Novel approach for semantic similarity cross ontology

Leila Benaissa Kaddar, Farah Ben-Naoum
Evolutionary Engineering and Distributed Information System Laboratory (EEDIS), Department of Computer Science,
Djillali Liabes University, Sidi-bel-Abbes, Algeria

Article Info	ABSTRACT
	Measuring semantic similarity between terms is a crucial step in information retrieval and integration since it necessitates semantic content matching. Even though several models have been proposed to measure semantic similarity, these models are not able to effectively quantify the weight of relevant items that affect the semantic similarity judgment process. In this study, we present a new method for measuring semantic similarity between cross-ontologies, that consists of hybridizing node-based approaches such as WuP and Reda with the weight of similarity computed using WordNet. The proposed approach has been experimented to show its efficiency with two ontologies, configuration management tool (CMT) and ConfOf, from the conference domaine in the web ontology language (OWL) ontologies benchmark OAEI 2015 and evaluated using two metrics: density and cohesion.

Corresponding Author:

Leila Benaissa Kaddar
Evolutionary Engineering and Distributed Information System Laboratory (EEDIS)
Department of Computer Science, Djillali Liabes University
Sidi-Bel-Abbes, Algeria
Email: Leila.benaissa_kaddar@univ-sba.dz

1. INTRODUCTION

Collaboration and competitiveness among businesses necessitate significant data exchanges, which are required for interoperability. Using ontologies, the companies describe the meaning of their data. The challenge that makes interoperability difficult is the lack of coherence between ontologies. In recent years, different tools and methods have been proposed for the reconciliation of the different ontologies. The evaluation of the degree of similarity between their concepts is the focus of most existing methods. In the literature, semantic similarity methods are classified into two main categories, single ontology similarity methods and cross ontology similarity methods.

First, consider a single ontology's semantic similarity measures [1]. They can be summarized as shown in: i) Ontology-based approaches [2]–[11] take into account the ontology's path length and depth, when applied to large-scale ontologies, the disadvantage of this approaches is that many inheritance relationships are ignored. Other factors that influence semantic similarity are not considered into account in this strategy. ii) Information content-based approaches [12]–[17] quantify the amount of information a concept expresses in the taxonomy by computing similarity in terms of the shortest path between target concepts in the taxonomy. This approach has the disadvantage of being greatly influenced by corpus. iii) The feature-based approach [18]–[20] considers the features that are shared by two concepts as well as the distinguishing features that are unique to each concept. This method has the advantage of being able to solve the problem of semantic similarity across ontologies. The disadvantage is that it is better suited to processing large ontologies with extensive semantic knowledge than small ontologies. iv) The hybrid-based approach combines different

sources of information to measure the level of similarity between concepts. Attribute similarity, ontology structure, information content, and node depth are all factors considered in these methods [21].

According to Elavarasi *et al.* [22], the main advantage of these approaches is that if the knowledge of an information source is inadequate, it may be derived from the alternate information sources. Which generally improves the quality of the similarity measure. Some representatives of this approach are [23]–[26]. The hybrid approach considers more factors than the single approach, but it mostly relies on expert experience and adopts the method of manual weight assignment to formulate the weight factors of each element.

Secondly, semantic similarity cross ontology. Elavarasi *et al.* [22] and Saruladha *et al.* [27] have developed measurements that compute similarity among concepts in different ontologies. The reason for this is the growing number of information sources on the web, which makes it difficult to evaluate the similarity between concepts. Furthermore, cross ontology and measurement match the words or concepts from different ontologies. Cross ontology often needs a hybrid or feature-based approach. This is due to the fact that the structure and information content of different ontologies cannot be directly compared [22]. The similarity measures between the concepts of different ontologies are classified in two main classes path length measure and feature based measure.

The first one is the approach path length based Information, which is similar to approach, used in the semantic similarity measures for single ontology. Al-Mubaid and Nguyen [28] Within the framework of the Unified Medical Language System, proposes a new ontology-structure-based technique for measuring semantic similarity in a single ontology and across ontologies in the biomedical domain (UMLS). To solve the problems that many existing semantic similarity measures that use on ontology structure as their primary source witch cannot measure semantic similarity between terms and concepts when multiple ontologies are used. His evaluation is based on three features: i) a new feature of common specificity of concepts in the ontology; ii) local granularity of ontology clusters; and iii) cross-modified path length between two concepts.

The second one is an approach based on features based on terms of information [29]. Tversky [30] propose a set-theoretical approach to similarity, where objects are represented as collections of features and similarity is described as a feature matching process. He demonstrates that the contrast model, which expresses similarity between items as a linear combination of measures of their common and distinctive features, is based on a set of qualitative assumptions. The author describes a method for computing semantic similarity that eliminates the need for a single ontology and takes into account differences in the levels of explicitness and formalization of distinct ontology specifications [31]. Using a matching process across synonym sets, semantic neighborhoods, and distinguishing features that are divided into parts, functions, and attributes, a similarity function determines related entity classes. Propose a method for computing semantic similarity that involves mapping concepts to ontologies and evaluating their links within those ontologies [32]. He investigated at methods for computing semantic similarity between natural language terms (usingWordNet as the underlying reference ontology) and medical terms (using the MeSH ontology of medical andbiomedical terms). His research also focuses on cross ontology approaches, which can compute the semantic similarity between concepts from different ontologies. An adaptive e-learning system with cross ontology similarity measure has been developed with automatically generated concept map as an application to the e-learning system [33]. The major feature of this method is the user personalization model, which assesses the student capability of learning. The other attraction of this method is the fact that it uses multiple ontologies for the evolution of the concept from a particular domain. Propose a hybrid approach for measuring semantic similarity between ontologies based on WordNet, denoted by WNOntoSim [34]. The semantic similarity across ontologies at the elemental level is calculated using WordNet. He compute semantic similarity between ontologies at the structural level by creating contexts of nodes in which the structure of the ontology is encoded, and then combining these scores to get a full semantic similarity between ontologies. The most disadvantage of WNOntoSim is that it cann't accurately compute semantic similarity between named entities because of coverage problem of WordNet. He presents a general approach for assessing similarity across multiple ontologies [35]. His strategies (the first focusing on high scalability and the second on high accuracy) aim to find an LCS that accurately represents the commonalities between terms among multiple ontologies.

