

# An Information Content and Set of Common Superconcepts-Based Algorithm to Estimate Similarity between Concepts of Ontologies

Gbede Sylvain Gbame<sup>1\*</sup>, Maho Wielfrid Morie<sup>2</sup>, Konan Marcelin Brou<sup>1</sup>

<sup>1</sup>Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire

<sup>2</sup>Laboratoire de Mathématiques Informatique, Université Nangui Abrogoua, Abidjan, Côte d'Ivoire

Email: \*gbede.gbame18@inphb.ci

**How to cite this paper:** Gbame, G.S., Morie, M.W. and Brou, K.M. (2023) An Information Content and Set of Common Superconcepts-Based Algorithm to Estimate Similarity between Concepts of Ontologies. *Open Journal of Applied Sciences*, 13, 1896-1909. <https://doi.org/10.4236/ojapps.2023.1311150>

**Received:** October 10, 2023

**Accepted:** November 7, 2023

**Published:** November 10, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Ontologies have been used for several years in life sciences to formally represent concepts and reason about knowledge bases in domains such as the semantic web, information retrieval and artificial intelligence. The exploration of these domains for the correspondence of semantic content requires calculation of the measure of semantic similarity between concepts. Semantic similarity is a measure on a set of documents, based on the similarity of their meanings, which refers to the similarity between two concepts belonging to one or more ontologies. The similarity between concepts is also a quantitative measure of information, calculated based on the properties of concepts and their relationships. This study proposes a method for finding similarity between concepts in two different ontologies based on feature, information content and structure. More specifically, this means proposing a hybrid method using two existing measures to find the similarity between two concepts from different ontologies based on information content and the set of common superconcepts, which represents the set of common parent concepts. We simulated our method on datasets. The results show that our measure provides similarity values that are better than those reported in the literature.

## Keywords

Ontology, Data Structure, Similarity Measure, Concepts, Information Content

## 1. Introduction

Ontologies are manifestations of the common understanding of a domain. His domain is agreed between several agents, whose agreement in turn facilitates re-

liable, meaningful communications. All these communications in turn lead to other benefits such as interoperability, reuse and sharing [1] [2]. Therefore, communication or sharing of information and knowledge between different concepts has become a fundamental trend in computational linguistics [3].

One of the most important activities in the study of ontologies is semantic similarity between words. This activity is necessary in many domains such as medicine [4], learning [5], large-scale storage systems and data management [6]. Semantic similarity is the set of correspondences between entities (e.g., classes, properties, and predicates) that form ontology. Semantic similarity is also a measure used to calculate the similarity between two concepts. In most existing works, the semantic similarity between two concepts of a same ontology was calculated.

However, for a given domain (e.g., transport, health), there are several ontologies designed whose exploration for the correspondence of the semantic content of concepts is increasingly difficult. Moreover, according to [7], by using more than one ontology, it provides additional knowledge which improves the study of similarity and solves all cases of semantic similarity measurement whose concepts are not represented in only one ontology.

Given two ontologies in the same domain and two concepts belonging respectively to each of these ontologies, the problem studied here is to find an efficient measure of semantic similarity between these two concepts.

To solve this problem, we propose a method of semantic similarity between concepts from different ontologies.

The main contributions of our study are as follows:

- The proposition of new versions of two existing measures (*i.e.*, Jiang & Conrah and Lin measures) to consider the fact that concepts belong to different ontologies.
- The proposition of a method that considers the Information Content (IC) and the set of Common Superconcepts (CS), which represents the set of concepts of common parents. Indeed, the knowledge revealed by the concepts is used to augment the information already presented in ontologies or taxonomies [8]. This method includes use of the two new measures proposed (Equation (8) and Equation (9)).

The remainder of this paper is organised as follows. Section 2 presents related works. Section 3 presents the problem formulation. Section 4 describes the proposed method. Sections 5 and 6 are related to the evaluated results, and the conclusion and future work respectively.

## 2. Related Works

Several methods have been proposed to determine semantic similarities. These methods are classified into two types: semantic similarity classification for a single ontology and semantic similarity classification for multiple ontologies [9]. In the first type, semantic similarity is studied for concepts belonging to a same

ontology. In the second type, semantic similarity is studied for concepts belonging to different ontologies. Semantic similarity for a single ontology includes four methods: Structure-based, information content-based, feature-based, and hybrid-based methods. While semantic similarity between concepts of different ontologies includes two methods: length-based and function-based methods.

