IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier XXXX

An Enriched Information-Theoretic Definition of Semantic Similarity in a Taxonomy

ANNA FORMICA, FRANCESCO TAGLINO
Istituto di Analisi dei Sistemi ed Informatica (IASI) "Antonio Ruberti", National Research Council, Via dei Taurini 19, 00185 Rome, Italy
(e-mail: {anna.formica, francesco.taglino}@iasi.cnr.it)

Corresponding author: Anna Formica (e-mail: anna.formica@iasi.cnr.it)

ABSTRACT This paper addresses the notion of semantic similarity between concepts organized according to a taxonomy, based on the well-known *information content* approach. This approach has been widely experimented in the literature over the years and, in general, outperforms other proposals which do not originate from it. However, it shows some limitations related to the notion of *generic sense* of a concept. In this paper we illustrate the problem arising by using the traditional approach, and a novel information-theoretic definition of semantic similarity in a taxonomy is proposed which also takes into account the *intended sense* of a concept in a given context. This proposal has been applied to some among the most representative state-of-the-art similarity measures based on the information content approach, and the experiment shows that it achieves very high correlation values with human judgment.

INDEX TERMS Semantic Similarity, Information Content, Taxonomy, Context, Concept Sense.

I. INTRODUCTION

The information-theoretic definitions of semantic similarity defined by Resnik in [37], [38] and by Lin in [32], more than two decades ago, have been extensively mentioned and investigated in the literature, and a significant amount of similarity measures have been proposed originating from them, by relying on the *information content* approach [11], [12]. It is based on a probabilistic model that can be applied not only to concepts organized according to an ISA taxonomy (taxonomy for short), but also to ordinal values, feature vectors, and words. In particular, with regard to concepts organized according to a taxonomy, which is the focus of this paper, the information content approach was proposed in order to overcome the limitations of alternative methods for evaluating concept similarity, such as the *edge-counting* approach [36]. The key idea on which it relies is the following: the more information two concepts share the more similar they are, and concept similarity is directly proportional to the maximum information content shared by the concepts. The similarity measures based on the information content approach have been widely investigated in the literature over the years and, in general, have shown a higher correlation with human judgment with respect to other proposals which do not originate from it [3], [8], as experimented also by the authors in [14].

The starting assumption of this paper is that semantic similarity has to be computed by taking into account not only the information contents of the concepts but also the *context*[1], because different contexts can lead to different similarity degrees among the same concepts. It could be argued that a taxonomy, in order to work properly for a given purpose in a given application domain, should reflect a specific point of view, also referred to as *perspective* in [32]. For instance, consider a taxonomy about animals. If the taxonomy distinguishes pets from wild animals, cats will result more similar to dogs than to lions, but if the taxonomy describes families of animals, such as felines and canids, cats will result more similar to lions than to dogs. However, building domain specific taxonomies is not a simple task, because it requires a specific background knowledge and a significant amount of effort from domain experts. Therefore, in many cases, it is preferable to adopt general purpose and widely accepted taxonomies (e.g., WordNet [31]), which do not rely on specific perspectives.

As shown in this paper, context (or perspective) is fundamental in evaluating semantic similarity, and its role is more

[1]In this paper the word "context" is used to indicate the application domain, which determines the meanings of the related concepts.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

evident if we focus for instance on *siblings*, i.e., concepts of the taxonomy with the same parent, which share the same information content. Note that also the approach in [32] is based on the notion of perspective, but it does not allow to evaluate similarity by addressing a single perspective at a time, and the proposed information-theoretic definition of similarity between concepts is interpreted as "a weighted average of their similarities computed from different perspectives". For this reason, in this work, we refer to the mentioned information content notion as the similarity between the concept *generic senses*, i.e., the senses of the concepts that are not related to any specific context.

In this paper, we show how the similarity based on concept generic senses is not adequate to capture the meanings of the concepts related to specific contexts, here referred to as their *intended senses*. The problem will be shown by using an example involving sibling concepts in a simple taxonomy. Therefore, we propose an enrichment of the information-theoretic definition of semantic similarity in a taxonomy, that takes also into account the concept intended senses, i.e., concept meanings according to the given application domain.

It is important to observe that, to our knowledge, the approach presented in this paper is novel and does not allow a comparison with existing proposals due to the inherently different assumptions on which it relies. However, in order to validate it, additional hypotheses have been made on the addressed benchmark dataset, and an experimentation involving some of the most significant information content similarity measures has been performed, based on the well-known Miller&Charles dataset [34]. Note that this proposal can be applied to existing semantic similarity measures, and the experiment shows that it achieves high correlation values with human judgment in line with the literature.

The paper is organized as follows. In Section II, the problem is informally introduced by using an example, and the rationale behind the proposed approach is illustrated. Successively, in Section III, the enriched similarity measure is formally presented. In Section IV, the experiment is presented, that includes the disambiguation of the intended concept senses and, furthermore, the evaluation of their relevance that is illustrated in Subsection IV-A. The related work follows in Section V, and Section VI concludes.

II. SEMANTIC SIMILARITY IN A TAXONOMY

In this section, the topic addressed in this paper is informally presented by using a running example.

A. THE INFORMATION CONTENT APPROACH

According to Resnik [37], [38], the notion of semantic similarity between concepts organized according to a taxonomy relies on concept frequencies in text corpora, e.g., huge collections of text samples of American English. As mentioned above, the basic assumption of the approach is the following: the more information two concepts share the more similar they are, and the similarity between concepts is given by the maximum information content shared by them, which is represented by the information content of their *most informative subsumer* (i.e., the most specific concept in the taxonomy that is more general than both of them). The root of the taxonomy is the concept whose information content is null by definition, since it represents the most abstract concept.

For the sake of simplicity, in this section we address an example involving siblings, i.e., concepts that in the taxonomy are direct descendants of the same node, that is their parent. Figure 1 shows a fragment of a taxonomy where the concept $person$ is the parent of the three concepts $student$, $employee$, and $planter$ (children).

