, the possibility of expressing the same concept in different ways in two distinct ontologies. The approach is substantiated in three stages describing the first stage as a Graph Derivation Representing (GDR) based approach in which each ontology's concept is derived through the application of a given set of rules. In the next stage, it will merge a set of GDR to produce an integrated GDR for the whole ontology. Finally, in the last phase, the ontologies' relationships are treated to homogenize concepts (e.g. identify and unify the expression of those concepts expressed differently in two or more ontologies). Thus, both GDRs could be compared in terms of a level of correlation between ontological entities.

On the other hand, in [20], the authors propose an unsupervised and simple approach to measuring semantic similarity between texts in different languages. Using three different datasets, the semantic similarity was measured on short texts. Thus, the authors show that the proposed approach achieves similar performance to supervised and resource-intensive methods, showing stability across different language pairs. The process consists of comparing two texts expressed in different languages and identifying how similar they are, using the Natural Language Processing (NLP). Initially, the language's corpus for each source language is obtained. Next, a word embedding model is built for each language. Thus, the embeddings language for each short text is obtained. Authors employ a translation matrix related to both languages for contrasting the similarity using a model derived from the short texts, obtaining a bilingual modeling space. From there, both sentences could be compared using the cosine distance to know the level of proximity between them semantically speaking. Taking into account the objective of our research, this article is selected after applying the SMS protocol, because it is

Table 7 Key aspects of the approaches related to the compared articles

Aspects	[19]	[20]	[21]	[22]	[23]	[24]
Aim	To establish the level of correlation between ontological entities	To compute the level of similarity for two short texts expressed in different languages	To compute the similarity between two ontology entities	To compute the non-linear similarity of two concepts hierarchically expressed by a multi-tree.	To compute the similarity between entities considering the related topics and representational hierarchical.	To compute the similarity between the geographic offered services and those demanded by a user
Perspective	Ontologies	NLP	OWL2 Annotations	Hierarchical representation of concepts based on ontologies or taxonomies	Knowledge Graph	Metadata contrasting between services and query
Input	Ontological entities from different ontologies	Short texts and associated language corpuses	Ontological entities from different ontologies	Concepts jointly with its ontology/taxonomy	Documents	Metadata document from services and user query
Distance	Ontology measurement and comparison based on GDR	Cosine	ColorSim (A measure of similarity)	Computing of the weighting for the root of the combined multi-tree obtained from the original concepts	The similarity of entities based on a knowledge graph articulated with a topic model	The Metadata-based similarity jointly with its descriptive attributes

interesting to describe the approach they take to measure semantic similarity which helps to answer previously defined research questions as well as to analyze the technique and procedure carried out in the measurement. It can be observed in Table 6 that this article describes concepts such as similarly and semantically, which are part of our search chain defined to execute SMS.

The authors in [21] propose ColorSim, which is a measure of similarity that considers the semantics of the OWL2 annotations, to accurately calculate the relationship of two commented ontology entities. They make a comparison between ColorSim and the most advanced approaches and then report the results obtained from exploiting the knowledge coded in ontologies to measure similarity. This article is selected in our research to learn how the authors approach the concept of measuring similarity using a measurement that allows precision to be calculated. Also, the proposal considers the hierarchical aspects of the ontological entities to compute the level of similarity between them, avoiding, in this way, high similarity rates when the structural meaning is divergent. Thus, contents and hierarchical relationships are jointly considered to avoid the isolation of relationships and/or contents when the level of similarity needs to be computed.

Finally, in [22], the authors describe a new technique for calculating semantic similarity between two entities. The proposed method is based on structured knowledge extracted from an ontology or taxonomy. A concept of multiple trees is defined, and a technique is described that uses a multiple tree similarity algorithm to measure the similarity of two multiple trees constructed from taxonomic relationships between entities in an ontology. The idea of selecting this article consists of analyzing the technique proposed by the authors to measure the semantic similarity between two entities. The initial assumption indicates that is possible to represent a concept from an ontology or taxonomy as a hierarchical tree that is ordered based on its features and relationships.