His approach is based on evidence found in background ontologies, both explicit (semantic) and implicit (structural). In this paper, we propose an original approach for computing semantic similarity between different ontologies. First, we merged these different ontologies in order to have a path between them and then we computed the weight similarity between concepts that are in the form of sentences using WordNet and finally we hybridized node-based approach such as WuP and Reda and Al with the weight of similarity computed before. This combination was necessary to integrate and reinforce the semantic factor. This paper is organized as shown in: Section 2 deals with the architecture of the proposed system. Section 3 experiments results and evaluation measures. Conclusion and perspective are given in section 4.

2.   THE PROPOSED APPROACH
Our approach is composed of four phases, which are:
- The preprocessing.
- Computing the similarity measure of Wu and Palmer and Reda and Al.
- Computing the Hybrid Similarity measure of WWP and WRA.
- Evaluation of our approach using the Cohesion and Density method.

Each phase is composed of several steps, which are explained in this section see Figure 1.
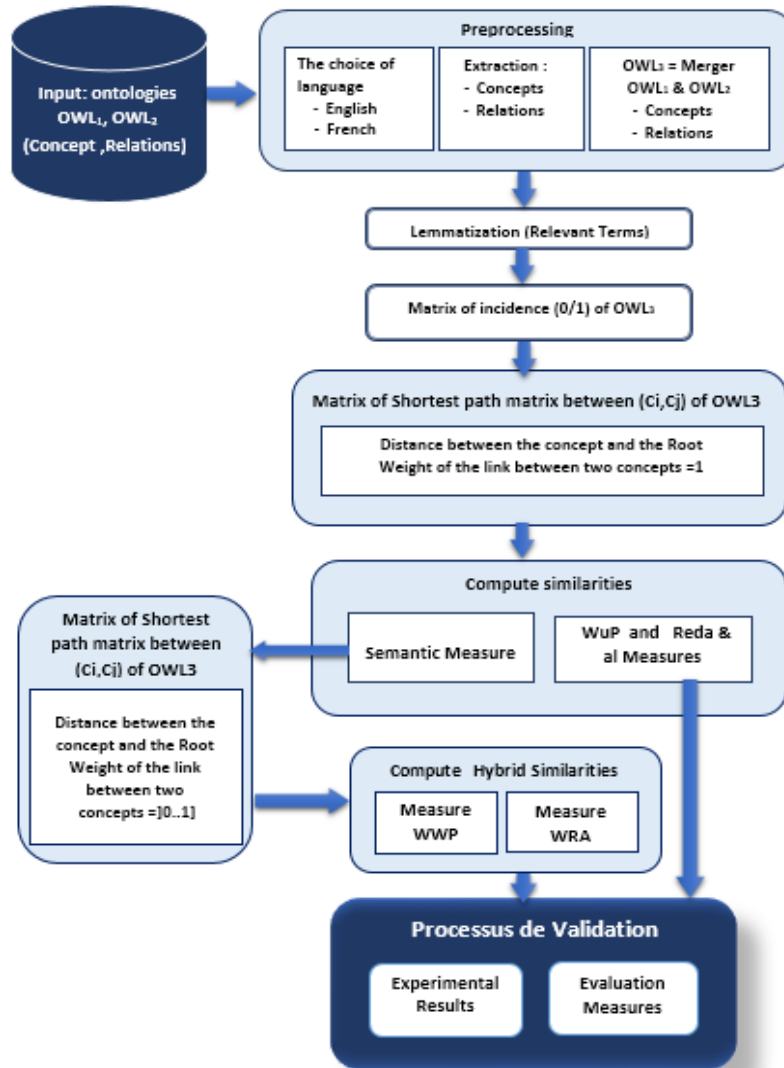

Figure 1. The proposed approach

2.1. Input OWL1 and OWL2 (dataset benchmark selection)
For the English dataset, benchmark, we used a collection of ontologies describing the domain of conference organization found in the OAEI [36] organizes evaluation campaigns aiming at evaluating ontology-matching technologies. These ontologies describe the conference domain. We justify the choice of this data set by the following points: i) the most well-known evaluation campaign for testing the performance of ontology matching systems is the OAEI campaign. ii) In our approach, we need to apply similarity criteria to different ontologies but from the same domain, which is possible with this dataset.