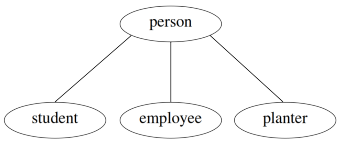

FIGURE 1. A simple taxonomy

The similarity between siblings is given by the information content associated with their parent, which is the maximum shared between them. For this reason siblings, in pairs, all have the same semantic similarity degrees. Therefore, in the example, the maximum information content shared by the pairs $(employee, student)$ and $(employee, planter)$ is the one associated with their parent, $person$, and the following holds:

$$sim(employee, student) = sim(employee, planter)$$

where sim stands for the similarity degree of the pair. Of course, this value also coincides with the one of the pair $(student, planter)$.

As a result, according to Resnik, siblings are indistinguishable from a similarity point of view, and the approach does not allow to capture further semantic aspects of the concepts, in order to have different pairs of siblings with different similarity degrees.

In the approach proposed by Lin [32], the notion of semantic similarity proposed by Resnik has been refined by also addressing the information contents of the compared concepts and, therefore, the related concept frequencies (or probabilities). Let us consider again the pairs of concepts $(employee, student)$ and $(employee, planter)$. Assume that the frequency of the concept $student$ in a text corpus is greater than the one of the concept $planter$ (but the opposite hypothesis can be taken as well). According to this assumption, the similarity degree between the concepts $employee$ and $student$ is greater than the one between $employee$ and $planter$ (see Section III where the similarity measure of Lin is formally recalled), i.e.:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

$$sim(employee, student) > sim(employee, planter).$$

Therefore, following this approach, given a set of sibling concepts in a taxonomy, one of them, in this case $employee$, is more similar to the "most frequent" sibling in a given corpus, i.e., $student$ in the example. With respect to the previous approach, pairs of siblings do not have the same similarity degrees, however similarity is evaluated by considering only concept frequencies and, in particular, the more frequent two siblings are the more similar they are. Indeed, as mentioned in the Introduction, this approach relies on the concept *generic senses*, i.e., meanings that are not related to any specific context. (Note that the above argumentation also holds in the case of evolved information content models, which are not based on concept frequencies in large-scale corpora, as discussed in the Related Work Section.)

In the next section, we propose a refinement of the information-theoretic definition of semantic similarity given in [32], by considering an additional element: the meanings of the concepts in a given application domain.

B. ENRICHING CONCEPTS WITH THE INTENDED SENSES

In our proposal, given a taxonomy of concepts and an application domain, we aim at "enriching" these concepts with other concepts of the same or another taxonomy, if there are any, that represent their meanings in that domain, as informally illustrated below.

Consider again the taxonomy of Figure 1, and suppose we have an application domain for which an important requirement for people is to spend several hours per day in a building. According to this perspective, we expect $employee$ to be more similar to $student$ rather than to $planter$, because an $employee$ and a $student$ are both characterized by the mentioned requirement better than the concepts $employee$ and $planter$. Therefore, we expect that the following holds:

$$sim(employee, student) > sim(employee, planter).$$

This is not the case if we consider another perspective, or application domain, where for instance it is more important to focus on people's income. Of course, in this second case, we expect that $employee$ will be more similar to $planter$ rather than to $student$, since the first two concepts share some form of $payment$. Therefore, in this second case, it is reasonable to expect the following:

$$sim(employee, student) < sim(employee, planter).$$

For these reasons, we propose to compute semantic similarity by also addressing the meanings that concepts have in the given domain, i.e., their *intended senses* in that domain. For instance, consider in Figure 2 an extension of the fragment of the taxonomy shown in Figure 1, where the concept $building$ has $office$ and $college$ as children, and $payment$ is the parent of $reward$ and $salary$. Now, in line with the first

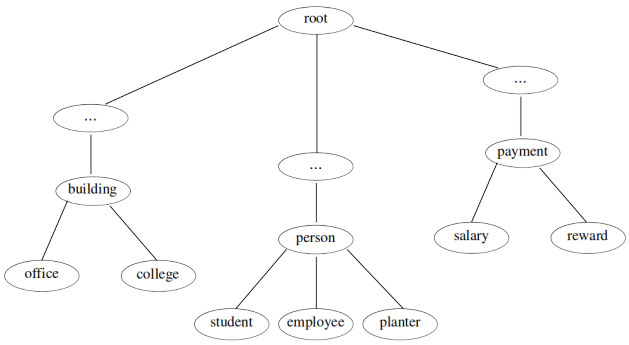

FIGURE 2. A simple taxonomy including concept senses

perspective illustrated above, suppose we have an application domain, say D_1, where it is important to characterize people on the basis of the time they spend in an edifice per day. Let $\mathcal{S}_{D_1}$ be the function associating the concepts of the taxonomy with their intended senses in the domain D_1, defined as follows:

$$\mathcal{S}_{D_1}(employee) = office$$
$$\mathcal{S}_{D_1}(student) = college$$
$$\mathcal{S}_{D_1}(planter) = reward$$

In the proposed approach, concept similarity is evaluated by addressing not only the maximum information content shared by the compared concepts, but also the one shared by their intended senses. Therefore, consider again the two pairs of siblings of our example. The intended senses of the concepts $employee$ and $student$ are $office$ and $college$, respectively, which have $building$, their parent, as maximum shared information content (see Figure 2). Whereas, with regard to $employee$ and $planter$, the most specific concept in the taxonomy that is more general than their meanings $office$ and $reward$ is the root, whose information content is null by definition. For this reason, for the related similarity degrees, we expect the following:

$$sim(employee, student) > sim(employee, planter).$$

In order to address the second scenario, where earnings are more relevant than workplaces, consider another application domain, say D_2, for which the intended sense of $employee$ is defined by the function $\mathcal{S}_{D_2}$ as follows:

$$\mathcal{S}_{D_2}(employee) = salary$$

while keeping the same definition for the concepts $student$ and $planter$, i.e.:

$$\mathcal{S}_{D_2}(student) = college$$
$$\mathcal{S}_{D_2}(planter) = reward.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

In this second perspective, since *salary* and *reward* share *payment* as concept with maximum information content, whereas *salary* and *college* share only the root, as shown in Figure 2, we expect that:

$$sim(employee, student) < sim(employee, planter).$$

In the next section the similarity measure proposed in this paper is introduced in formal terms. It allows an enrichment of the traditional information-theoretic definition of semantic similarity, with the intended senses of the concepts according to a given application domain.