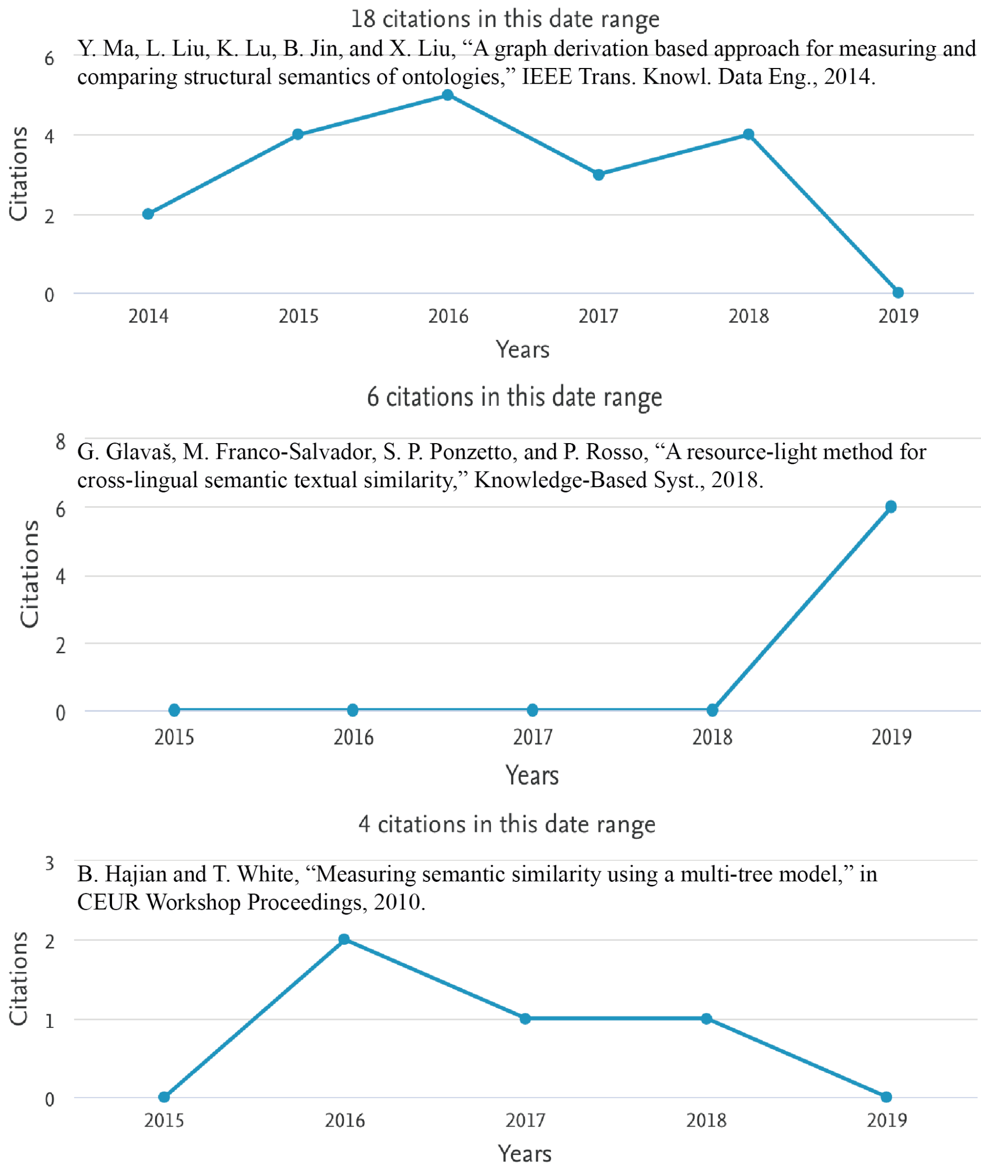

Figure 2 Comparative citation index extracted from the scopus database at October 14 of 2019

Thus, given two concepts and their corresponding ontologies, each concept could be expressed as a tree using its associated features and relationships indicated within is ontology. In this way, having two different concepts expressed as a tree, a new tree could be built from the merging of the previous ones to compute the joint weighting from the leaf to the root. Such weighting represents the level of similarity between the original trees. This strategy is non-linear and as a difference with other linear approaches (e.g. cosine distance, etc), this strategy considers the feature meaning, but also the relationships between them for computing the similarity.

Table 7 shows the aim, perspective, inputs and how the distance is computed in all the alternatives (even the excluded before). On the one hand, about the distance computing, no one uses the same way to compute such similarity, while [19] propose the computing based on ontologies and comparing a GDR, [20] uses the cosine distance based on Word2Vec models, [21] employs ColorSim based on OWL 2's annotations, [22] boards the computing considering thee hierarchy by multi-trees, [23] takes into consideration a knowledge graph and topic model for obtaining the similarity, and [24] computes the similarity based on metadata documents from services. On the other hand, the used perspective for each one is based on ontologies for [19, 21, 22], while it is NLP in [20], knowledge graph and topic models [23], and metadata documents [24].

