I.J. Information Technology and Computer Science, **2015, 07, 48-56**
Published Online June 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2015.07.06

A New Measure of the Calculation of Semantic Distance between Ontology Concepts

Abdeslem DENNAI
EEDIS Laboratory, Djillali Liabes University, Sidi Bel Abbes Algeria
de_selam@yahoo.fr

Sidi Mohammed BENSLIMANE
EEDIS Laboratory, Djillali Liabes University, Sidi Bel Abbes Algeria
Benslimane@univ-sba.dz

Abstract**—** Semantic similarity calculation models are found in many applications, with the aim to give additional knowledge to reason about their data. The choice of a similarity measure is quite crucial for a successful implementation of reasoning. In this work, we present an update of similarity calculation presented by Wu and Palmer which is considered the fastest in time generation of similarity. The results obtained show that the measure produced provides a significant improvement in the relevance of the values produced for the similarity of two concepts in ontology.

Index Terms— Ontology, Similarity Measure, Semantic Distance, Semantic Web, Semantic Association.

I. Introduction

Identifying the similarity was considered a highly recommended research topic in various domains like: semantic web [1], artificial intelligence and linguistic literature. The choice of a similarity measure is quite crucial for a successful implementation of reasoning [2]. This is, in fact, to find the best match between the goal and the kind of manipulated knowledge. The similarity identification between the data resulting from the extraction and the concepts of a domain ontology is a fundamental phase in a reverse engineering approach that is adopted by several techniques such as clustering, data mining, semantic web and, in particular, the information research domain. This latter is largely based on measures for the similarity identification between the documents [3]; [4]. In the Semantic Web field, where ontologies intervene to knowledge modeling, measurement of [5], for example, has the advantage of being simple to implement and have also performed well compared to other measures similarity [6].

Rummaging through the different similarity measure methods, we can deduce the limit of these methods in several application domains which led us to a general synthesis of these methods, completed by our contribution to the updating of a calculating the similarity method between the concepts of an ontology.

The calculation of the semantic similarity between concepts from different systems or domains is becoming an increasingly important task [2]. It plays a key role in, among others, information retrieval; service oriented computing, language automatic treatment and bioinformatics.

A. Information retrieval

Information retrieval is largely based on similarity measures to identify the similarity between the documents [3]; [4]. The majority of approaches to research information do not take into account only single words and / or fragments of words to search for relevant documents and ignore the essential idea that considers the ontological relations of words. These can be detected by a calculation process of similarity between pairs of objects [1].

B. Service oriented computing

With the application of semantic Web service, the similarity measures between services are more and more important in the processing of service matching. By formally defining the similarity of semantic services, useful information can be obtained about their similarity and compatibility. The determination of the semantic services similarity makes it possible to obtain useful information concerning their accountancies. [7] Propose metric to measure the semantic services similarity annotated with an ontology OWL. The proposed similarity measure is based on the intuition that the similar objects share the most common descriptive information.

C. Language automatic treatment

Several studies on the similarity measure were motivated by the automatic language processing. Among the works in this area are: the work of [8] which uses semantic similarity for measuring semantic similarity between all the senses of the word of a given pair of words and disambiguate well in a given context. [9] combined the use of a thesaurus automatically acquired from the raw text corpora and WordNet (based on the metric of similarity) to find predominant meaning of words in unstructured text. The authors of work [10] applied the WordNet semantic similarity measures for evaluating the relevance of expressions, given a specific dialogue and automatically build summaries of spoken dialogue, in this same domain, you can consult work of [11].

D.Bioinformatics

The large-scale effort in developing, maintaining and making biomedical ontology available motivates the application of similarity measures to compare ontology concepts or, by extension, the entities described therein. A common approach, known as semantic similarity, compares ontology concepts through the information content they share in the ontology. Ontology-based similarity has become a prominent approach to compare biomedical entities based on their biomedical activity [12]. A variation of similarity measure based on the informational content is adopted to find a better way to organize and interrogate a Gene Ontology data (GO). To calculate the semantic similarity between proteins, rather than the terms of GO ontology, authors in [13] combined between three similarities measures [14]; [6]; [14]. Furthermore, many similarity measures have been applied to biomedical ontology and compared against, traditional structural similarity measures [9]; [15]; [16].

Our contribution is shown in this domain by using an extract of this gene ontology (GO) [17], in addition to another extraction of travel ontology [18] in the fourth part of this work. While excavating in the various existing methods of similarity measure, we can deduce the limit from these methods in some domains of application what us led to make a general synthesis of these methods, completed by our contribution in the update of a method of similarity calculating between ontology concepts.

The reset of the paper is organized as follows: The second section presents a classification of the main approaches to measuring similarity. The third section describes some related work. In the fourth section, we present our contribution before concluding with some future perspectives in the fifth section.