III. THE ENRICHED SEMANTIC SIMILARITY MEASURE

The information content approach was proposed in [37] as an alternative to the edge-counting method [36], whose drawback is the assumption that links in a taxonomy represent uniform distances. Indeed, the former is based on a probabilistic model that is not sensitive to the problem of link distances. Below, it is briefly recalled in formal terms.

Consider a set of concepts C of an ISA taxonomy (taxonomy for short), and a function p:

$$p\colon C \rightarrow [0,1]$$

such that, for any $c \in C$, $p(c)$ is the *probability* of the concept c computed on the basis of the relative concept *frequency*, $freq(c)$, evaluated from large collections of multidisciplinary texts, such as the Brown Corpus of American English [17]. In particular, the probability of a concept c is defined as:

$$p(c) = freq(c)/N$$

where N is the total number of concepts in the corpus.

According to [39], the information content of a concept c, indicated as $IC(c)$, is computed as:

$$IC(c) = -log\ p(c)$$

which means that, intuitively, as the probability increases the informativeness decreases and, therefore, the more abstract a concept the lower its information content. Given two concepts $c_i, c_j \in C$, the notion of semantic similarity proposed by Resnik, $sim_R(c_i, c_j)$, relies on the assumption that the more information two concepts share, the more similar they are, and is defined as follows:

$$sim_R(c_i, c_j) = \max_{c \in S(c_i, c_j)} [-log\ p(c)]$$

where $S(c_i, c_j)$ is the set of concepts that *subsume* (are more general of) both c_i, c_j. The concept corresponding to the maximum value above is referred to as the *least common subsumer* (*lcs*) (the most informative subsumer in [37]) of the concepts c_i, c_j. Therefore:

$$sim_R(c_i, c_j) = -log\ p(lcs(c_i, c_j))$$

and therefore:

$$sim_R(c_i, c_j) = IC(lcs(c_i, c_j))$$

Successively, in [32] this notion was refined and, in particular, given two concepts $c_i, c_j \in C$, the concept semantic similarity proposed by Lin, $sim_L(c_i, c_j)$, is defined as follows:

$$sim_L(c_i, c_j) = \frac{2 \times IC(lcs(c_i, c_j))}{IC(c_i) + IC(c_j)} = \frac{2 \times sim_R(c_i, c_j)}{IC(c_i) + IC(c_j)}$$

where, with respect to the approach proposed by Resnik, the information contents of the compared concepts are both considered as an essential contribution in the evaluation of their semantic similarity.

However, both the Resnik's and Lin's approaches, as well as the similarity methods originating from them that will be addressed in the experiment of Section IV, do not consider the semantic similarity of the meanings of concepts according to a given context. In this paper, an enrichment of the information content based methods is proposed by allowing to further characterize the meanings of the compared concepts with respect to a given application domain. Below it is illustrated by using the Lin's formula, but it can be applied to any information content measure, as shown below.

Suppose we have an application domain, say D_k, the semantic similarity of the concepts $c_i, c_j \in C$, indicated as $sim_{D_k}(c_i, c_j)$, is defined as follows:

$$sim_{D_k}(c_i, c_j) = \frac{2 \times IC(lcs(c_i, c_j))}{IC(c_i) + IC(c_j)} * (1 - \omega_k) + \frac{2 \times IC(lcs(\mathcal{S}_{D_k}(c_i), \mathcal{S}_{D_k}(c_j)))}{IC(\mathcal{S}_{D_k}(c_i)) + IC(\mathcal{S}_{D_k}(c_j))} * \omega_k$$

where ω_k is a weight, $0 \leq \omega_k \leq 1$, defined by the domain expert according to D_k, and $\mathcal{S}_{D_k}$ is a function from C to C, referred to as the *intended sense* function, associating a concept with its meaning according to D_k, i.e.:

$$\mathcal{S}_{D_k}\colon C \rightarrow C$$

and:

$$\mathcal{S}_{D_k}(c) = \begin{cases} s & \text{if } s \in C \text{ is the } intended\ sense \text{ of } c \text{ in } D_k \\ c & \text{otherwise} \end{cases}$$

The above formula can be rewritten and generalized by using any information content based semantic similarity measure, $sim(c_i, c_j)$, as follows:

$$sim_{D_k}(c_i, c_j) = sim(c_i, c_j) * (1 - \omega_k) + sim(\mathcal{S}_{D_k}(c_i), \mathcal{S}_{D_k}(c_j))) * \omega_k \quad (1)$$

where the weight ω_k, depending on D_k, allows a balance between the roles of the generic senses and the intended

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

TABLE 1. Correlation with *HJ* of the selected seven methods by using the Miller&Charles ($M\&C$) dataset.