## **2.1. Semantic Similarity Based on a Single Ontology**

### **2.1.1. Structure-Based Method**

The structure-based method considers the shortest path between concepts in an ontology to calculate the semantic similarity. This method calculates the shortest path while the degree of similarity is determined according to the path length. There are various measurements on structure based method which have been used by [10] [11]. [10] [11] proposed a measure, called distance on the power set of nodes in a semantic network. The distance is the average minimum path length over all combinations of nodes between two subsets of nodes. While [12] [13] [14] discuss the depth method which takes into account the depth of the edges connecting two concepts in the ontology structure. This method calculates the depth between the root and the target concept.

### **2.1.2. Information Content-Based Method**

The information content-based method is also known as the corpus-based method. It determines the similarity between two concepts based on the probabilities of each concept in the ontology and the occurrences of words in a given corpus [8]. This knowledge revealed by the corpus is used to augment the information already presented in ontologies or taxonomies. It consists of different measures including the Resnik [15], Lin measure [16], Jiang and Conrath measures [17].

### **2.1.3. Feature-Based Method**

Feature-based methods have attempted to overcome the limitations of structure-based method. Because the taxonomic links in an ontology do not necessarily represent uniform distances. The problem of structure-based method is solved by considering the degree of overlap between the ontological feature sets. As a result, they are more general and can potentially be applied in ontology similarity estimation contexts, where pairs of concepts belong to two different ontologies. A situation in which edge counting methods cannot be applied directly [18]. Thus, unlike edge counting measures which are based on the notion of minimum path distance, feature-based approaches evaluate the similarity between concepts according to their properties [19].

### **2.1.4. Hybrid-Based Method**

The hybrid method considers different sources of information to calculate the similarity score between concepts. This method is a mixture of several characteristics, namely: attribute similarity, ontology structure, information content and Lowest Common Ancestor (LCA) node depth. This method has been the

subject of several studies, including [20] which proposed a hybrid method that combines term-based and concept-based methods. These measures concern structure and contextual relevance measures. The major advantage of this approach is that if the knowledge of an information source is inadequate, it can be derived from other sources [21]. Thus, the quality of the similarity measure would be improved. Among the works based on this method, we can cite [22] and [23].

The information content method is the most widely used for estimating the similarity between two concepts in the same ontology. However, two concepts from multiple ontologies can be similar.

## 2.2. Semantic Similarity Based on Multiple Ontologies

The last ten years, the increase in the number of resources on the web has led researchers to introduce methods that calculate the similarity between concepts belonging to several ontologies.

These methods are:

- Structure-based method.
- Feature-based method.

### 2.2.1. Structure-Based Method

These methods have been proposed to evaluate the similarity of concepts by exploiting different knowledge. Some of these methods have been adapted to the biomedical domain to study clinical information [4]. Works in [4] presented and analysed methods to determine their advantages and disadvantages. Using an input such as SNOMED CT [24], a new method based on the exploitation of the taxonomic structure of a biomedical ontology was proposed. This proposed similarity method achieves a similar level of precision as corpus-based approaches, while retaining the low computational complexity and lack of constraints of path-based method.

### 2.2.2. Feature-Based Method

Works in [24] proposed a method for measuring semantic similarity between two concepts  $c_1$  and  $c_2$  that belong to two different ontologies, each of which models knowledge from a different perspective. This method is based on the evaluation of the union of the set of superconcepts of  $c_1$  and  $c_2$  in each ontology and on the search for equivalences between them. The set of shared superconcepts for  $c_1$  belonging to ontology  $O_1$  and  $c_2$  to ontology  $O_2$  is composed of the superconcepts of  $c_1$  and  $c_2$  having the same label, as well as the subsumers of these equivalent superconcepts. Three different sets are highlighted:

- The set of superconcepts defined by Equation (1) and Equation (2) were proposed in [24]

$$T_{O_1}(c_1) = \{c_j \in C_{O_1} \mid c_j > c_1\} \cup \{c_1\} \quad (1)$$

$$T_{O_2}(c_2) = \{c_j \in C_{O_2} \mid c_j > c_2\} \cup \{c_2\} \quad (2)$$

The set of superconcepts of a given concept  $c_i$  is all the concepts that are linked to  $c_i$  from the root to this concept  $c_i$ , including  $c_i$  itself.

- The set of terminologically Equivalent Superconcepts (ES) in  $T_{O_1}(c_1) \cup T_{O_2}(c_2)$  defined by were proposed in [24]:

$$ES = \{c_i \in T_{O_1}(c_1) \mid \exists c_j \in T_{O_2}(c_2) \text{ and } c_i \equiv c_j\} \quad (3)$$

In Equation (3),  $\equiv$  means terminological correspondence.