II. TAXONOMY OF TECHNIQUES FOR MEASURING SIMILARITY

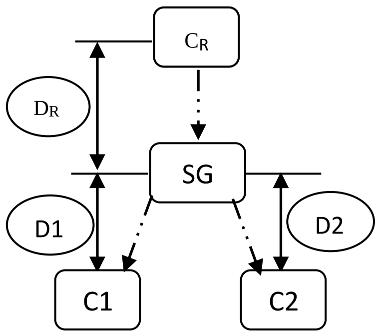
In this section, we present a classification of the main approaches to measuring similarity.

A. Techniques based on the arcs

The most intuitive similarity measure between objects in ontology is their distances [19]; [20]; [5]. Obviously, an object X is more similar to an object Y than an object Z, this similarity is evaluated by the distance between the objects in the ontology.

These measures use the hierarchical structure of the ontology to determine the semantic similarity between concepts. Calculating distances in the ontology is based on a specialization of object graph. In each graph, the distance of the ontology must be characterized by the shortest path that involves a common ancestor or the smallest generalizing (SG), potentially connecting two objects through common descendants. Among the works classified under this banner, we have:

Table 1. Similarity measures based on the arcs.

Similarity Measure	Description	Mathematical formula and commentaries
Wu and Palmer. [5]	It was used by [21] to organize web documents into clusters. It was also used in [22] to evaluate the semantic proximity of the two concepts of a HTML page to a relatively thesaurus through indexation of website by ontology. It's based on the following principle (See figure 1): Given an ontology formed by a set of nodes and a root node (C_R). C1 and C2 represent two ontology elements for which we will calculate the similarity. The principle of similarity calculation is based on the distances ($D1+D_R$) and ($D2+D_R$) separating the C1 and C2 nodes from the node C_R and the distance (D_R) separating the subsuming concept or the smallest generalizing of C1 and C2 of node C_R.	 Fig. 1. Example of an ontology extract. $SimWP(C1, C2) = \frac{2 \times D_R}{D1 + D2 + 2 \times D_R}$ (1)
Rada and al. [19]	This measure is adopted in a semantic network and it is based on the fact that we can calculate the similarity based on the hierarchical relationships "is-a". To calculate the similarity of two concepts in ontology, we must calculate the number of minimum arcs between them. This measure, based on the calculation of the distance between nodes by the shortest path, has an average of more obvious to evaluate the semantic similarity in a hierarchical ontology.	In the biomedical domain, [15] proposed the first semantic similarity measure by using path length between biomedical terms in the MeSH (Medical Subject Heading, [23]) ontology as a measure of semantic.
Ehrig and al. [24]	The similarity of the entities is measured in data by considering the simple or complex data type (integer, character). The semantic relations between the entities are measured at the layer of the ontology. Finally the context of the layer specifies how the entities of the ontology used in some external context, specifically, the context of the application.	This work introduces three layers: data, ontology and context.

B. Techniques based on the nodes

These techniques adopt a new measure in terms of the entropy measure of information theory [6] [26]. The probability P (.) for identifying the use of a class or its descendents in a corpus refers to the class information. The following formula defines the entropy of a class:

$$E(c) = -log(P) \qquad (2)$$

Where P is the probability of finding an instance of the concept c. the probability of concept c is calculated by dividing the number of instances of c by the total number of the instances. By associating the concepts of probability of taxonomy, it is possible to avoid the unreliability of the distances arcs. This quantitative characteristic of the information provides a new way of measuring the semantic similarity. More information is shared by two concepts, more they are similar. Among the work found in the literature under this banner, we can cite (see the following table):

Table 2. Similarity measure based on the nodes.

Similarity Measure	Description	Mathematical formula and commentaries
Resnik. [26]	The notion of the Informational Contents (IC) was initially introduced by [26] who proved that an object (word) is defined by the number of the specified classes and that the semantic similarity between two concepts is measured by the quantity of information they share. To evaluate the relevance of an object, calculate the IC. The IC is obtained by calculating the object frequency in the corpus (by using WordNet for example) [12].	$SimR(C1, C2) = Max\left[E\left(CS(C1, C2)\right)\right]$ $= Max\left[-log\left(p\left(CS(c1, c2)\right)\right)\right] \qquad (3)$ CS (C1, C2) represents the most specific concept (which maximizes the similarity value) which subsumes (located at a higher hierarchical level) the two concepts C1 and C2 in ontology.
Lin. [6]	It uses a hybrid approach that combines two different source of knowledge (Thesaurus, corpus). In addition, it represents the similarity as probabilistic degree of overlap of descendents concepts C1 and C2. The work of [27] evaluated this measure through an experiment that uses human subjects to evaluate the similarity between 30 pairs of names; it appears that this method offers a significant improvement [12].	$SimL\ (C1, C2) = \frac{2 \times log\left(P(AC(C1, C2))\right)}{log\left(P(C1)\right) + log\left(P(C2)\right)} \qquad (4)$
Hirst. [28]	The idea of this measure is that two lexicalized concepts are semantically narrow if their set of synonymous (synsets) in WordNet are connected by a path that is not too long and that "do not change direction too often". With this measure, all the relations contained in a WordNet network are considered. In the work of [28], the authors have classified of the links towards the top link (superclass), down link (subclass) and horizontal link (antonym) [12]. The calculation of similarity with this method, is between objects (words) by the weight of the shortest path from one term to another, in addition to classifications which indicate the changes of direction [29]; [30].	$SimH = T - SPD - K \times nd \qquad (5)$ T and K are constants, SWD is the Shortest Path Distance in number of arc and nd the number of direction changes.