$concept_1$, $concept_2$	*HJ*	sim_R	$sim_{W\&P}$	sim_L	$sim_{J\&C}$	$sim_{P\&S}$	sim_A	$sim_{A\&M}$
car, automobile	3.92	11.630	1.00	1.00	30.000	1.000	0.826	0.815
gem, jewel	3.84	15.634	1.00	1.00	30.000	1.000	0.824	0.806
journey, voyage	3.84	11.806	0.91	0.89	27.497	0.800	0.766	0.753
boy, lad	3.76	7.003	0.90	0.85	25.839	0.823	0.673	0.633
coast, shore	3.70	9.375	0.90	0.93	28.702	0.928	0.645	0.587
asylum, madhouse	3.61	13.517	0.93	0.97	28.138	0.978	0.864	0.849
magician, wizard	3.50	8.744	1.00	1.00	30.000	0.950	0.721	0.680
midday, noon	3.42	11.773	1.00	1.00	30.000	1.000	0.877	0.863
furnace, stove	3.11	2.246	0.41	0.18	17.792	0.368	0.231	0.207
food, fruit	3.08	1.703	0.33	0.24	23.775	0.468	0.153	0.103
bird, cock	3.05	8.202	0.91	0.83	26.303	0.618	0.598	0.587
bird, crane	2.97	8.202	0.78	0.67	24.452	0.618	0.589	0.578
tool, implement	2.95	6.136	0.90	0.80	29.311	0.790	0.561	0.535
brother, monk	2.82	1.722	0.50	0.16	19.969	0.515	0.744	0.714
crane, implement	1.68	3.263	0.63	0.39	19.579	0.488	0.339	0.316
lad,brother	1.66	1.722	0.55	0.20	20.326	0.293	0.201	0.170
journey, car	1.16	0.000	0.00	0.00	17.649	0.228	0.000	0.000
monk, oracle	1.10	1.722	0.41	0.14	18.611	0.310	0.195	0.163
food, rooster	0.89	0.538	0.70	0.04	17.657	0.310	0.063	0.000
coast, hill	0.87	6.329	0.63	0.58	25.461	0.650	0.390	0.325
forest, graveyard	0.84	0.00	0.00	0.00	14.520	0.173	0.111	0.078
monk, slave	0.55	1.722	0.55	0.18	20.887	0.343	0.208	0.176
coast, forest	0.42	1.703	0.33	0.16	15.538	0.200	0.117	0.083
lad, wizard	0.42	1.722	0.55	0.20	20.717	0.310	0.203	0.172
chord, smile	0.13	2.947	0.41	0.20	17.535	0.173	0.184	0.122
glass, magician	0.11	0.538	0.11	0.06	17.098	0.283	0.153	0.131
noon, string	0.08	0.000	0.00	0.00	12.987	0.195	0.061	0.000
rooster, voyage	0.08	0.000	0.00	0.00	12.506	0.048	0.000	0.000
Correl. with *HJ*	1.00	0.795	0.777	0.834	0.836	0.874	0.857	0.858

senses of the concepts, according to the relevance they have in the domain D_k.

Note that, given an application domain D_k, in this proposal both the weight ω_k and the function $\mathcal{S}_{D_k}$ are defined according to domain expert judgments. However, as mentioned in Section VI, in future work we are planning to extend this approach to the framework of Linked Data [9], in order to support the domain expert not only in the evaluation of ω_k, but also in the selection of the intended senses of concepts according to the addressed domain.

IV. EXPERIMENTAL RESULTS

As mentioned above, this paper focuses on semantic similarity of concepts organized according to an ISA taxonomy. For this reason, as also discussed in the Related Work Section, in this experiment all the methods for computing the more general notion of semantic *relatedness*, i.e., concerning non-taxonomic relations [28], have not been addressed. The same also holds for the similarity methods relying on, for instance, Wikipedia [22], since the automatic extraction of the ISA taxonomy from it requires additional ad-hoc algorithms and, therefore, a further level of correlation to be analyzed, that necessarily impacts on the overall evaluation of the methods and goes beyond the scope of this work [6].

It is important to remark that this is a novel approach for which, to our knowledge, there are no comparable proposals in the literature. Therefore, in order to validate it, in the experiment below additional assumptions are required on the addressed benchmark datasets. In fact, with respect to the traditional experimentations where a dataset composed of a set of pairs of concepts suffices, for this proposal we need further pairs of concepts, i.e. concept senses, representing contexts.

In order to arrange the experiment, we focused on the well-known Miller&Charles ($M\&C$) dataset [34], which is still considered a reference for comparing semantic similarity methods [3] and, for each pair of concepts of this dataset, we considered all the pairs of concepts of the same dataset as possible contexts. With regard to the state-of-the-art, we selected six information content based approaches, which are among the most significant methods for evaluating semantic similarity in a taxonomy, that are recalled in the Related Work Section. In particular, besides the Resnik (sim_R), and Lin (sim_L) milestones, we applied our proposal to the measures of Jiang and Conrath ($sim_{J\&C}$) [26], Pirrò and Seco ($sim_{P\&S}$) [35], Adhikari et al. (sim_A) [3] and, finally, the measure proposed by Adhikari et al. with the information content model computed as Meng ($sim_{A\&M}$) [2]. In addition, also the Wu and Palmer method ($sim_{W\&P}$) has been addressed [46], as representative of the edge-counting approach [36], that can be seen as a special case of [32].