2.2. Preprocessing
Firstly, in order to carry out our approach, we load two ontologies from a dataset file; this file contains the same ontologies but in different languages, after the system extracts all constructors from the

web ontology language (OWL) ontology file (concepts and relations between each concept in the same ontology). Secondly, we merge two ontologies selected in the previous step in order to find paths connecting between the concepts of the first ontology with a concept of the second ontology using protégé 2000. Protégé 2000 is one of the greatest ontology management software programs currently available. This application's performance is attributable to the efficiency of its integrated tools, such as PROMPT Suite [37]. It is constituted of a set of tools that are useful for merging and mapping ontologies. One of the PROMPT tools is iPROMPT, which performs basic ontology merging operations. The algorithm's first stage requires two ontologies as input and returns a list of first suggestions for matches based on the concept names' lexical equivalence [38]. The algorithm then moves on to the next step, in which users do an action of their choosing.

This operation is a task of the algorithm, which is done after human intervention. The choice of operation is made by selecting one of the suggestions or by specifying the required operation using the ontology-editing environment. The next step of iPROMPT automatically executes the modifications according to the previously selected operation. Then, iPROMPT generates again a list of suggestions based on the structure of the ontology, the inconsistencies and problems resolved after execution of the operation. Finally, iPROMPT proposes solutions for these problems and generates the merger ontology.

However, the iPROMPT tools have some limitations:
- The semi-automation of the merger algorithm.
- The iPROMPT considers the structure of the ontology, but does not consider the treatment of the relations between the concepts and the pertinence of concepts.

Figure 2 represente screenshot of the graph node of OWL1 (the red one), OWL2 (the blue one) and OWL3 which represents the merging of OWL1 and OWL2 (red and blue).

Figure 2. Screenshot of the graph node of OWL1, OWL2 and OWL3

2.3. Lemmatization

After analyzing the concepts of the two ontologies, we notice that each concept is a sentence (set of words). Before applying the similarity measure, we used Tree Tagger [39] to lemmatize all of the concepts in order to keep only the relevant words of each concept, as shown in Figure 3. The type of relevant words chosen is noun, adjective and verb.
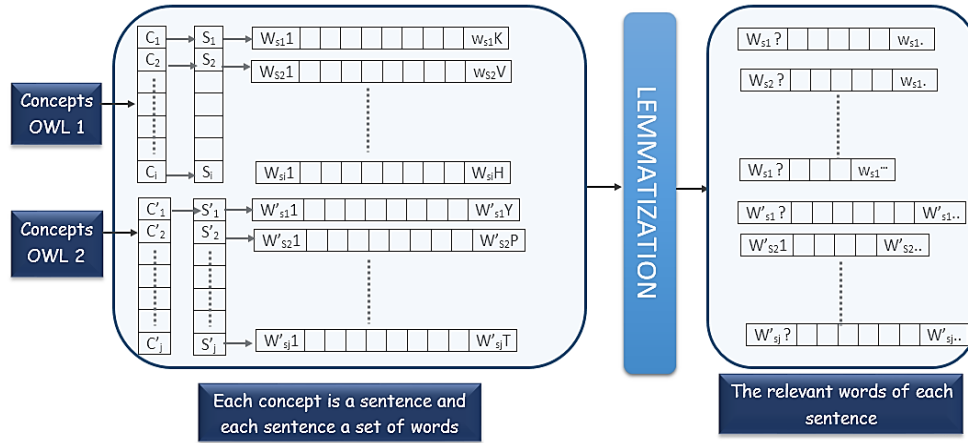

Figure 3. Preprocessing on concepts

The TreeTagger is a tool for annotating text with information about parts of speech and lemma. Schmid [39] created it as part of the TC project at the Institute for Computational Linguistics at the University of Stuttgart. German, English, French, Italian, Danish, Swedish, Norwegian, Dutch, Spanish, Bulgarian, Russian, Portuguese, Galician, Greek, Chinese, Swahili, Slovak, Slovenian, Latin, Estonian, Polish, Persian, Romanian, Czech, Coptic, and old French texts in English have all been successfully tagged with TheTreeTagger. In our approach, we have used this tool for the English and French language. We have chosen relevant terms in our approach (Noun, Adjective, Verb), to compute the weights of similarity between concepts (C_i, C_j) using Algorithm 1.

Algorithm 1. Lemmatization using TreeTagger

```
Input: Set of concepts Ci// knowing that each concept is a sentence and each sentence a set
of words.
Output: Vrt : vector containing relevant terms for each concept
Method:
Vrt←φ
For each Ci do
begin
Let VtokenVector of relevant terms Ti of the conceptCi
For each terms Ti of the vector Vtoken do
begin
        Lemmatization (Vtoken(Ti))
        Let token← term Ti
        let pos← type Ti
        let lemma← original Ti
                ifposЄ{ Noun, Adjective, Verb }
                        Vrt←Vrt∪lemma
end
returnVrt
end
```

2.4. Computing matrixes

In this step, we compute the incidence matrix on OWL3 by setting "0" if there is not a pathbetween two concepts (C_i, C_j), otherwise set "1". According to this matrix, we apply the Dijkstra algorithm; it is one of the most widely used algorithms for finding the shortest paths from a particular source node to any other node. Where the weight of the link is equal to "1". We all compute the distance between a concept and the root (root R). The reason for this is that some similarity measures used in this work are based on the computation of distances that separate the desired nodes from the root node R and the distance that separates

the subsuming concept (CS) from these nodes. An example is the Wu and Palmer measure. This phase generates a minimal distance between a concept and its root node R, see Algorithm 2.