- Finally, the set of Common Superconcepts (CS) in  $T_{O_1}(c_1) \cup T_{O_2}(c_2)$  defined by were proposed in [24]:

$$CS(c_1, c_2) = \bigcup_{c_i \in ES} (T_{O_1}(c_i) \cup T_{O_2}(c_i)) \quad (4)$$

$T_{O_1}(c_1)$ : The set of superconcepts of concept  $c_1$  in concept hierarchy  $H_{c_1}^C$  of concepts  $c_{O_1}$  of ontology  $O_1$  (Equation (1)).

$T_{O_2}(c_2)$ : The set of superconcepts of concept  $c_2$  in hierarchy  $H_{c_2}^C$  of concepts  $c_{O_2}$  of ontology  $O_2$  (Equation (2)).

The set of common superconcepts (CS) in  $T_{O_1}(c_1) \cup T_{O_2}(c_2)$  is composed of the elements in ES and all superconcepts of the elements in ES (Equation (3) and Equation (4)). Unlike structure-based methods for calculating similarity, this similarity method considers all parents. It also evaluates sets of superconcepts instead of paths. This method exploits the taxonomic structure of the biomedical ontology. Therefore, it does not consider taxonomic depths such as information content [24] [25].

### 3. Problem Specification

In this section, the assumptions, notations, and definitions of the problem are given.

#### 3.1. Problem Statement

Generally, for a given domain, several ontologies are relevant to each other. In many works most similarity methods did not support more than one input ontology. However, for the same domain, some concepts belonging to different ontologies are semantically equivalent [7]. Therefore, the problem is to estimate a similarity measure between two concepts belonging to different ontologies.

#### 3.2. Assumption and Notation

We make the following assumptions:

- Assumption 1: All ontologies belong to the same domain. However, the two concepts studied belong to two different ontologies.
- Assumption 2: The concepts whose similarity we want to estimate have at least two terminologically equivalent superconcepts.

#### 3.3. Problem Formulation

**Given:** Two ontologies  $O_1$  and  $O_2$ , two concepts  $c_1$  and  $c_2$  respectively of  $O_1$  and  $O_2$ .

**Goal:** Estimate a similarity measure between two concepts  $c_1$  and  $c_2$  from multiple ontologies.

#### 4. Proposed Method

This section presents our method to solve the semantic similarity problem.

Consider the set of ontologies  $O = \{O_1, O_2\}$ . Let  $c_1$  and  $c_2$  be two concepts such that  $c_1 \in O_1$  and  $c_2 \in O_2$ . The algorithm determines the semantic similarity of two concepts  $c_1$  and  $c_2$  from two different ontologies  $O_1$  and  $O_2$  respectively. Indeed, the algorithm uses the set of common superconcepts, information content IC values and terminologically equivalent superconcepts ES.

Information content (IC) is used in our method because it determines the similarity between two concepts based on the probabilities of each concept in the ontology and the occurrences of words in a given corpus [8]. This knowledge revealed by the corpus is used to augment the information already presented in ontologies or taxonomies.

When we refer to ontologies  $O_1$  and  $O_2$ , their common superconcepts  $CS(c_1, c_2)$  is equal to  $\{root_1, s_1, s_3, s_{LCS}, root_2, r_3, r_1\}$ . Let  $Q$  be the sum of the information content of each concept belonging to the CS such that:

$$Q = IC(root_1) + IC(s_1) + IC(s_3) + IC(s_{LCS}) + IC(root_2) + IC(r_3) + IC(r_1) \quad (5)$$

- In [21], the expression for the semantic similarity measure between two concepts  $c_1$  and  $c_2$  in a single ontology according to Lin and Jiang & Conrath is represented respectively by Equation (6) and Equation (7) were proposed in [26]:

$$Sim_{ACS-Lin}(c_i, c_j) = \frac{2 * \sum_{k=1}^n IC(c_k)}{IC((c_i)) + IC((c_j))} \quad (6)$$

$$Sim_{ACS-Jc}(c_i, c_j) = \mathbb{P} - \frac{[IC(c_i) + IC(c_j)] - 2 * \sum_{k=1}^n IC(c_k)}{2} \quad (7)$$

where:

- $IC(x)$  represents the information content of the concept  $x$ .
- In our work,  $CS(c_1, c_2)$  is equal to  $\{root_1, s_1, s_3, s_{LCS}, root_2, r_3, r_1\}$ .
- The set of common superconcepts CS is defined as the most specific superconcepts (the one with the depth) that define the path between the pair of evaluated concepts. It is the union of the set of superconcepts  $T_{O_1}(c_1)$ , of concept  $c_1$ , and that of  $T_{O_2}(c_2)$  of concept  $c_2$ .
- $\sum_{k=1}^n IC(c_k)$  represents the sum of the information content of all common parents along the paths for the concepts whose similarity is to be calculated in a single ontology.