Analogously to [32], let us consider 28 pairs of concepts of the $M\&C$ dataset, and the correlations with human judgment (*HJ*) of the seven methods above, that are shown in Table

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

TABLE 2. The 28 contexts for the pair of concepts $(journey, voyage)$

context	*sense*$_1$, *sense*$_2$	*HJ*	*R*	*W&P*	*L*	*J&C*	*P&S*	*A*	*A&M*	r_1	r_2	ω_k
D_1	car, automobile	3.88	11.72	0.96	0.95	28.75	0.90	0.80	0.78	0.50	0.50	0.50
D_2	gem, jewel	3.84	12.39	0.92	0.91	27.88	0.83	0.77	0.76	0.15	0.15	0.15
D_3	journey, voyage	3.84	11.81	0.91	0.89	27.50	0.80	0.77	0.75	1.00	1.00	1.00
D_4	boy, lad	3.83	11.07	0.91	0.88	27.24	0.80	0.75	0.73	0.15	0.15	0.15
D_5	coast, shore	3.81	11.27	0.91	0.90	27.76	0.83	0.74	0.72	0.44	0.00	0.22
D_6	asylum, madhouse	3.82	11.94	0.91	0.90	27.55	0.81	0.77	0.76	0.00	0.15	0.08
D_7	magician, wizard	3.72	10.73	0.94	0.93	28.38	0.85	0.75	0.73	0.35	0.35	0.35
D_8	midday, noon	3.78	11.80	0.92	0.91	27.88	0.83	0.78	0.77	0.15	0.15	0.15
D_9	furnace, stove	3.67	9.58	0.79	0.72	25.24	0.70	0.64	0.63	0.31	0.15	0.23
D_{10}	food, fruit	3.54	7.78	0.68	0.63	26.01	0.67	0.52	0.49	0.48	0.32	0.40
D_{11}	bird, cock	3.55	10.50	0.91	0.87	27.06	0.73	0.71	0.69	0.30	0.42	0.36
D_{12}	bird, crane	3.64	10.98	0.88	0.84	26.80	0.76	0.73	0.71	0.30	0.15	0.23
D_{13}	tool, implement	3.51	9.68	0.91	0.86	28.18	0.80	0.69	0.67	0.37	0.37	0.37
D_{14}	brother, monk	3.50	8.45	0.77	0.65	24.99	0.71	0.76	0.74	0.35	0.32	0.33
D_{15}	crane, implement	3.04	8.62	0.81	0.70	24.55	0.68	0.61	0.59	0.37	0.37	0.37
D_{16}	lad, brother	3.30	9.29	0.82	0.72	25.71	0.67	0.62	0.61	0.15	0.35	0.25
D_{17}	journey, car	1.83	2.94	0.23	0.22	20.10	0.37	0.19	0.19	1.00	0.50	0.75
D_{18}	monk, oracle	3.19	9.43	0.79	0.71	25.40	0.68	0.63	0.61	0.32	0.15	0.24
D_{19}	food, rooster	2.51	6.74	0.82	0.51	23.07	0.58	0.45	0.41	0.48	0.42	0.45
D_{20}	coast, hill	3.19	10.61	0.85	0.82	27.05	0.77	0.68	0.66	0.44	0.00	0.22
D_{21}	forest, graveyard	2.87	7.98	0.62	0.60	23.29	0.60	0.55	0.53	0.34	0.3	0.32
D_{22}	monk, slave	2.77	8.51	0.79	0.66	25.34	0.65	0.58	0.56	0.32	0.33	0.33
D_{23}	coast, forest	2.50	7.86	0.68	0.60	22.82	0.57	0.51	0.49	0.44	0.34	0.39
D_{24}	lad, wizard	2.98	9.27	0.82	0.72	25.79	0.68	0.62	0.61	0.15	0.35	0.25
D_{25}	chord, smile	3.44	10.85	0.86	0.82	26.42	0.73	0.70	0.69	0.00	0.22	0.11
D_{26}	glass, magician	2.58	7.99	0.64	0.61	23.98	0.63	0.56	0.54	0.33	0.35	0.34
D_{27}	noon, string	3.55	10.9	0.84	0.82	26.39	0.75	0.71	0.70	0.15	0.00	0.08
D_{28}	rooster, voyage	1.17	3.42	0.26	0.26	16.85	0.27	0.22	0.22	0.42	1.00	0.71
Correl. with *HJ*		1.00	0.90	0.81	0.91	0.90	0.92	0.92	0.92			

1. As mentioned above, the same dataset was addressed in order to associate each pair of the dataset with 28 possible application domains D_k, k = 1...28, in the following referred to as contexts (therefore we evaluated 28*28 = 784 similarity scores). For instance, for the pair of concepts $(journey, voyage)$, the 28 contexts are shown in Table 2, and are:

$\mathcal{S}_{D_1}(journey) = car$
$\mathcal{S}_{D_1}(voyage) = automobile$
$\mathcal{S}_{D_2}(journey) = gem$
$\mathcal{S}_{D_2}(voyage) = gewel$
...
$\mathcal{S}_{D_{28}}(journey) = rooster$
$\mathcal{S}_{D_{28}}(voyage) = voyage.$

We have seen in Section III that, in general, the intended senses of concepts are supposed to be estimated by domain experts, together with the related weight ω_k in the given context D_k, according to Formula (1). In this experiment, in order to quantify such a weight, which represents the relevance of a pair of senses with respect to the pair of contrasted concepts, we leveraged the existing methods for evaluating the semantic relatedness of concepts. In fact, for this purpose, we do not have to restrict our attention to concept similarity, but we also need to consider non-taxonomic relations, e.g., thematic relations [28]. Therefore, in the available literature, the method proposed in [41] has been selected because it exploits the large amounts of semantic relations encoded within DBpedia semantic network[2]. Furthermore, it achieves competitive performances in computing the semantic distances of concepts by relying on the information content approach. In particular, given a pair of concepts c_i,c_j and a context D_k, we assumed ω_k as the average of the semantic relatedness of a concept of that pair with the corresponding concept of the associated pair of senses $(\mathcal{S}_{D_k}(c_i), \mathcal{S}_{D_k}(c_j))$, i.e.:

$\omega_k = (r_1 + r_2)/2$

where $r_1 = rel(c_i, \mathcal{S}_{D_k}(c_i))$ and $r_2 = rel(c_j, \mathcal{S}_{D_k}(c_j))$, and rel is the relatedness degree computed according to [41]. For instance, for the pair of concepts $(journey, voyage)$ consider the pair of senses $(food, fruit)$, corresponding to the context D_{10} in Table 2, i.e.:

$\mathcal{S}_{D_{10}}(journey) = food$
$\mathcal{S}_{D_{10}}(voyage) = fruit.$

In this case we have:

$rel(journey, food) = 0.48$, $rel(voyage, fruit) = 0.32$, and therefore:

$\omega_{10} = (0.48 + 0.32)/2 = 0.40.$

The similarity values of Table 2 have been obtained by applying Formula (1) in Section III to the selected state-of-the-art measures, as well as to the human judgment, as shown by the following example. For instance, in the case

[2]We recall that DBpedia (http://dbpedia.org) is the result of the ongoing project of representing Wikipedia content in RDF (Resource Description Framework) [45], in order to make it compliant with the Linked Data principles [9].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

of the context D_{10}, the measure proposed by Adhikari et al., $sim_{A,D_{10}}$, has been computed according to the values given in Table 1, as follows[3]:

$$sim_{A,D_{10}}(journey, voyage) = sim_A(journey, voyage) * 0.60 + sim_A(food, fruit) * 0.40 = 0.52$$

In order to compute 28 tables[4], one for each pair of the $M\&C$ dataset, each table containing 28 possible contexts for that pair, a disambiguation step has to be performed. In fact, it is well-known that in Wikipedia, and consequently in DBpedia, terms are addressed with the possible meanings they have, i.e., a term is associated with multiple senses. For this reason, in this experiment the disambiguation is necessary in order to address senses in line with the *HJ* evaluation in the $M\&C$ experiment. For instance, $crane$ in Wikipedia has two main senses, that are $bird$ and $machine$. Table 3 shows the results concerning the average weights in the 28 contexts, ω_{avg}, of $crane$ before and after the disambiguation steps. In particular, when paired with $implement$ and $bird$, it is disambiguated by using the senses $machine$ and $bird$, respectively. Note that in the case of the pair $(crane, implement)$, the average weight significantly increases (from 0.08 to 0.32) if $crane$ stands for a machine, and $implement$ stands for a tool. Analogously, for the pair $(bird, crane)$, the average weight increases after disambiguating it with the bird sense.

In the next subsection, a data analysis concerning the senses of the concepts with respect to the concepts to be compared is performed.

TABLE 3. Disambiguation of $crane$

$sense_1$, $sense_2$	ω_{avg}
crane, implement	0.08
crane_(machine), tool	0.32
bird, crane	0.20
bird, crane_(bird)	0.36

A. RELEVANCE OF THE INTENDED CONCEPT SENSES

In the experiment, in associating a given pair of concepts with a pair of possible concept senses, in some cases the weigh ω_k, for a given context D_k, is null. Within these cases, there are some particular situations for which both the concept senses do not have any relevance with the concepts to be compared, i.e., both the values r_1, r_2 above are null. In other words, for some pairs of concepts, there are contexts (or perspectives) that do not apply to both the compared concepts, i.e., they do not correspond to any specific point of view and, for this reason, in the experimentation these contexts have been ignored. This is for instance the case of the pair of concepts $(coast, shore)$, when associated with the pairs of senses $(brother, monk)$, or (boy, lad).

The same also holds in the case of concept senses with low similarity values, such as for instance the pair $(noon, string)$, or $(chord, smile)$. Therefore, in order to analyze significant contexts, a threshold for *HJ* in Table 1 has been introduced, in this case equal to 0.5 (in the scale from 0 to 4). It is important to observe that this threshold has been applied only in the case of concept *senses*, whereas the experiment concerns all the 28 pairs of concepts, including the ones with *HJ* less than 0.5. The correlations with *HJ* for the pair $(journey, voyage)$ in the addressed contexts is shown in Table 2, whereas the average correlations for all the 28 pairs are shown in Table 4. Furthermore, in Table 5, the correlations of some specific pairs of concepts are illustrated, with the related average weights, starting from a pair of very similar concepts, such as $(car, automobile)$, to the pair of concepts $(journey, car)$ which are related but not similar [28]. This issue is also discussed in the next subsection.

1) Reliability of concept senses

It is interesting to observe how the correlation behaves if we further restrict our attention to pairs of concept senses which correspond to "reliable" contexts, as illustrated below.

Consider again the pair of concepts $(journey, voyage)$. Among the 28 contexts, including for instance (boy, lad), or $(midday, noon)$ that have low relevance weights ω_k, let us focus on the five contexts shown in Table 6, where the similarity values for the methods L, $J\&C$, $P\&S$, A, and $A\&M$ are given. In the table, besides the correlation obtained according to the data analysis illustrated above (*Correl.*), also the one obtained by applying only the disambiguation step, indicated as $Correl_d$, is shown. Note that in the case of D_3, the context does not provide any additional information about the intended senses of the concepts $journey$ and $voyage$ and, therefore, their similarity coincides with the one of their generic senses. In fact ω_3 = 1.00 and, for all the methods, the similarity values in Table 6 are the same of Table 1. Consider now the context D_{20}, where the concept $journey$ stands for a trip up the $coast$, whereas $voyage$ is a travel through the $hill$. In this case similarity always decreases, since for all the methods the intended senses of $journey$ and $voyage$ have similarity values less than the ones computed by addressing their generic senses (see Table 1). This is more evident if we consider the context D_{23}, which associates with $journey$ the same meaning it has in D_{20}, i.e., $coast$, whereas $voyage$ stands for a trip in the $forest$. This is not the case of the context D_5, where the intended meaning of $voyage$ is a trip along the $shore$. In fact, the similarity of $journey$ and $voyage$ increases except for the methods A and $A\&M$, as expected according to the corresponding values of the related senses.

It is interesting to observe the case of the context D_{17}, where the intended senses for $(journey, voyage)$ are represented by the pair $(journey, car)$. For all the methods

[3]In Tables 2, 4, 5, and 6, R, $W\&P$, L, $J\&C$, $P\&S$, A, $A\&M$, stand respectively for the methods of Resnik, Wu and Palmer, Lin, Jiang and Conrath, Pirrò and Seco, Adhikari et al., and Adhikari et al. with information content computed as Meng, all evaluated according to Formula (1) in Section III.