C. Hybrid techniques

These techniques are based on a model that combines between the approaches based on arcs (distances) in addition to the information contents that are considered as a decision factor.

Table 3. Hybrid techniques.

Similarity Measure	Description	Mathematical formula and commentaries
Jiang and Conrath. [14]	To remedy the problem presented at the Resnik measurement, authors has brought a new formula of combining the entropy (Informational Contents) the specific concept to those concepts which we seeks the similarity (combines between the techniques based on the arcs and those based on the nodes which consist in counting the arcs to improve results through the nodes based calculations). The measure adopting this method is based on the combination of a rich source of knowledge (thesaurus) with a source of poor knowledge (corpus) [30].	$SimJC(C1, C2) = \frac{1}{Distance(C1, C2)} \qquad (6)$ The distance between C1 and C2 is calculated by using the following formula: $Distance(C1, C2) = E(C1) + E(C2)$ $-\left(2 \times E\left(CS(C1, C2)\right)\right) \qquad (7)$
Leacock and Chodorow. [31]	The authors presented a method that combines between the counting method arcs and information content method. The measure proposed is based on the shortest path length between two WordNet synsets. The authors limited their attention to reporting relationships "is-a" and the path length through the overall depth P taxonomy.	$SimLC(C1, C2) = -Log\left(\frac{cd(C1, C2)}{2 \times M}\right) \qquad (8)$ M is the length of the longest path between the root concept of ontology and the lowermost concept. cd(C1, C2) is the length of the shortest path between C1 and C2.

D. Techniques based on the vector space

In the information retrieval domain, the vector space models are widely adopted [3]; [4]. These approaches use a characteristic vector, in a dimensional space, to represent each object and calculate the similarity based on the cosine measure or the Euclidean distance.

The vector space model is used for an arrangement of complex objects as the representatives of the K-dimensional vectors. The definition of the similarity between two vectors of objects is obtained by their internal contents. Here are some approaches mentioned in the literature (see the following table):

Table 4. Similarity measure based on the vector space.

Similarity measure	Description	Mathematical formula and commentaries
Jaccard	It's defined by the number of common objects divided by the total number of objects minus the number of common objects [30].	$SimJ(X,Y) = \frac{x \times y}{\|\|x\|\|_2^2 + \|\|y\|\|_2^2 - x \times y}$ (9) Such that x and y are vectors extracted from the concepts C1 and C2. $\|\|x\|\| = \varphi^{x}{}_{i=1}$ denotes the norm of vector x and $\|\|x\|\|_2 = \sqrt{\sum_{i=1}^{i=n} \|x_i\|^2}$
Cosine	It uses the complete vector representation, that is to say the objects frequency (words). Two objects (documents) are similar if their vectors are combined. If two objects are not similar, their vectors form an angle (X, Y) whose cosine is the value of similarity. This measure therefore quantifies the similarity between the two vectors such as the cosine of the angle between the two vectors [30].	The formula is defined by the ratio of the scalar product of the vectors x and y, and the product of the norm of x and y. $SimC(X,Y) = \cos(X,Y) = \frac{x \times y}{\|\|x\|\|_2 \times \|\|y\|\|_2}$ (10)
Euclidean	It's based on the ratio of the Euclidean Distance (ED) increased by 1 [30].	$ED = \|x - y\|$ $SimE(C1, C2) = \frac{1}{1 + ED}$ (11)
Dice	It's defined by the number of the common objects multiplied by 2 to the total number of objects [30].	$SimD(C1, C2) = \frac{2 \times x \times y}{\|\|x\|\|_2^2 + \|\|y\|\|_2^2}$ (12)

III. Related Works

A similarity measure between the concepts of a hierarchical ontology permits the validation of a domain data and the enrichment of these data in the same way by others of domain ontology.

Zargayouna and Salotti, in [29], define a function Spec (C1, C2) which calculates the specificity of two concepts in relation to the lowest of the ontology concept (Bottom, virtual concept which symbolizes the end of the ontology) as shown in figure 2.