To calculate the similarity of concepts from two ontologies, it is necessary to consider the most specific superconcepts which define the path between the pair of concepts being evaluated (in our case,  $c_1$  and  $c_2$ ). Our work proposes integrating the information content of  $Q$  superconcepts into Equations (6) and (7). It

results our proposed measures through Equation (8) and Equation (9):

$$Sim_{SCS-Lin}(c_1, c_2) = \frac{2 * Q}{IC((c_1)) + IC((c_2))} \quad (8)$$

$$Sim_{SCS-Jc}(c_1, c_2) = 1 - \frac{[IC(c_i) + IC(c_j)] - 2 * Q}{2} \quad (9)$$

### Description of Our Algorithm

The literature review revealed that in [27], an algorithm can determine the semantic similarity between two concepts  $c_1$  and  $c_2$  from the same ontology. This algorithm is based on the information content of concepts, therefore by re-checking the path between the concepts. However, the formulation of this algorithm does not allow to determine the semantic similarity of two concepts from different ontologies, which is addressed here. In other words, using the proposed algorithm in [27], finding the semantic similarity between two concepts from two ontologies is almost impossible. Because if the information content alone is used for the determination of the semantic similarity in two different ontologies, it will not succeed in establishing the taxonomic link between the two concepts studied.

Recall that [24] proposes a measure based on the exploitation of the taxonomic structure of a biomedical ontology. The algorithm chooses two concepts  $c_1$  and  $c_2$  respectively in the ontologies  $O_1$  and  $O_2$ . These two concepts are the objects of similarity measurement research.

Our algorithm consists of three steps. Indeed, the algorithm chooses two concepts  $c_1$  and  $c_2$  respectively in the ontologies  $O_1$  and  $O_2$ . These two concepts are the objects of similarity measurement research. The proposed algorithm consists of 4 main steps:

- Step 1 (Refer to Line 3) calculates the information content of a concept. For each concept  $c_i$  belonging to the set of concepts  $C$  or  $C'$ , the information content of  $c_i$  is then calculated from the number of descendants of  $c_i$  and the total number of concepts Max in ontology  $O$ .
- Step 2 (lines 4 to 6) involves extracting for concepts  $c_1$  and  $c_2$ , the set of superconcepts  $T_{O_1}(c_1)$ , and  $T_{O_2}(c_2)$ , the set of terminologically equivalent superconcept ( $ES$ ), then the set of common superconcepts ( $CS$ ).
- Step 3 (Line 7) calculates  $Q$ , which is the sum of the information content  $IC$  of each concept in the set of common superconcepts  $CS$ .
- Step 4 (line 8) calculates the semantic similarity between concepts  $c_1$  and  $c_2$  according to Equation (8) and Equation (9).

**Algorithm 1** allows us to provide experimental results to support our approach in the next section of our paper (Section 5.1).

### 5. Performance Evaluation

We conducted a simulation to compare the efficiency of the methods. Python was used as the programming language during the simulation. This simulation

**Algorithm 1.** SemSim\_SCS\_  $c_1, c_2$ .

**Input:**  $O_1, O_2$  (input ontologies),  $c_1, c_2$  (any two concepts from  $O_1, O_2$  respectively)

**Output:** Value of Semantic similarity measures between  $c_1$  and  $c_2$

**Begin**

1. Let  $C$  and  $C'$  be the ordered set of all concepts respectively of  $O_1$  and  $O_2$ ,  $root_1$  and  $root_2$  the root respectively  $O_1$  and  $O_2$ .
2. Let  $MaxOnt(O_i)$  be the number of concepts in  $O_i$  (where  $O_i \in \{O_1, O_2\}$ )
3. For  $c_i \in C \cup C'$ 
  - 3.1  $num\_desc(c_i)$  = The number of descendants of concept  $c_i$ ;  $Max = 0$
  - 3.2 If  $c_i \in C$  then
    - 3.2.1  $Max = MaxOnt(O_1)$
  - 3.3 Else
    - 3.3.1  $Max = MaxOnt(O_2)$
  - 3.4 Compute  $IC(c_i) = 1 - \frac{\log(num\_desc(c_i) + 1)}{Max}$
4. Extract  $T_{O_1}(c_1)$  and  $T_{O_2}(c_2)$ , set of superconcepts of  $c_1$  and  $c_2$  respectively.
5. Extract ES in  $T_{O_1}(c_1) \cup T_{O_2}(c_2)$  according to Equation (3)
6. Extract CS (set of common superconcepts) in  $T_{O_1}(c_1) \cap T_{O_2}(c_2)$  using ES and according to Equation (4)
7. Compute  $Q$  with IC of all CS concept
8. Using  $IC(c_1)$  and  $IC(c_2)$  values calculated in 3.4 and  $Q$  to find the similarity  $Sim_{SCS-Lin}(c_1, c_2)$  and  $Sim_{SCS-Jc}(c_1, c_2)$  according to Equation (8) and Equation (9)

**End.**

was performed on a computer equipped with a CPU, i7-2620M @ 2.70 GHz RAM, and 128 GB SSD with a windows 10 operating system.