[4]The data about the experiment are available at Taglino, F., Formica, A., (2021), "Miller_and_Charles_with_concepts_senses", Mendeley Data, http://dx.doi.org/10.17632/thtndvvp9s

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

TABLE 4. Average correlations in 28 contexts

$concept_1$, $concept_2$	R	$W\&P$	L	$J\&C$	$P\&S$	A	$A\&M$
car, automobile	0.84	0.77	0.86	0.85	0.87	0.87	0.87
gem, jewel	0.76	0.68	0.79	0.77	0.82	0.82	0.83
journey, voyage	0.90	0.81	0.91	0.90	0.92	0.92	0.92
boy, lad	0.85	0.75	0.89	0.85	0.92	0.89	0.89
coast, shore	0.82	0.79	0.87	0.85	0.81	0.88	0.88
asylum, madhouse	0.88	0.75	0.90	0.89	0.92	0.89	0.88
magician, wizard	0.80	0.69	0.85	0.87	0.89	0.86	0.85
midday, noon	0.86	0.71	0.88	0.88	0.88	0.87	0.87
furnace, stove	0.63	0.45	0.61	0.57	0.64	0.66	0.67
food, fruit	0.78	0.52	0.81	0.82	0.82	0.84	0.84
bird, cock	0.83	0.69	0.86	0.84	0.85	0.88	0.88
bird, crane	0.78	0.72	0.82	0.80	0.84	0.84	0.84
tool, implement	0.77	0.62	0.81	0.80	0.82	0.80	0.80
brother, monk	0.78	0.70	0.82	0.83	0.89	0.86	0.86
crane, implement	0.72	0.63	0.75	0.72	0.77	0.76	0.77
lad, brother	0.80	0.73	0.87	0.83	0.90	0.88	0.88
journey, car	0.89	0.84	0.90	0.88	0.90	0.91	0.91
monk, oracle	0.76	0.63	0.80	0.75	0.84	0.83	0.83
food, rooster	0.75	0.53	0.81	0.84	0.81	0.79	0.80
coast, hill	0.67	0.63	0.76	0.69	0.76	0.73	0.73
forest, graveyard	0.78	0.71	0.81	0.76	0.81	0.84	0.84
monk, slave	0.75	0.66	0.79	0.74	0.82	0.82	0.82
coast, forest	0.75	0.67	0.79	0.76	0.79	0.77	0.77
lad, wizard	0.77	0.72	0.85	0.79	0.88	0.86	0.85
chord, smile	0.74	0.71	0.85	0.80	0.89	0.82	0.81
glass, magician	0.87	0.85	0.92	0.92	0.90	0.89	0.88
noon, string	0.93	0.85	0.94	0.92	0.94	0.94	0.94
rooster, voyage	0.90	0.85	0.90	0.93	0.93	0.92	0.91
Avg Correl.	0.80	0.70	0.84	0.82	0.85	0.84	0.84

the similarity values of the contrasted concepts considerably decrease, although the weight w_5 is high (0.75). Indeed, this result is expected due to the semantically different kind of relation the concepts $journey$ and car have, since they are *related* concepts, linked by a thematic relation, that are not considered *similar* [28]. In fact, it is important to observe that, in Table 1, according to L, A, and $A\&M$, the similarity values of $journey$ and car are null, whereas this does not hold for the methods $J\&C$ and $P\&S$. Overall, in Table 6, with respect to $Correl_d$, the methods L, A, and $A\&M$ show correlation values slightly better than $J\&C$, and $P\&S$ (0.98 vs 0.97), with an increase of the average weights (0.55 vs 0.52). On the basis of these results, a further investigation about the impact of non-taxonomic relations on this proposal may be worthwhile.

The methods addressed in this experimentation are better illustrated in the Related Work Section below.

V. RELATED WORK

In the literature, there is a significant amount of works addressing semantic similarity [11], [12]. With the advent of Wikipedia, the most widely used and up-to-date knowledge repository, several approaches have been proposed by exploiting its features, such as articles, hyperlinks, categories, etc.. (see for instance [22], [25], [27], [31]). As mentioned above, semantic similarity is a special case of semantic relatedness [28]. The latter also concerns thematic relations (and, more in general, non-taxonomic relations) and, in the literature, there are several proposals investigating it, by relying on general purpose knowledge resources, such as Wikipedia or WordNet [5], [20], [21], [49]. However, all the approaches in the mentioned literature do not address the intended senses of a pair of concepts in a given application domain, or context, but they consider all the combinations of all possible senses for that pair, and then select the highest values. For this reason, as also mentioned in the previous section, in order to compare this proposal with the mentioned state-of-the-art, not only ad-hoc algorithms for the automatic extraction of the ISA taxonomy are needed, but also for the identification and the extraction of the concept senses in the given application domains. Therefore, in order to have reliable benchmark datasets, further parameters have to necessarily be investigated [6], whose analysis goes beyond the scope of this paper.

With regard to semantic relatedness, we remark that in this proposal it has been addressed in the experiment for the different purpose of evaluating ω_k, i.e., the relevance of the intended sense of a concept in a domain D_k. As mentioned above, among the existing proposals, the method defined in [41] has been selected since it shows competitive performances by relying on the information content approach.