Table 8 Comparative between the citation indexes

Articles	Publishing Year	Accumulated Citation	Average by year	Citations bounded to [2018-2019]
"A graph derivation-based approach for measuring and comparing structural semantics of ontologies" [19]	2014	18	$\frac{18}{(2019-2014)} = 3.60$	4
A resource-light method for cross-lingual semantic textual similarity" [20]	2018	6	$\frac{6}{(2019-2018)} = 6.00$	6
"Measuring semantic similarity using a multi-tree model" [22]	2010	4	$\frac{4}{(2019-2010)} = 0.44$	1
"Topic Model Based Knowledge Graph for Entity Similarity Measuring" [23]	2018	0	0	0
"Exploiting semantics from ontologies to enhance accuracy of similarity measures" [21]	2015	0	0	0
Using semantic similarity model to improve OGC web services matching accuracy" [24]	2015	0	0	0

Comparing the selected articles, all of them refer to the underlying concepts associated with entities, similarity, and semantic, proposing different kinds of techniques, methods or models which allow measuring the semantic similarity among entities.

Figure 2 introduces comparatively the citation index of each retained article, analyzing the citation evolution. The articles [21, 23, 24] are not present in the previous figure because they do not have citation associated, in other words, the citation index is zero.

From the three documents, the document [19] is the most cited article, reaching a citation peak in 2016 with five references. In this sense, it is worthy to mention that it has had a continuous citation from 2014 to 2018, being without citation during 2019. The minimum citation was in 2014 with two citations, remaining the rest of the years above this citation level (with exception of 2019). However, the document [20] has six citations until now limited to 2019 only. In other words, even when [19] has an accumulated citation index upper than [20], the last one has a better impact per year. In contrast, the document [22] accumulates 4 citations distributed between 2016 and 2018, without peaks that allow standing out from the previous ones.

It is possible to establish a scoring for the articles based on the citation indexes in order to provide a relative importance associated with the specific interest of each one. In this sense, taking into consideration the level of citation, the historical scoring model would provide the following order:

1. *"A graph derivation based approach for measuring and comparing structural semantics of ontologies"* with 18 citations.
2. *"A resource-light method for cross-lingual semantic textual similarity"* with 6 citations.
3. *"Measuring semantic similarity using a multi-tree model"* with 4 citations.

In addition to the previous scoring and based on Figure 2, it should be highlighted that taking the average by the year or the relative importance bounded to the last two years, the order previously indicated could change the first two positions such as it is synthesized in Table 8.

That is to say, considering the average by year the new order would be [20] (6 citations/year), [19] (3.6 citations/year), [22] (0.44 citation/year), and [21, 23, 24] (0 citations/year). Even, considering the citations indexes bounded to the last two years (i.e. 2019 and 2018) the mentioned order is kept as follows [20] (6 citations), [19] (4 citations), [22] (1 citation), and [21, 23, 24] (0 citations/year).

Thus, it is should be mentioned that the accumulated citation index provides a historical perspective, while the average by year represents an interest average rate which is interesting to be contrasted to analyze tendencies and peaks.

6. Conclusions and future work

This paper presented the application of a systematic mapping study. In this kind of study, all the evidence could be limited and plotted for a domain at a high level of granularity. The main and alternative terms introduced as part of the collecting method were "entity", "entities", "similarity semantic", "coefficient similarity", "measure similarity", "similarity between text", and "semantic measurement text".

The underlying idea related to the entities semantically similar is to reuse the knowledge and previous experiences when some similar entity has no evidence, or it is very limited. Thus, by the reuse of the knowledge and previous experiences, a recommendation system based on the organizational memory could deliver analogous courses of action for decision making.

In this mapping, a total of 2,122 documents were found, reduced to 111 by applying the exclusion and analysis criteria. Thus, a set of filters was applied focusing on computer science and engineering areas, among other aspects. Finally, just four works simultaneously satisfied the requirements associated with the objective of the research in concordance with the applied method.

The semantic similarity applied to entities under monitoring in the measurement and evaluation projects is a challenge. Real-time decision making depends on the obtained measures, the monitored entity, and the context in which it is immersed.

As a future work, a comparative analysis among the strategies, functionalities, and algorithms related to the semantic similarity computing will be carried forward.

7. Declaration of competing interest

We declare that we have no significant competing interests including financial or non-financial, professional, or personal interests interfering with the full and objective presentation of the work described in this manuscript.

8. Acknowledgements

This research is partially supported by the Project accredited by resolution CD 312/18 belonging to the Economy School of the National University of La Pampa.

References

[1] M. J. Diván. (2012) Fundamentos sobre Tecnología de la información para las Ciencias Económicas. [La Pampa: Facultad de Ciencias Económicas y Jurídicas (Universidad Nacional de La Pampa)]. [Online]. Available: https://bit.ly/2xJj7J3

[2] G. Rossi and D. Schwabe, "Modeling and implementing web applications with oohdm," in *Web Engineering: Modelling and Implementing Web Applications*, G.Rossi, O. Pastor, D. Schwabe, and L. Olsina, Eds. Springer, 2008, pp. 109–156.