### 5.1. Simulation Process

The simulation process followed the same steps as those described in the literature. Lin and Jiang & Conrath proposed a comparison of existing measures on 15 pairs of concepts from two reference datasets [26]. The simulation of their proposal confirmed that some concepts are equivalent or roughly equivalent. For example, the comparison shows that for concepts such as cheese and ice cream, eye and chin, lung and liver, the similarity value is the same for the existing and proposed measures, because the concepts are subsumed by numerous super-concepts.

In our simulation, in addition to  $s_3 \equiv r_3$  for ontologies  $O_1$  and  $O_2$ , we considered the Least Common Subsumer (LCS) of the concepts defined in **Tables 1-3**. The LCS represents the superconcept whose maximum depth defines the path between the pair of concepts evaluated (**Figure 1**). These are  $s_{LCS}$  for ontology  $O_1$  and  $r_{LCS}$  for ontology  $O_2$ . Next, we compared the pairs of concepts derived from the descendants of  $s_3$  and  $r_3$ .





**Table 3.** Comparison of proposed and existing measures for non-similar pairs.

Pair of Concepts	Sim_Lin	Sim_SCS_Lin	Sim_JC	Sim_SCS_JC
( $s_{17}, r_{25}$ )	0.193561	0.679437	0.188217	0.646341
( $s_{16}, r_{11}$ )	0.189912	0.666629	0.186405	0.644530
( $s_{23}, c_3$ )	0.195409	0.685924	0.189108	0.647233
( $s_{14}, r_{18}$ )	0.195421	0.685966	0.189114	0.647239
( $c_4, r_{19}$ )	0.192314	0.675060	0.187605	0.645730
( $s_{23}, r_{21}$ )	0.193500	0.679226	0.188187	0.646312

## 5.2. Results

After the simulation, the results are as follows. The simulation performed obeys the condition of Section 5.1, which is  $s_3 \equiv r_3$ .

1) The similarity measure is calculated using all concepts that have a relationship in each ontology with each of the equal concepts including  $s_3$  in the ontology  $O_1$  and  $r_3$  in ontology  $O_2$ . Thus, in the ontology  $O_1$  the concepts concerned are:

$$Tab\_sim\_O_1 = [s_4, s_{10}, c_1, c_4, s_{12}, s_{13}, s_{14}, s_{15}, s_{16}, s_{17}, s_{18}, s_{19}, s_{20}, s_{21}, s_{22}, s_{23}, s_{24}, s_{25}, s_{26}, s_{27}, s_{28}, s_{29}, s_{30}]$$

In ontology  $O_2$ :

$$Tab\_sim\_O_2 = [r_4, r_{11}, r_{12}, c_2, r_{10}, r_{17}, r_{18}, r_{19}, r_{20}, r_{21}, r_{22}, r_{23}, r_{24}, r_{25}, r_{26}, r_{27}, r_{28}, r_{29}, r_{30}, r_{31}]$$

2) Each concept of  $Tab\_sim\_O_1$  is associated with  $Tab\_sim\_O_2$  for the calculation of the similarity. First, the first concept  $s_4$  of  $Tab\_sim\_O_1$  is associated with the first concept  $r_4$  of  $Tab\_sim\_O_2$ . After  $r_4$  of  $Tab\_sim\_O_2$ ,  $s_4$  is always associated with  $r_{11}$  of  $Tab\_sim\_O_2$  to calculate the similarity. The same process is repeated between  $s_4$  of  $Tab\_sim\_O_1$  and all the other concepts of  $Tab\_sim\_O_2$ , until the last concept of  $Tab\_sim\_O_2$ . After  $s_4$ , the next concept  $s_{10}$  of  $Tab\_sim\_O_1$  is chosen to perform the same process.

3) After having traversed all the concepts of  $Tab\_sim\_O_1$  and  $Tab\_sim\_O_2$ , the various results of the similarity calculation are compared. The results of the semantic similarities of each pair were calculated. Then, the highest value is kept for examination [27].