Algorithm 2. For finding the shortest path between the concepts

```
Input: - Max: number of all concepts in the merging ontology OWL3,
- Mat: integer matrix of size (max) x (max),
- MI: it is a matrix of link between the concepts, equal to 1 if there is a link otherwise
-1
Output: CS: matrix of all paths from each start (d) to each arrived (a)
Method:
fori=0 : Max do
for j=0 : Max do
        if (MI[i,j] == -1)
        Mat[i,j] ← 100000
        d← 0
        while (d< Number of concepts) do
        begin
        Dijkstrada // da is instance of Dijkstra
        da set of all paths from start d to arrive a in Mat matrix
        fora =1 : Maxdo
           begin
           Let V a vector of all way from d to arrive to a
           Let N be an integer vector of size V
           path_da← Ø //path is a set of node where d is the first node and a is the last node
        initialized at empty set
           fori =0 : sizeof(N) do
               begin
               N[i] ←Get(V (i)) // receive value of V[i]
               path_da←path_daU {N[i]}
               end
           storage path_da in the CS matrix at line d and column a
           end
        d←d+1
        end while
```

2.5. Similarity measure computing
In order to show the effectiveness of our approach, we have implemented two measures Wu and Palmer [5] and Rada *et al.* [3]. This is in order to make a comparison. This evaluation is discussed in section 3.1.

2.5.1. Wu and palmer
The edge counting method proposed by Wu and Palmer [5] is defined as shown in: OWL3 is composed of a set of nodes and a root node (R) as shown in Figure 4. C1 and C2 are two concepts of the ontology for which we will compute similarity. The distances (N1 and N2) between nodes C1 and C2 and the root node, as well as the distance (N) between the closest common ancestor (CS) of C1 and C2 from the node R, are used to compute similarity. The similarity measure proposed by Wu and Palmer [5] is defined as (1).

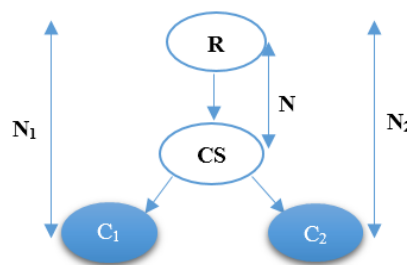$$Sim_{wp}(C_1, C_2) = \frac{2*N}{N_1+N_2} \qquad (1)$$

Figure 4. Example of a concept hierarchy

The issue with this measure is that in ontology, arcs represent equal distances, meaning that all semantic links have the same weight. After analyzing a comparison between similarity measurement methods this comparison shows that, the Wu and Palmer present the advantage of being simple to compute, while

remaining as expressive as the others do. This is the reason that influenced us to adopt this measure as the basis of our hybrid approach.

2.5.2. Reda and Al

In ontology, the number of minimal arcs that separate two concepts must be computed to estimate their similarity. To discover the shortest path between two concepts, this measure uses the edge-counting method. The (2) defines the Rada *et al.* [3] measure:

$$Sim_{Ra}(C_1, C_2) = \frac{1}{1+dist(C_1,C_2)} \tag{2}$$

hence, dist (C1, C2) corresponds to the number of arcs that must be traversed in the ontology to connect the concepts C1 and C2.

2.6. The weight similarity matrix

The objective of this step is to compute the similarity weight between each concept of OWL_1 with the concepts of OWL2. Knowing that the concepts of our two ontologies are in the form of sentences, in order to calculate the weight similarity between these concepts, we use the lemmatization step, which was applied previously. This allowed us to obtain a set of relevant terms for every concept, in order to compute first the weight of similarity between the word pairs before going to the phrase.

We have used WorldNet. WorldNet [40] is a free English electronic dictionary created by cognitive scientists at Princeton University, directed by Miller. There are 150,000 words in the WorldNet 2.0, organized into 115,000 synonym sets. There is a total of 207,000 words sense groups. Each synonym represents a fundamental semantic concept and is linked with lexical relations and conceptual-semantic. We use WorldNet to compute the weight of similarity between two words. The value obtained will be used in the (3) to compute the weight of similarity between concepts. After we use the formula [41] to calculate the similarity between two concepts, the first concept comes from OWL_1, the second from OWL2.

$$Sim(C_{O1}, C_{O2}) = \frac{\sum_{i=0}^{n} MaxSim(wi(w'1,...,w'm))}{2n} + \frac{\sum_{j=0}^{m} MaxSim(w'j,(w1,...,wn))}{2m} \tag{3}$$

The weight of similarity between the concepts allows to build the following matrix as shown in Figure 5, which represents the semantic similarity between these two ontologies.

OWL 1 \ OWL 2				C_1			C_2		C_3	C_4				C_5			C_6	C_7		C_8
				w_1	w_2	w_3	w_1	w_2	w_1	w_1	w_2	w_3	w_4	w_1	w_2	w_3	w_1	w_1	w_2	w_1
C'_1	$w'1$	$w'2$	$w'3$	$w'4$																
C'_2	$w'1$		$w'2$																	
C'_3	$w'1$																			
C'_4	$w'1$		$w'2$	$w'3$																
C'_5	$w'1$	$w'2$	$w'3$	$w'4$	$w'5$															
C'_6	$w'1$																			
C'_7	$w'1$		$w'2$																	

Computation of similarity for each two concepts browsing OWL 3

Figure 5. Matrix of the weight of similarity

2.7. Computing hybrid matrix

In this step, the incidence matrix is updated by computing the union of the incidence matrix computed in section 2.4 with the matrix of the weight of similarity computed in section 2.6. After, the algorithm is modified to integrate the update of the incidence matrix. This hybrid similarity matrix allows us to have the shortest path between concepts.