[3] M. G. Mendonça and V. R. Basili, "Validation of an approach for improving existing measurement frameworks," *IEEE Transactions on Software Engineering*, vol. 26, no. 6, June 2000. [Online]. Available: https://doi/10.1109/32.852739

[4] F. Coallier, "A vision for international standardization in software and systems engineering," in *6th International Conference on the Quality of Information and Communications Technology (QUATIC 2007)*, Lisbon, Portugal, 2007, pp. 3–11.

[5] P. Becker, P. Lew, and L. Olsina, "Strategy to improve quality for software applications:A process view]," in *ICSSP '11: Proceedings of the 2011 International Conference on Software and Systems Process*, 2011, pp. 129–138.

[6] M. Diván and M. L. Sanchéz, "Fostering the interoperability of the measurement and evaluation project definitions in PAbMM," in *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2019, pp. 231–238.

[7] M. J. Divan, "Processing architecture based on measurement metadata," in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2016, pp. 6–15.

[8] M. A. Martin and M. J. Divan, "Case based organizational memory for processing architecture based on measurement metadata," in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2016.

[9] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empir. Softw. Eng.*, vol. 14, no. 131, 2009. [Online]. Available: https://doi.org/10.1007/s10664-008-9102-8

[10] J. M. Verner, J. Sampson, V. Tosic, N. A. Abu, and B. A. Kitchenham, "Guidelines for industrially-based multiple case studies in software engineering," in *2009 Third International Conference on Research Challenges in Information Science*, Fez, Morocco, 2009, pp. 313–324.

[11] D. Budgen and P. Brereton, "Performing systematic literature reviews in software engineering," in *Proceeding of the 28th international conference on Software engineering - ICSE '06*, Shanghai, China, 2006, pp. 1051–1052.

[12] P. Runeson and M. Höst, "Using mapping studies as the basis for further research – A participant-observer case study," *Inf. Softw. Technol.*, vol. 53, no. 6, June 2011. [Online]. Available: https://doi.org/10.1016/j.infsof.2010.12.011

[13] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, August 2015. [Online]. Available: https://doi.org/10.1016/j.infsof.2015.03.007

[14] D. Carrizo and C. Ortiz, "Models of requirements elicitation process: A systematic mapping," *Ing. Desarro.*, vol. 34, no. 1, January 2016. [Online]. Available: https://doi.org/10.14482/indc.33.2.6368

[15] A. R. Cárdenas and *et al*, "A systematic mapping study of process mining," *Enterp. Inf. Syst.*, vol. 12, no. 5, November 2018. [Online]. Available: https://doi.org/10.1080/17517575.2017.1402371

[16] A. Idri, I. Abnane, and A. Abran, "Systematic mapping study of missing values techniques in software engineering data," in *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Takamatsu, Japan, 2015, pp. 1–8.

[17] V. H. Durelli and *et al*, "Machine learning applied to software testing: A systematic mapping study," *IEEE Trans. Reliab.*, vol. 68, no. 3, September 11 2019. [Online]. Available: https://doi.org/10.1109/TR.2019.2892517

[18] I. D. López, A. Figueroa, and J. C. Corrales, "Un mapeo sistemático sobre predicción de calidad del agua mediante técnicas de inteligencia computacional," *Rev. Ing. Univ. Medellín*, vol. 15, no. 28, June 2016. [Online]. Available: http://dx.doi.org/10.22395/rium.v15n28a2

[19] Y. Ma, L. Liu, K. Lu, B. Jin, and X. Liu, "A graph derivation based approach for measuring and comparing structural semantics of ontologies," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, May 2014. [Online]. Available: http://dx.doi.org/10.1109/TKDE.2013.120

[20] G. Glavaš, M. Franco, S. P. Ponzetto, and P. Rosso, "A resource-light method for cross-lingual semantic textual

similarity," *Knowledge-Based Syst.*, vol. 143, March 1 2018. [Online]. Available: https://doi.org/10.1016/j.knosys.2017.11.041

[21] I. Traverso, "Exploiting semantics from ontologies to enhance accuracy of similarity measures," in *Proceedings of the 12th European Semantic Web Conference on The Semantic Web. Latest Advances and New Domains*, 2015, pp. 795–805.

[22] B. Hajian and T. White, "Measuring semantic similarity using a multi-tree model," in *CEUR Workshop Proceedings*, 2010, pp. 7–14.

[23] H. Sun and *et al*, "Topic model based knowledge graph for entity similarity measuring," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, Xi'an, China, 2018, pp. 94–101.

[24] L. Miao, Y. Zhou, W. Cheng, and J. Guo, "Using semantic similarity model to improve ogc web services matching accuracy," in *2015 Fourth International Conference on Agro-Geoinformatics (Agro-geoinformatics)*, Istanbul, Turkey, 2015, pp. 217–220.