## 6. Discussion

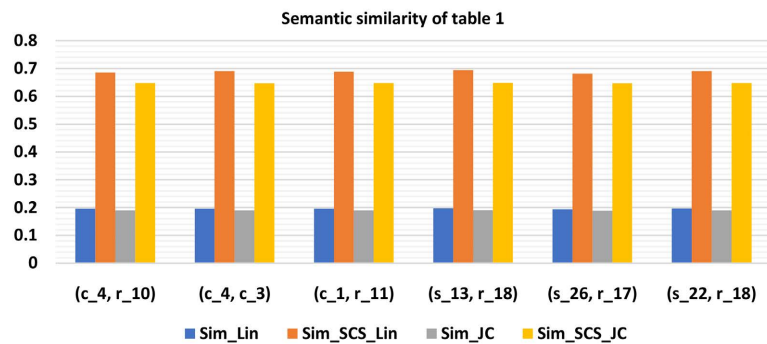
In **Table 1**, all concepts are subsumed by numerous super concepts, except the last concepts of each ontology. This allows our proposed measure to provide good results. The pairs ( $c_4, r_{10}$ ), ( $c_4, c_3$ ) and ( $c_1, r_{11}$ ) have the same similarity value of Lin and JC (0.195925 for Lin and 0.189355 for JC), while the similarity values of our approach SCS\_Lin and SCS\_JC of these pairs vary and are larger than those reported in the literature.

In **Table 2**, the pairs have the same similarity values for the proposed and the existing measure: ( $s_{24}, r_{22}$ ) and ( $s_{21}, r_{22}$ ) whose values are [0.195059 for Lin,

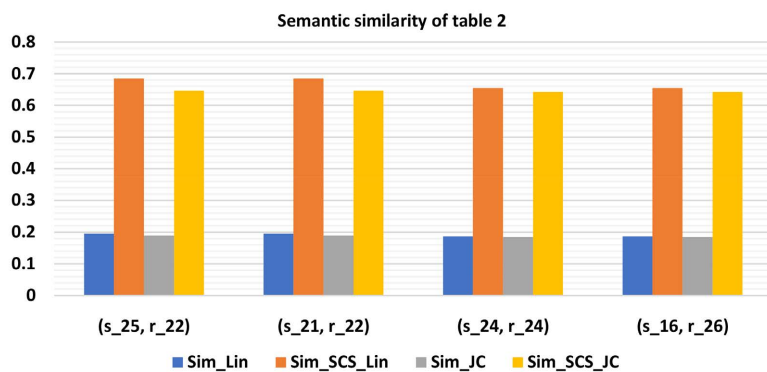
0.188941 for J & C, 0.684695 for SCS\_Lin, 0.647065 for SCS\_JC]. The same is true for the pairs  $(s_{26}, r_{26})$  and  $(s_{16}, r_{26})$  whose values are [0.186517 for Lin, 0.184656 for J & C, 0.184656 for SCS\_Lin, 0.642780 for SCS\_JC]. Our measure shows better results of the similarity measure for concepts that are subsumed by many superconcepts.

In **Table 3**, all values of similarity measures for the proposed and the existing measure are different for all pairs. However, the proposed measures are higher than the existing measures.

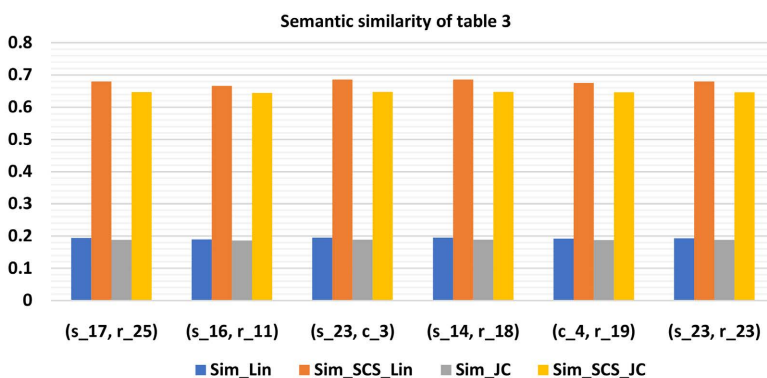
The different comparison graphs of the concepts explained above are illustrated in **Figures 2-4**. It results that the similarity values from our measures



**Figure 2.** 1st round of comparison of proposed and existing measures.



**Figure 3.** 2nd round of comparison of proposed and existing measures.



**Figure 4.** 3rd round of comparison of proposed and existing measures.

(SCS\_Lin and SCS\_J & C) are better than those existing (Lin and Jin). These different results lead us to the conclusion that our method is more efficient to calculate the semantic similarity between two concepts belonging to two different ontologies.