Algorithm 3. For optimizing the shortest path between concepts

```
Input: - Max: number of all concepts in the merging ontology OWL3,
- Mat: integer matrix of size (max) x (max),
- MI: it is a matrix of link between the concepts, equal to 1 if there is a link otherwise
-1
- MP: matrix of weight similarity between all concepts.
Output: CSsh : Updating matrix of hybrid similarity of all paths from each start (d) to each
arrived (a)
Method:
Let MIup: union matrix between MI and MP
for i=0 : Max do
for j=0 : Max do
        if (MIup [i,j] == -1)
        Mat[i,j] ← 100000
        d← 0
        while (d< Number of concepts) do
        begin
        Dijkstrada // da is instance of Dijkstra
        da set of all paths from start d to arrive a in Mat matrix
        fora =1 : Maxdo
            begin
            Let V a vector of all way from d to arrive to a
            Let N be an integer vector of size V
            pathda← Ø //path is a set of node where d is the first node and a is the last
            node initialized at empty set.
            fori =0 : sizeof(N) do
                Begin
                N[i] ←Get(V (i)) // receive value of V[i]
                pathda←pathdaυ {N[i]}
                end
                storagepathda in the CSsh matrix at line d and column a
            end
        d←d+1
        end while
```

2.8. Computing the hybrid similarity

The principle of computing similarity with node-based approaches such as Wu and Palmer, Reda is based on the idea that a shorter path between two nodes makes them more similar. Another point about these approaches is that the arcs represent uniform distances. Therefore, this approach has the disadvantage that all semantic links have the same weight, which imposes difficulty in defining and controlling the linking distances. This is the reason why we have chosen to hybrid these approaches with the weight of similarity to reinforce the semantic aspect.

2.8.1. Wordnet Wu and Palmer (WWP) approach

In this approach, we apply the same principle as wup as shown in Figure 6 but we replace the values of the semantic links, which are uniform and have the same weight in this method, with the weight similarity between each concept in owl3 computed in the Weight similarity matrix. After we use the hybrid matrix computed in 2.7 to update the shortest path between concepts. This modification allows highlighting the semantic aspect in the computing of the similarity.
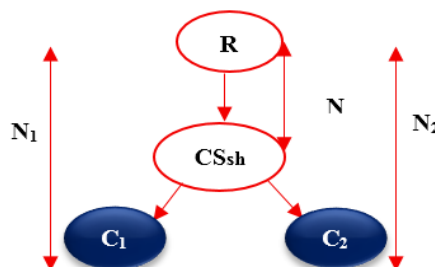

Figure 6. Example of a concept hierarchy in WWP

2.8.2. WRA approach (Wordnet Reda and Al)

In this approach, we use the formula of Reda and Al, but with an updated computation of the distance between C1 and C2 as shown in (4). Considering that distsh (C1, C2) corresponds to the number of arcs, taking into account the weight of similarity between the concepts that must be crossed in the ontology to connect the concepts C1 and C2.

$$Sim_{wra}(C_1, C_2) = \frac{1}{1+dist_{sh}(C_1,C_2)} \tag{4}$$

3. EXPERIMENTAL EVALUATION

We performed our experiment between two ontologies. In the first one, we use Cmt, which contains 29 concepts and 76 relations. The second one, ConfOF, contains 38 concepts and 107 relations. These two ontologies describe the conference domain. Our system analyzed and extracted the OWL constructors of the two ontologies in 20 seconds.

3.1. Experimental results

Some experimental evaluations of the proposed approach are described in this section. All of the tests are executed on a laptop with an Intel Core (TM) i3 Duo 2.30 GHz processor and 4GB of RAM running Windows 7 and Java Netbeans 8.2. We report the results of our experiment by presenting evaluation measures and comparisons between node-based approaches and our approach. Table 1 describes comparison between the performances of our WWP and WRA measures with two popular measure of similarity WuP and Reda. From the analysis, it is clear that the suggested approach produces better results than other methods for the sam two ontologies used which are cmt and con-of from the conference domain. Figure 7 represente two graphs the first one comparaison between the performances of our WWP measures with WuP and the second one WRA measures with Reda. The number of arcs has increased at a rate of 9% for this experimentation. This is due to the impact of integrating the weight of similarity in the computation of the shortest path between the concepts.
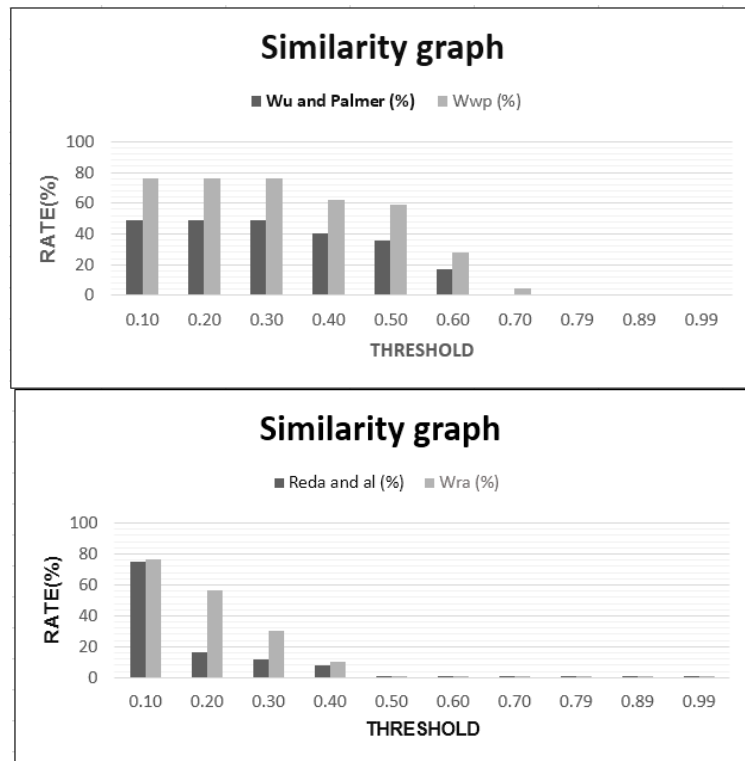