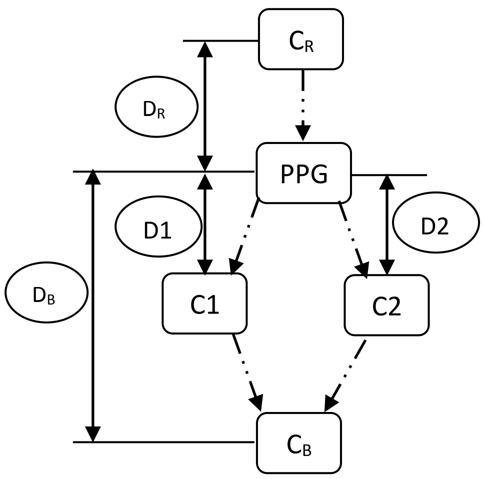


Fig. 2. Example of an extract of ontology.

$$Spec(C1, C2) = D_B \times D1 \times D2 \quad (13)$$

Spec (C1, C2) is equal to zero if C1 is ancestor of C2 or vice versa. Will therefore be penalized neighboring concepts C1 or C2, thus, the measure of Wu and Palmer becomes:

$$SimZS(C1, C2) = \frac{2 \times D_R}{D1 + D2 + Spec(C1, C2)} \quad (14)$$

$$SimZS(C1, C2) = \frac{2 \times D_R}{D1 + D2 + D_B \times D1 \times D2} \quad (15)$$

The cares of this measure are:

1. The measure of Wu and Palmer was changed by changing the distance (2 * D_R) by (D_B * D1 * D2) in the denominator of the formula (Here, we are talking about a new formula instead of an extension of the Wu and Palmer measure).
2. The obligation to seek the lowest concept of ontology (C_B).
3. The obligation to introduce and calculate another distance D_B.
4. The measure has been tested on a single instance of a single ontology.

T. Slimani and al in [30], Give an extension of Wu and Palmer measure represented by the following formula:

$$SimTBK = \frac{2 \times N}{N1 + N2_B} \times PF(C1, C2) \quad (16)$$

Let PF (C1, C2) be the penalization factor of two concepts C1 and C2 placed in the neighborhood.

If C1 and C2 are ⊂ in the same way,

$$PF(C1, C2) = (1 - \lambda)(Min(N1, N2) - N) + \lambda(|N1 - N2| + 1)^{-1} \quad (17)$$

Let N1 and N2 be the distances which separate nodes C1 and C2 from the root node, and N the distance which separates the closest common ancestor of C1 and C2 from the root node. C1 and C2 are the concepts for which the similarity is computed. The coefficient λ is a Boolean value indicating 0 or 1, with 0 indicating two concepts in the same hierarchy and 1 indicating two concepts in neighborhood, respectively. Min (N1, N2) represent the minimum value between C1 and C2. The ratio PF (C1, C2) =1 if C1 is ancestor of C2 or the reverse.

Among the cares of this formula:

1. Mathematically, the formula is complicated more because of:
   a. We must, always, to calculate the *Min (N1, N2)*,
   b. We must, always, to calculate the absolute value |N1 - N2|,
   c. A logical variable besides in the formula (λ).
2. The new measure has been tested on a single instance of ontology (To validate the measure, it is necessary to test it with more of an example).

IV. Our Contribution

A. Exposition of th problem

The Wu & Palmer measurement is interesting but has a limit because it primarily aims to detect the similarity between two concepts in relation to their distance of their SG. More this subsuming is general less they are similar (and vice versa). However, it does not capture the same similarities that symbolic conceptual similarity. Thus we can have Sim(A, f) < Sim(A, B), where f is one of the sons of A, and B one of the brothers of A. That which is inadequate for our senses in the information retrieval framework where it is necessary to bring back all the son of a concept (i.e. query) before its neighborhood. This measure has the advantage of the execution time speed, but the disadvantage of the production of a similarity value of two nearby concepts that exceed the value of two concepts in the same hierarchy.

As reference, figure 3 represents a graph showing a part of a hierarchy of gene ontology concepts in biology. The concepts contained in this ontology intuitively represent a set of various conceptual distances are compared two by two.

For example, the concept "cellular process" and "cellular component organization" present a similarity value equal to 0 in the case of the use of traditional similarity traditional measures that include external information in the hierarchy such as measures of [25]; [32].

By against, adopting an approach based on the hierarchy gives a similarity value different from 0 for these two same concepts. In addition, the similarity value of the two concepts "cellular process" and "cellular component organization" is less low than that of the concepts "cellular process" and "cell cycle".

However, we judge that the concept "cellular process" is closer to the concept "cell cycle" than the concept "cellular component organization".

These precise details are very interesting to research the semantic similarity of a set of concepts in ontology.

These intuitive distances can be used, for example, to improve search engines in terms of efficiency and accuracy of responses to customer requests. The simplest structure supporting the reasoning on type hierarchy is that which we can be found in a conceptual graphs support. In this structure, the subsumption links grouped types according to the definitional characteristics which they share.

B. Solution Suggested

For example, we can obtain with the Wu and Palmer measurement, a similarity value between the concept "M phase of mitotic cell cycle" and "mitosis" which exceeds the similarity value between "cellular process" and "M phase of mitotic cell cycle". However, this measure provides a higher similarity between a concept and its neighborhood compared to the same concept and son concept (see example of application below).