## 7. Conclusions

The problem studied in this paper is to find the semantic similarity between the concepts of two different ontologies. Existing works such as [27] have proposed an approach to calculate the semantic similarity of two concepts in the same ontology. However, several concepts are from different ontologies.

To overcome this difficulty, we proposed:

- A hybrid method based on the method of [24] and [27]. Our method is based on information content and set of common superconcepts. This method uses our proposed versions of existing measures.

The simulation confirmed that this method is more effective in solving the problem studied.

According to [2] by using more than one ontology, it provides additional knowledge that improves the similarity study and solves all cases where the terms are not represented in an ontology. Therefore, our method is an essential tool that can be used by ontology developers to evaluate several ontologies in the decision process.

Our method can provide semantic similarity between only two concepts from two ontologies.

- However, taking a lot of concepts from a lot of ontologies as input can be tricky to solve. So, future work could explore this issue.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Chandrasekaran, D. and Mago, V. (2022) Evolution of Semantic Similarity—A Survey. *ACM Computing Surveys*, **54**, 1-37. <https://doi.org/10.1145/3440755>
- [2] Saruladha, K., Aghila, G. and Bhuvaneshwary, A. (2011) Information Content Based Semantic Similarity Approaches for Multiple Biomedical Ontologies. In: Abraham, A., Lloret Mauri, J., Buford, J.F., Suzuki, J. and Thampi, S.M., Éds., *Advances in Computing and Communications*, Springer, Berlin, 327-336. [https://doi.org/10.1007/978-3-642-22714-1\\_34](https://doi.org/10.1007/978-3-642-22714-1_34)
- [3] Agarwal, P. (2004) Contested Nature of Place: Knowledge Mapping for Resolving Ontological Distinctions between Geographical Concepts. In: Egenhofer, M.J., Freksa, C. and Miller, H.J., Éds., *GIScience 2004: Geographic Information Science*, Springer, Berlin, 1-21. [https://doi.org/10.1007/978-3-540-30231-5\\_1](https://doi.org/10.1007/978-3-540-30231-5_1)
- [4] Yang, F., Xing, Y., Sun, H., Sun, T. and Chen, S. (2015) An Ontology-Based Semantic Similarity Measure Considering Multi-Inheritance in Biomedicine. *Mathematical Problems in Engineering*, **2015**, Article ID: 305369.

- <https://doi.org/10.1155/2015/305369>
- [5] Dagnaw, G., Zhou, K. and Wang, H. (2019) dCACH: Content Aware Clustered and Hierarchical Distributed Deduplication. *Journal of Software Engineering and Applications*, **12**, 460-490.
- [6] Hardman, J. and Set, B. (2021) Investigating the Double-Move in Pedagogy in a Grade 4 Namibian Science Classroom: A Cultural Historical Analysis. *Creative Education*, **12**, 453-469.
- [7] Al-Mubaid, H. and Nguyen, H.A. (2009) Measuring Semantic Similarity between Biomedical Concepts within Multiple Ontologies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **39**, 389-398.  
<https://doi.org/10.1109/TSMCC.2009.2020689>
- [8] Gottlieb, A., Stein, G.Y., Ruppim, E. and Sharan, R. (2011) PREDICT: A Method for Inferring Novel Drug Indications with Application to Personalized Medicine. *Molecular Systems Biology*, **7**, 496. <https://doi.org/10.1038/msb.2011.26>
- [9] Ibrahim, R., Zeebaree, S. and Jacksi, K. (2019) Survey on Semantic Similarity Based on Document Clustering. *Advances in Science, Technology and Engineering Systems Journal*, **4**, 115-122. <https://doi.org/10.25046/aj040515>
- [10] Bulskov, H., Knappe, R. and Andreasen, T. (2002) On Measuring Similarity for Conceptual Querying. In: Carbonell, J.G., Siekmann, J., Andreasen, T., Christiansen, H., Motro, A. and Legind Larsen, H., Eds., *FQAS 2002: Flexible Query Answering Systems*, Springer, Berlin, 100-111. [https://doi.org/10.1007/3-540-36109-X\\_8](https://doi.org/10.1007/3-540-36109-X_8)
- [11] Bulskov, H., Knappe, R. and Andreasen, T. (2004) On Querying Ontologies and Databases. In: Christiansen, H., Hacid, M.S., Andreasen, T. and Larsen, H.L., Eds., *FQAS 2004: Flexible Query Answering Systems*, Springer, Berlin, 191-202.  
[https://doi.org/10.1007/978-3-540-25957-2\\_16](https://doi.org/10.1007/978-3-540-25957-2_16)
- [12] Wu, Z.B. and Palmer, M. (1994) Verb Semantics and Lexical Selection. *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, Las Cruces, 27-30 June 1994, 133-138. <https://doi.org/10.3115/981732.981751>
- [13] de Oliveira, A.V., Pimentel, C.M.O., Godina, R., de Oliveira Matias, J.C. and Garrido, S.M.P. (2022) Improvement of the Logistics Flows in the Receiving Process of a Warehouse. *Logistics*, **6**, Article 22. <https://doi.org/10.3390/logistics6010022>
- [14] Luna, D. (2005) Ontological Modelling of Information Systems from Bunge's Contributions.  
[https://www.academia.edu/4581161/Ontological\\_Modelling\\_of\\_Information\\_Systems\\_from\\_Bunges\\_Contributions](https://www.academia.edu/4581161/Ontological_Modelling_of_Information_Systems_from_Bunges_Contributions)
- [15] Vaid, A.J., Chaudhary, R., Vaid, A.J. and Chaudhary, R. (2022) Review Paper on Impact of Behavioral Biases in Financial Decision-Making. *World Journal of Advanced Research and Reviews*, **16**, 989-997.  
<https://doi.org/10.30574/wjarr.2022.16.2.1236>
- [16] Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.J. (1990) Introduction to WordNet: An On-Line Lexical Database. *International Journal of Lexicography*, **3**, 235-244. <https://doi.org/10.1093/ijl/3.4.235>
- [17] Leacock, C., Towell, G. and Voorhees, E. (1993) Corpus-Based Statistical Sense Resolution. *Human Language Technology. Proceedings of a Workshop*, Plainsboro, 21-24 March 1993, 260-265.
- [18] Petrakis, E., Varelas, G., Chliaoutakis, A. and Raftopoulou, P. (2006) X-Similarity: Computing Semantic Similarity between Concepts from Different Ontologies. *JDIM: Journal of Digital Information Management*, **4**, 233-237.
- [19] Tversky, A. (1977) Features of Similarity. *Psychological Review*, **84**, 327-352.