Figure 7. Graph of comparaisonbetween the performances of our WWP and WRA measures with WuP and Reda

3.2. Evaluation measures

Although there are many evaluation methods available today, the cohesion and density evaluation methods are used in this paper. Figures 8 and 9 show the impact of cohesion and density between the proposed approach and other methods (Wu and Palmer and Reda and A1). We compute the density and the cohesion between pairs of concepts (Ci, Cj) by treating these concepts as context units. Concerning the density, we use the probability of terms appearing in these units as well as explicit semantic relations to determine the density. Comparing the results obtained, we notice that the density is higher in our approach than in the other methods as shown in Figure 8. The degree of relatedness of OWL concepts, which are semantically related by the property's relatedness of items in ontologies, is referred to as cohesion, and we notice that the cohesion in our approach is stronger than Wu and Palmer's and Reda and Al's approaches as shown in Figure 9.

Table 1. Comparaison between the performances of our WWP and WRA methods with WuP and Reda

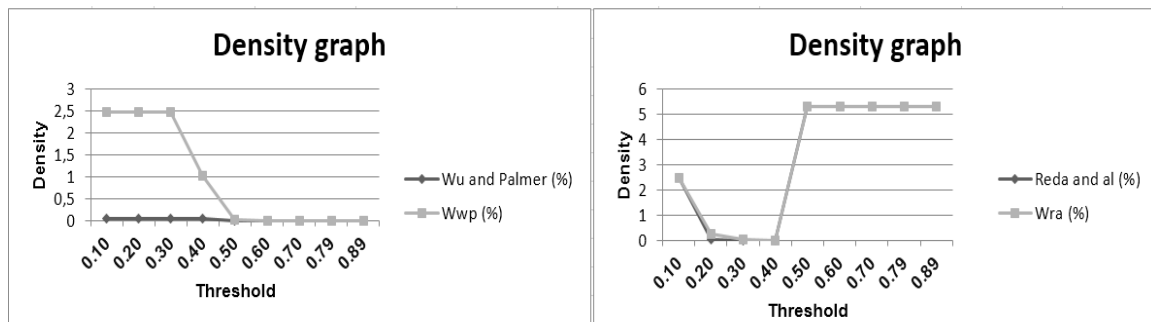| No	Threshold	Wu and Palmer (%)	Reda and Al (%)	Wwp (%)	Wra (%)
1	0.10	48.911	74.733	76.225	76.225
2	0.20	48.911	16.333	76.225	56.261
3	0.30	48.911	11.887	76.225	30.217
4	0.40	40.108	7.894	62.159	10.254
5	0.50	35.571	0.725	58.711	0.725
6	0.60	16.878	0.725	28.13	0.725
7	0.70	0.544	0.725	4.264	0.725
8	0.79	0.544	0.725	0.544	0.725
9	0.89	0.0	0.725	0.0	0.725
10	0.99	0.0	0.725	0.0	0.725

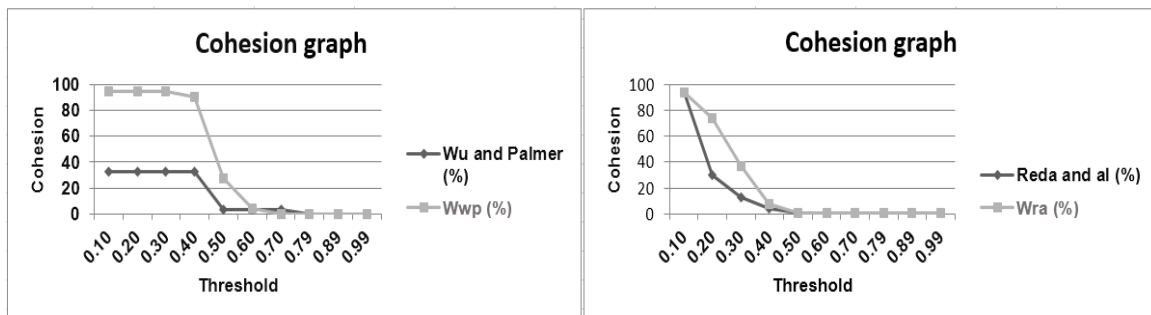Figure 8. Graph evaluating the density

Figure 9. Graph evaluating the cohesion

4. CONCLUSION

In this paper, we present a system to compute the semantic similarity between two different ontologies but in the same domain. These ontologies are represented in OWL. The main objective of this paper was to demonstrate the impact of integrating the weight of similarity using Wordnet between concepts in node-based approaches. This allowed us to optimize the shortest path and to reinforce the semantic factor

between these concepts. We compared our two methods (WWP and WRA) with the methods (WuP and Reda) respectively, and the results obtained are encouraging. For future work, there are some possible steps, which we can focus on. One of these is the merging ontology algorithm. Another direction to explore is how this measure influences research effectiveness on the same ontologies but in different languages, french and Arabic.