Let the ontology of Figure 3, denoted by C1, C2 and C3 the concepts "cellular process", "M phase of mitotic cell cycle" and "mitosis". By applying the Wu and Palmer measure, the similarity value is calculated as follows:

$$SimWP(C1, C2) = \frac{2 * 1}{0 + 4 + 2 * 1} = 0.33$$

$$SimWP(C2, C3) = \frac{2 * 3}{2 + 1 + 2 * 3} = 0.66$$

The values obtained by the Wu and Palmer measure show that the neighboring concepts C2 and C3 are more similar than the concepts C1 and C2 located in the same hierarchy what is inadequate within the framework of the semantic information retrieval.

We propose a new measure which updates the Wu and Palmer measure, whose expression is represented by the following formula:

$$SimDB(C1, C2) = \frac{2 \times D}{D1 + D2 + 2 \times D + FPD\_SG(C1, C2)} \quad (18)$$

With:

$$FPD\_SG(C1, C2) = \begin{cases} 0 & \text{If } Cond1 \\ (D+D1)*(D+D2) & \text{If } Cond2 \end{cases}$$

***Cond1* → C1 is ancestor of C2 or conversely.**

***Cond2* → C1 and C2 are close by a CS.**

FPD_SG (Function Produces Depths by Smaller Generalizing) is a function which makes it possible to penalize the similarity of two neighboring concepts which are not located in the same hierarchy. In the case of neighboring concepts, FPD_SG gives the distance in number of arcs equal to the product of depths of the two concepts compared to the ontology root via a CS. More and more that the distances D or Di (where D is the distance between CS and the root and Di represent the distance between a concept Ci and it CS) are distant, more and more SimDB decreases. With this function, the similarity measure between two hierarchical concepts is higher than the similarity between two neighboring concepts by a CS.

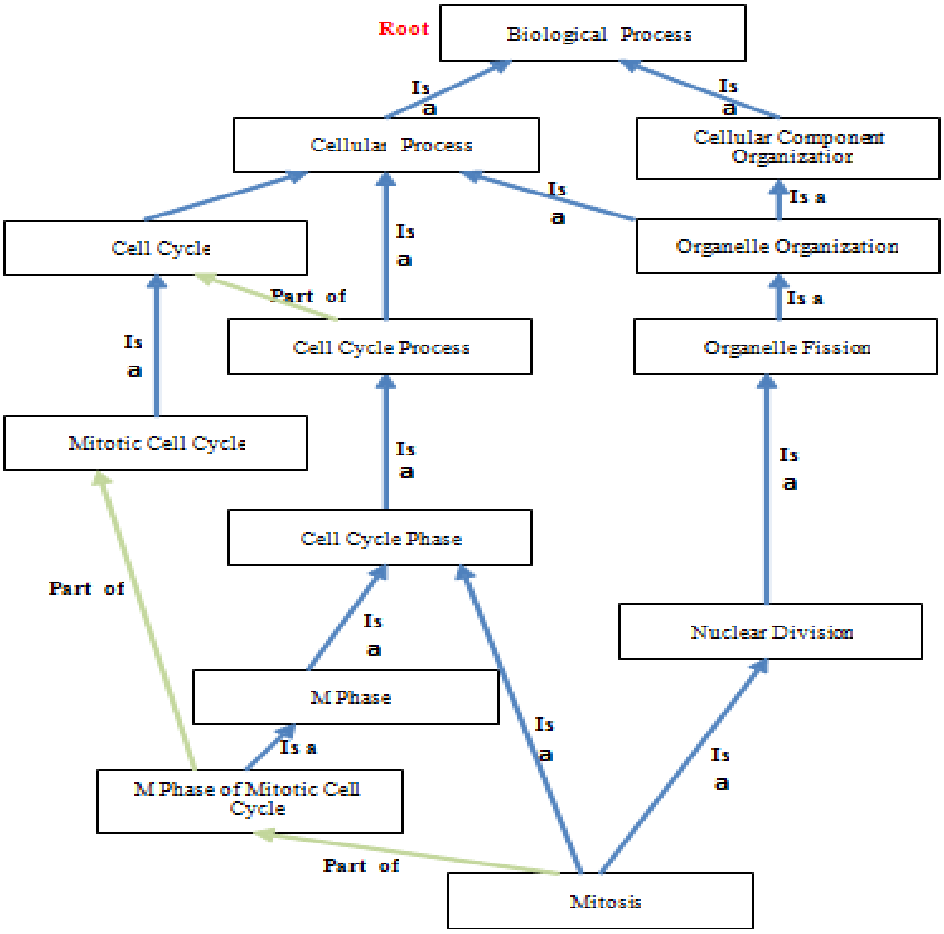


Fig. 3. Graph representing a hierarchical part of Gene Ontology in Biology 17].