- <https://doi.org/10.1037/0033-295X.84.4.327>
- [20] Sumathy, K.L. and Chidambaram, D. (2016) A Hybrid Approach for Measuring Semantic Similarity between Documents and Its Application in Mining the Knowledge Repositories. *International Journal of Advanced Computer Science and Applications*, **7**, 231-237. <https://doi.org/10.14569/IJACSA.2016.070831>
- [21] Elavarasan, K., Naveen Kumar, V. and Shamasundar, B.A. (2014) Antioxidant and Functional Properties of Fish Protein Hydrolysates from Fresh Water Carp (*Catla catla*) as Influenced by the Nature of Enzyme. *Journal of Food Processing and Preservation*, **38**, 1207-1214. <https://doi.org/10.1111/jfpp.12081>
- [22] Gonçalves, R.S., Kamdar, M.R. and Musen, M.A. (2019) Aligning Biomedical Metadata with Ontologies Using Clustering and Embeddings. In: Hitzler, P., et al. Eds., *ESWC 2019: The Semantic Web*, Springer, Cham, 146-161. [https://doi.org/10.1007/978-3-030-21348-0\\_10](https://doi.org/10.1007/978-3-030-21348-0_10)
- [23] Chiang, J.H. and Yu, H.C. (2005) Literature Extraction of Protein Functions Using Sentence Pattern Mining. *IEEE Transactions on Knowledge and Data Engineering*, **17**, 1088-1098. <https://doi.org/10.1109/TKDE.2005.132>
- [24] Batet, M., Sánchez, D., Valls, A. and Gibert, K. (2013) Semantic Similarity Estimation from Multiple Ontologies. *Applied Intelligence*, **38**, 29-44. <https://doi.org/10.1007/s10489-012-0355-y>
- [25] Harispe, S., Imoussaten, A., Troussset, F. and Montmain, J. (2015) Utilisation des fonctions de croyance pour l'estimation du contenu informationnel des concepts d'une ontologie. Rencontre francophone sur la Logique Floue et ses applications LFA, Poitiers.
- [26] Mohd, A. (2020) Information Content Based Semantic Similarity Measure. *International Journal Web Applications*, **7**, 85-94.
- [27] Banu, A. (2015) Information Content Based Semantic Similarity Measure for Concepts Subsumed. *International Journal of Web Application*, **7**, 85-94. [https://www.dline.info/ijwa/fulltext/v7n3/v7n3\\_1.pdf](https://www.dline.info/ijwa/fulltext/v7n3/v7n3_1.pdf)