REFERENCES

[1] T. Slimani, "Description and evaluation of semantic similarity measures approaches," *International Journal of Computer Applications*, vol. 80, no. 10, pp. 25–33, Oct. 2013, doi: 10.5120/13897-1851.
[2] A. Budanitsky and G. Hirst, "Evaluating wordnet-based measures of lexical semantic relatedness," *Computational Linguistics*, vol. 32, no. 1, pp. 13–47, Mar. 2006, doi: 10.1162/coli.2006.32.1.13.
[3] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and application of a metric on semantic nets," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 1, pp. 17–30, 1989, doi: 10.1109/21.24528.
[4] F. Fakhar, "Semantic constraints satisfaction based improved quality of ontology alignment," *Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 3, pp. 182–189, 2013, doi: 10.12928/eei.v2i3.202.
[5] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics -*, 1994, pp. 133–138, doi: 10.3115/981732.981751.
[6] M. Biniz, R. El Ayachi, and M. Fakir, "Ontology matching using babelnet dictionary and word sense disambiguation algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 1, pp. 196–205, Jan. 2017, doi: 10.11591/ijeecs.v5.i1.pp196-205.
[7] C. Leacock and M. Chodorow, *Combining Local Context and WordNet Similarity for Word Sense Identification*. MIT Press, 1998.
[8] M. A. Hadj Taieb, M. Ben Aouicha, and A. Ben Hamadou, "Ontology-based approach for measuring semantic similarity," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 238–261, Nov. 2014, doi: 10.1016/j.engappai.2014.07.015.
[9] X. Zhou, X. Xie, Y. Xue, and B. Xue, "Ontology-based probabilistic estimation for assessing semantic similarity of land use/land cover classification systems," *Land*, vol. 10, no. 9, p. 920, Aug. 2021, doi: 10.3390/land10090920.
[10] A. Dennai and S. M. Benslimane, "A new measure of the calculation of semantic distance between ontology concepts," *International Journal of Information Technology and Computer Science*, vol. 7, no. 7, pp. 48–56, Jun. 2015, doi: 10.5815/ijitcs.2015.07.06.
[11] H. Q. Dung, L. T. Q. Le, N. H. H. Tho, T. Q. Truong, and C. H. Nguyen-Dinh, "A novel ontology framework supporting model-based tourism recommender," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4, pp. 1060–1068, Dec. 2021, doi: 10.11591/IJAI.V10.I4.PP1060-1068.
[12] S. Tongphu, B. Suntisrivaraporn, and P. Aimmanee, "Toward semantic similarity measure between concepts in an ontology," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, p. 1356, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1356-1372.
[13] A. A. I. N. E. Karyawati, E. Winarko, Azhari, and A. Harjoko, "Ontology-based sentence extraction for answering why-question," in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Sep. 2017, vol. 2017-December, pp. 1–6, doi: 10.1109/EECSI.2017.8239127.
[14] P. Resnik, "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language," *Journal of Artificial Intelligence Research*, vol. 11, pp. 95–130, Jul. 1999, doi: 10.1613/jair.514.
[15] L. Meng, J. Gu, and Z. Zhou, "A new model of information content based on concept's topology for measuring semantic similarity in WordNet," *International Journal of Grid and Distributed ...*, vol. 5, no. 3, pp. 81–94, 2012, [Online]. Available: http://www.sersc.org/journals/IJGDC/vol5_no3/6.pdf
[16] J. J. Lastra-Díaz and A. García-Serrano, "A novel family of IC-based similarity measures with a detailed experimental survey on WordNet," *Engineering Applications of Artificial Intelligence*, vol. 46, pp. 140–153, Nov. 2015, doi: 10.1016/j.engappai.2015.09.006.
[17] P. Verschaffelt *et al.*, "MegaGO: A fast yet powerful approach to assess functional gene ontology similarity across meta-omics data sets," *Journal of Proteome Research*, vol. 20, no. 4, pp. 2083–2088, Apr. 2021, doi: 10.1021/acs.jproteome.0c00926.
[18] G. Pirró, "A semantic similarity metric combining features and intrinsic information content," *Data and Knowledge Engineering*, vol. 68, no. 11, pp. 1289–1308, Nov. 2009, doi: 10.1016/j.datak.2009.06.008.
[19] N. Seco, T. Veale, and J. Hayes, "An intrinsic information content metric for semantic similarity in word net," in *Frontiers in Artificial Intelligence and Applications*, 2004, vol. 110, pp. 1089–1090.
[20] D. Sánchez, M. Batet, D. Isern, and A. Valls, "Ontology-based semantic similarity: A new feature-based approach," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7718–7728, Jul. 2012, doi: 10.1016/j.eswa.2012.01.082.
[21] M. Gan, X. Dou, and R. Jiang, "From ontology to semantic similarity: Calculation of ontology-based semantic similarity," *The Scientific World Journal*, vol. 2013, pp. 1–11, 2013, doi: 10.1155/2013/793091.
[22] S. A. Elavarasi, J. Akilandeswari, and K. Menaga, "A survey on semantic similarity measure," *International Journal of Research in Advent Technology*, vol. 2, no. 3, pp. 389–398, 2014.
[23] V. Schickel-Zuber and B. Faltings, "OSS: A semantic similarity function based on hierarchical ontologies," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 551–556, 2007.
[24] Y. Li, D. McLean, Z. A. Bandar, J. D. O'Shea, and K. Crockett, "Sentence similarity based on semantic nets and corpus statistics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 8, pp. 1138–1150, Aug. 2006, doi: 10.1109/TKDE.2006.130.
[25] Z. Zhou, Y. Wang, and J. Gu, "New model of semantic similarity measuring in WordNet," in *Proceedings of 2008 3rd International Conference on Intelligent System and Knowledge Engineering, ISKE 2008*, Nov. 2008, pp. 256–261, doi: 10.1109/ISKE.2008.4730937.
[26] Y. Li, Z. A. Bandar, and D. McLean, "An approach for measuring semantic similarity between words using multiple information sources," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 871–882, Jul. 2003, doi: 10.1109/TKDE.2003.1209005.
[27] K. Saruladha, G. Aghila, and A. Bhuvaneswary, "COSS: Cross ontology semantic similarity measure An information content based approach," in *International Conference on Recent Trends in Information Technology, ICRTIT 2011*, Jun. 2011, pp. 485–490, doi: 10.1109/ICRTIT.2011.5972360.