C. Example of application

By taking the same previous example with the same concepts C1, C2 and C3, and by applying our measure and the Wu and Palmer measure, the similarity values between C1 and C2 and between C2 and C3 are indicated in the tables below:

Table 5. Similarity values calculated by Wu & Palmer measure, and our measure (Hierarchical concepts).

C1	C2	**SimWP**	**SimDB**
cellular process	M phase of mitotic cell cycle	**0.33**	**0.33**

Table 6. Similarity values calculated by Wu & Palmer measure, and our measure (neighboring concepts).

C2	C3	**SimWP**	**SimDB**
M phase of mitotic cell cycle	Mitosis	**0.66**	**0.20**

$$SimDB(C1, C2) = \frac{2*1}{0+4+2*1+0} = 0.33$$

$$SimDB(C2, C3) = \frac{2*3}{2+1+2*3+((3+2)*(3+1))} = 0.20$$

D. Properties of our measure

Let three concepts C1, C2 and C3 of any one ontology. Here are some properties satisfieded by our measure.

- Non-negativity: SimDB (C1, C2) $\geq$ 0.
- Identity: SimDB (C1, C1) = SimDB (C2, C2) = SimDB (C3, C3) = 1.
- Symmetry: SimDB(C1, C2) = SimDB(C2, C1).
- Uniqueness: SimDB (C1, C2) = 1 ➔ C1 = C2.
- Different: SimDB (C1, C2) = 0 ➔ C1 $\neq$ C2.
- Interval of definition: SimDB (C1, C2) $\subset$ [0..1].

E. Comparison between our measure and that of Wu and Palmer

The objective of this paper is to implement and test an update of a method of similarity measure can advance research in the field of ontology and simulation conceptual distances.

In an OWL ontology, each object is described by some RDF [30] reports. Let O be an object in OWL ontology. O is characterized by a set of descriptions which contains all the reports described. A set of description for O is defined by: Descr (O) = {(s, p, o) ϵ O}, where s, p and o are an RDF triple indicating the subject, the predicate and the object. RDF (Resource Description Framework) is now used as a standard for the exchange of the metadata between various applications. It facilitates the work of search engines to seek the documents in an efficient manner.

To verify the validity of our measure, it is judicious to test its relevance calculation compared to the Wu and Palmer measure which was considered to be fastest in terms of the similarity generation time [6]. The impact of the change in the Wu and Palmer measure and the result to our measure must be evaluated to judge its relevance.

In tables 7 and 8, we chose a representation per concepts pairs contained in ontology in order to calculate the similarities values. The calculation is performed respectively by the Wu and Palmer measure and by our measure.

Table 7 relates to hierarchical concepts while table 8 examines nearby concepts.

Table 7. Representation by pair of hierarchical concepts.

Concepts		SimWP	SimDB
C1	C2		
cellular process	M phase of mitotic cell cycle	**0.33**	**0.33**
cellular process	cell cycle	**0.66**	**0.66**
cell cycle phase	M phase	**0.54**	**0.54**
cellular process	cell cycle process	**0.66**	**0.66**
cell cycle phase	Mitosis	**0.54**	**0.54**

Table 8. Representation by pair of neighboring concepts.

Concepts		SimWP	SimDB
C2	C3		
M phase of mitotic cell cycle	Mitosis	**0.66**	**0.20**
cell cycle	cell cycle process	**0.50**	**0.25**
M phase	Mitosis	**0.75**	**0.25**
cell cycle process	organelle organization	**0.50**	**0.25**
Mitosis	M phase of mitotic cell cycle	**0.66**	**0.20**

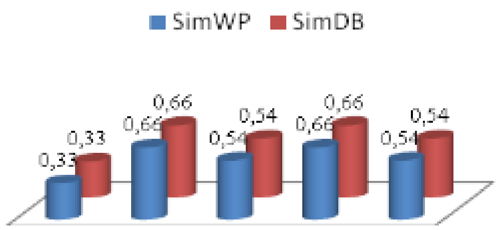


Fig. 4. Wu & Palmer measure unchanged by our measure (Hierarchical concepts).

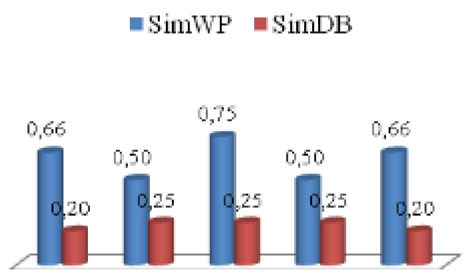


Fig. 5. Our measure relevance (Neighboring concepts).

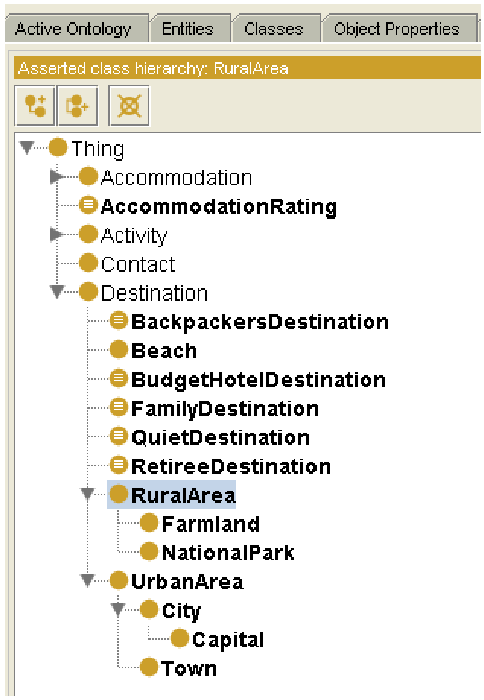


Fig. 6. Graph representing a hierarchical part of Travel Ontology [18].

Table 9. Representation by pair of hierarchical concepts.

Concepts		SimWP	SimDB
C1	C2		
Destination	Capital	**0.40**	**0.40**
Destination	NationalPark	**0.50**	**0.50**
UrbanArea	Capital	**0.33**	**0.33**

Table 10. Representation by pair of neighboring concepts.

Concepts		SimWP	SimDB
C2	C3		
Capital	Town	**0.57**	**0.21**
NationalPark	Farmland	**0.66**	**0.26**
Capital	Town	**0.40**	**0.21**

F. Another example (2nd Test)

The relevance of our measure compared to the Wu and Palmer measure is localized on the level of two concepts located in a hierarchy from which the subsuming concept is different.

Increasing the distance between the direct subsuming concepts is far more than the similarity value decreases. Comparing the relevance of the values found in Table 7 is shown in Figure 5. The results obtained show that there is an increase in the relevance provided by our measure.

Remark: Table 7, figure 4 and table 9 show that our measure has not changed the Wu and Palmer measure in the case of hierarchical concepts.

V. Conclusions and Future Work

In this work, we presented an update calculation of similarity presented by Wu and Palmer. We compared our measure with that of Wu and Palmer considered the fastest. The results obtained show that the measure produced ensures the relevance of the values produced for the similarity of two concepts.

The importance of this measure increases, in addition, in the case of hierarchical ontology that presents “is-a” relationships which allows a clearer precision for relationships. This can be adopted in the domain of semantic association identification where the current approaches relate to associations not giving a precision on the degree of association accuracy.

References

[1] V. Jain and M. Singh, “Ontology based information retrieval in semantic web: A survey”, International Journal of Information Technology and Computer Science (IJITCS), pp. 62-69, 2013.
[2] G. Bisson, “La similarité: Une notion symbolique / numérique”, journal apprentissage symbolique-numérique (Tome 2), pp. 169-201, 2000.
[3] R. Baeza-Yates and B. Ribeiro-Neto, “Modern Information Retrieval”, ACM Press and Addison-Wesley, New York USA, 1999.
[4] G. Salton and M. J. McGill, “Introduction to modern information retrieval”, Book, McGraw-Hill Inc, New York USA, 1986.
[5] Z. Wu and M. Palmer, “Verb semantics and lexical selection”, in proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics (ACL’94), pp. 133-138, Las Cruces, New Mexico, 1994.
[6] D. Lin, “An Information-Theoretic Definition of similarity”, in proceedings of the 15th International Conference on Machine Learning (ICML'98), pp. 296-304, Morgan-Kaufmann, San Francisco USA, 1998.
[7] H. Jeffrey, L. William and D. John, “A Semantic Similarity Measure for Semantic Web Services”, in proceedings of WWW 2005, Chiba Japan, 2005.
[8] P. Siddharth, S. Banerjee and T. Pedersen, “Using measures of semantic relatedness for word sense disambiguation”, in proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, Mexico, pp. 241-257, 2003.
[9] D. McCarthy, R. Koeling, J. Weeds and J. Carroll, “Finding predominant senses in untagged text”, in proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona Spain, pp. 280 – 287, 2004.
[10] I. Gurevych and M. Strube, “Semantic similarity applied to spoken dialogue summarization”, in proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, pp. 764-770, 2004.
[11] G. Hirst and A. Budanitsky, “Correcting real-word spelling errors by restoring lexical cohesion”, Natural Language Engineering Journal, vol. 11 Issue 1, pp. 87-111, 2004.
[12] M. Naeem, S. Gillani, M. Abdul Qadir and S. Asghar, “gSemSim: semantic similarity measure for intra gene ontology terms”, International Journal of Information Technology and Computer Science (IJITCS), pp. 32-40, 2013.
[13] P.W. Lord, R.D. Stevens, A. Brass and C. A. Goble, “Semantic Similarity Measures as Tools for Exploring the Gene Ontology”, Pacific Symposium on Biocomputing 8, Lihue Hawaii USA, pp. 601-612, 2003.
[14] J. J. Jiang and D. W. Conrath, “Semantic similarity based on corpus statistics and lexical taxonomy”, in Proceedings of International Conference Research on Computational Linguistics (ROCLING X), Taiwan, 1997.
[15] T. Pedersen, S. V. S. Pakhomov, S. Patwardhan and C. G. Chute, “Measures of semantic similarity and relatedness in the biomedical domain”, Journal of Biomedical Informatics, vol. 40, pp. 288–299, 2007.
[16] C. Pesquita, D. Faria, AO. Falcão, P. Lord and F. M. Couto, “Semantic similarity in biomedical ontologies”, PLOS Computational Biology Journal vol. 5 issue 7, 2009.
[17] Http:// www.geneontology.org
[18] Http://www.protege.stanford.edu/plugins/owl/owl-library/travel.owl
[19] R. Rada, H. Mili, E. Bicknell and M. Blettner, “Development and application of a metric on semantic nets”, IEEE Transaction on Systems, Man and Cybernetics, vol. 5 issue 1, pp. 17-30, 1989.
[20] J. H. Lee, M.H. Kim and Y. J. Lee, “Information Retrieval Based on Conceptual Distance in IS-A Hierarchies”, Journal of Documentation, vol. 49 issue 2, pp. 188-207, 1993.
[21] M. Halkidi, B. Nguyen, I. Varlamis and M. Vazirgiannis, thesus: ‘’Organizing web document collections based on link semantics”, Journal on Very Large Databases, Special Edition on the Semantic Web, Nov. 2003.