[28]   H. Al-Mubaid and H. A. Nguyen, "Measuring semantic similarity between biomedical concepts within multiple ontologies," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 39, no. 4, pp. 389–398, Jul. 2009, doi: 10.1109/TSMCC.2009.2020689.

[29]   N. A. Omar, S. Kasim, M. F. Md Fudzee, A. A. Ramli, H. Mahdin, and S. C. Sen, "A review on feature based approach in semantic similarity for multiple ontology," *Acta Informatica Malaysia*, vol. 1, no. 1, pp. 7–9, Feb. 2017, doi: 10.26480/aim.01.2017.07.09.

[30]   A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, no. 4, pp. 327–352, 1977, doi: 10.1037/0033-295X.84.4.327.

[31]   M. A. Rodríguez and M. J. Egenhofer, "Determining semantic similarity among entity classes from different ontologies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 2, pp. 442–456, Mar. 2003, doi: 10.1109/TKDE.2003.1185844.

[32]   E. G. M. Petrakis, G. Varelas, A. Hliaoutakis, and P. Raftopoulou, "X-Similarity: Computing semantic similarity between concepts from different ontologies," *Journal of Digital Information Management*, vol. 4, no. 4, pp. 233–237, 2006.

[33]   B. Saleena and S. K. Srivatsa, "Using concept similarity in cross ontology for adaptive e-Learning systems," *Journal of King Saud University - Computer and Information Sciences*, vol. 27, no. 1, pp. 1–12, Jan. 2015, doi: 10.1016/j.jksuci.2014.03.007.

[34]   W. He, X. Yang, and D. Huang, "A hybrid approach for measuring semantic similarity between ontologies based on WordNet," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7091 LNAI, IEEE, 2011, pp. 68–78.

[35]   D. Sánchez, A. Solé-Ribalta, M. Batet, and F. Serratosa, "Enabling semantic similarity estimation across multiple ontologies: An evaluation in the biomedical domain," *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 141–155, Feb. 2012, doi: 10.1016/j.jbi.2011.10.005.

[36]   M. Cheatham and P. Hitzler, "Conference v2.0: An uncertain version of the OAEI conference benchmark," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8797, pp. 33–48, 2014, doi: 10.1007/978-3-319-11915-1_3.

[37]   N. F. Noy and M. A. Musen, "The PROMPT suite: Interactive tools for ontology merging and mapping," *International Journal of Human Computer Studies*, vol. 59, no. 6, pp. 983–1024, Dec. 2003, doi: 10.1016/j.ijhcs.2003.08.002.

[38]   N. F. Noy and M. A. Musen, "{PROMPT}: Algorithm and tool for automated ontology merging and alignment," *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on on Innovative Applications of Artificial Intelligence, July 30 - August 3, 2000, Austin, Texas, USA*, pp. 450–455, 2000.

[39]   H. Schmid, "Probabilistic part-of-speech tagging using decision trees," *Proceedings of the International Conference on New Methods in Language Processing*, 1994.

[40]   G. A. Miller, "WordNet: A lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995, doi: 10.1145/219717.219748.

[41]   Y. Bounab, M. Oussalah, and A. Taleb-ahmed, "Semantic similarity approach between two sentences," *ResearchGate*, no. October, pp. 1–7, 2019.

BIOGRAPHIES OF AUTHORS

Leila Benaissa Kaddar 🆔 📊 SC Ⓟ Master assistant in computer science at the University of Mustapha Stambouli, Mascara (Algeria). She is a teacher in the department of process engineering. She received her master degrees from University of Sciences and Technology of Oran, Mohamed-Boudiaf USTOMB, Algeria. Currently, she is preparing her Phd in computer science in Djillali liabes University Sidi Bel Abbes, Algeria. She can be contacted at email: Leila.benaissa_kaddar@univ-sba.dz.

Farah Ben-Naoum 🆔 📊 SC Ⓟ received an engineering degree in computer status, the master's degree, a PhD degree and HDR in artificial intelligence from DjillaliLiabès University, Algeria, in 1997, 2002, 2010, and 2016, respectively. Now, she is a professor and researcher at the Department of Computer Science of DjillaliLiabès University. Since 2003, she is the member of Evolutionary Engineering and Distributed Information Systems Laboratory. Her research interests include architectural plant modeling, tree compression, tree editing distance, and heuristics for bioinformatics. She can be contacted at email: farah.bennaoum@univ-sba.dz.