[22] E. Desmontils and C. Jacquin, ‘’Des ontologies pour indexer un site Web’’, actes des journées francophones d’Ingénierie des Connaissance, Nantes France, 2001.
[23] Http://www.nlm.nih.gov/mesh/
[24] M. Ehrig, P. Haase, M. Hefke and N. Stojanovic, “Similarity for ontology-a comprehensive framework”, in Workshop on Ontology and Enterprise Modelling: Ingredients for Interoperability In Conjunction with 5th International Conference on Practical Aspects of Knowledge Management, Vienna Austria, 2004.
[25] T. Eiter and H. Mannila, “Distance measures for point sets and their computation”, Acta Informatica Journal, vol. 34, pp. 109-133, Springer-Verlag, 1997.
[26] 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, 1999.
[27] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K. Miller, “Introduction to WordNet: An On-line Lexical Database”, Cognitive Science Laboratory, Princeton University, Princeton USA, Technical Report, 1993.
[28] G. Hirst and D. St-Onge, “Lexical chains as representations of context for the detection and correction of malapropisms”, Christiane Fellbaum editor, Cambridge, MA:The MIT Press, 1998.
[29] H. Zargayouna and S. Salotti, “SemIndex: A model of semantic indexing on XML documents”, in 26th European Conference on Information Retrieval (ECIR'2004), VOL. 2, 2004.
[30] T. Slimani, B. Ben Yaghlane and K. Mellouli, “A New Similarity Measure based on Edge Counting”, in World Academy of Science, Engineering and Technology, vol. 23, pp. 773-777, 2008.
[31] C. Leacock and M. Chodorow, “Combining Local Context and WordNet Similarity for Word Sense Identification, in WordNet: An Electronic Lexical Database”, MIT Press, 1998.
[32] J. Green, N. Horne, E. Orlowska and P. Siemens, “A Rough Set Model of Information Retrieval”, Fundamenta Infomaticae Journal, vol. 28, pp. 273-296, 1996.

Authors’ Profiles

Abdeslem DENNAI is a PhD student in sixth year Computer Science, University of Sidi Bel Abbes, Algeria. In 1994, he received the diploma of engineering in Computer Science from the University of Sidi Bel Abbes, Algeria. In 2008, he received the diploma of teaching in Computer Science from the University of Bechar, Algeria. He is a lecturer at the University of Bechar, Algeria. His research interests are in the field of semantic web, web applications and ontology.

Sidi Mohamed Benslimane is an Associate Professor at the Computer Science Department of Sidi Bel Abbes University, Algeria. He received his PhD degree in computer science from Sidi Bel Abbes University in 2007. He also received a M.S. and a technical engineer degree in computer science in 2001 and 1994 respectively from the Computer Science Department of Sidi Bel Abbes University, Algeria. He is currently Head of Research Team 'Service Oriented Computing' at the Evolutionary Engineering and Distributed Information Systems Laboratory (EEDIS). His research interests include, semantic web, service oriented computing, ontology engineering, information and knowledge management, distributed and heterogeneous information systems and context-aware computing.

How to cite this paper: Abdeslem DENNAI, Sidi Mohammed BENSLIMANE,"A New Measure of the Calculation of Semantic Distance between Ontology Concepts", International Journal of Information Technology and Computer Science(IJITCS), vol.7, no.7, pp.48-56, 2015. DOI: 10.5815/ijitcs.2015.07.06