Within the semantic similarity approaches, as for instance the one recently presented in [47] for Neural Networks, or hybrid similarity measures combining the shortest path lengths and the depths of subsumers [30], below we restrict our attention to the methods based on the information content

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3096598, IEEE Access

TABLE 5. Correlations of specific pairs of concepts in 28 contexts with the related average weights

$concept_1$, $concept_2$	R	$W\&P$	L	$J\&C$	$P\&S$	A	$A\&M$	ω_{avg}
car, automobile	0.84	0.77	0.86	0.85	**0.87**	**0.87**	**0.87**	0.41
coast, shore	0.82	0.79	0.87	0.85	0.81	**0.88**	**0.88**	0.31
coast, hill	0.67	0.63	**0.76**	0.69	**0.76**	0.73	0.73	0.33
journey, car	0.89	0.84	0.90	0.88	0.90	**0.91**	**0.91**	0.75

TABLE 6. Similarity of $(journey, voyage)$ in reliable contexts

context	$sense_1$, $sense_2$	*HJ*	L	$J\&C$	$P\&S$	A	$A\&M$	ω_k
D_3	journey, voyage	3.84	0.89	27.50	0.80	0.77	0.75	1.00
D_5	coast, shore	3.81	0.90	27.76	0.83	0.74	0.72	0.22
D_{17}	journey, car	1.83	0.22	20.10	0.37	0.19	0.19	0.75
D_{20}	coast, hill	3.19	0.82	27.05	0.77	0.68	0.66	0.22
D_{23}	coast, forest	2.50	0.60	22.82	0.57	0.51	0.49	0.39
ω_{avg}								0.55
Correl.		1.00	**0.98**	0.97	0.97	**0.98**	**0.98**	
ω_{avg}								0.52
$Correl_d$		1.00	0.96	**0.97**	**0.97**	0.96	**0.97**	

(IC) approach, which has been employed in different research areas, such as Natural Language Processing [4], Semantic Web [14], [33], Formal Concept Analysis [13], [43], IFCA (Formal Concept Analysis with Interval Type-2 Fuzzy sets) [15], Geographical Information Systems [16], [40], and different application domains, such as health [1], [24], and network security [44], just to mention a few examples. However the IC approach, although recognized as "the state of the art on semantic similarity" [3], [8], has shown some limitations, as discussed below. In the following, we start by recalling the IC based approaches addressed in the Experimental Results Section.

With regard to the works of Resnik [37] and Lin [32] (R, and L, respectively, in the tables above), which have been analyzed in details in Section III, we briefly summarize the following. According to the former, concept similarity in a taxonomy is computed by considering only concept commonalities (i.e., concept least common subsumers). Therefore, it shows some limitations since pairs of concepts having the same least common subsumers have the same similarity degrees. The latter, according to [10], can be reconducted to the well-known Tversky linear contrast model of similarity [42], which addresses both concept commonalities and differences. In particular, also in [32] the importance of observing an object from different perspectives is emphasized, but the proposed resulting similarity degrees are considered as weighted averages of the similarity values obtained from such perspectives. As a result, this approach does not allow to estimate concept similarity by considering a single specific perspective at a time. Successively, in the late 1990s, in [26] a proposal combining the IC with the edge-counting approach has been presented ($J\&C$ in our experiment), showing better performances with respect to the previous methods. However, one objection to the early IC based measures relies on the use of large-scale corpora [3], [7], [8], [23], [48]. In fact, evaluating the IC on the basis of statistical information taken from textual corpora requires a huge amount of manual effort at level of both design and maintenance of the corpus. For this reason, in the literature, an evolution of the IC notion has been extensively investigated, referred to as *intrinsic information content* (IIC), although there is a lack of a statistically significant difference between the performances of the IIC models and the corpus-based ones [29]. In particular, the IIC is evaluated independently of textual corpora, and in accordance with the intrinsic structure of the taxonomy, i.e., on the basis of the number of hyponyms and/or hypernyms of the concepts. Along this direction, Adhikari et Al. propose a method in [3] (A in our experiment), arguing that by relying only on the maximum among the ICs of the least common subsumers leads to ignore some common subsumers that can be relevant in order to evaluate semantic similarity. For this reason, in the mentioned paper, the IC is estimated according to an IIC approach by introducing a new notion, referred to as *Disjoint Common Subsumers*. A variant of this approach based on Meng model has also been proposed in [2], that shows slightly better performances with respect to the other measure ($A\&M$ in our experiment). Both the models they present achieve high correlation values when applied to the state-of-the-art measures addressed in our experiment. Finally, in [35] ($P\&S$ in our experiment), an IIC approach for semantic similarity has been proposed by relying on Tversky contrast model, that shows high correlation with human judgment with respect to the state-of-the-art. As illustrated above, this measure also shows high correlation values in our experiment, although the impact of non-taxonomic relations on our proposal should be better investigated. Besides the methods addressed in our experimentation, it is worth mentioning that in [8] the IIC notion is revised by using not only concept hyponyms and hypernyms, but also leveraging synonymy and polysemy in WordNet. In [7], [48], the authors claim that most of IC computing models have been developed for single-parent taxonomies, therefore they propose a new

IIC computing model in the presence of multiple inheritance hierarchies, such as in WordNet.

With respect to all the aforementioned works, in this paper we do not present a new IC computing model, and our proposal is independent of the IIC models recalled above. In fact, although these approaches show a high accuracy in the similarity evaluation, they do not involve concept meaning and, in particular, the related similarity measures do not address the intended senses of concepts according to a given application domain.

Note that the semantic similarity measure proposed in [23] originates from the need to overcome one of the limitations we highlighted in this paper, i.e., that pairs of concepts sharing the same least common subsumers have the same similarity degrees. However, the authors base their solution on the whole WordNet ontology, by associating the different kinds of relationships (e.g., ISA and PartOf) with different weights, which is again a proposal independent of the concept intended senses.

The notion of sense has been addressed by Resnik in [38], where semantic similarity is used to identify and select the appropriate sense of a concept when it appears in a group of related terms. Analogously, in [18] the semantic similarity of Lin and the MeSH thesaurus have been employed in order to determine the adequate sense of an ambiguous biomedical term. However, both these papers address word sense disambiguation in the field of computational linguistics, where semantic similarity is not the objective of the works but is used in order to associate a noun with the right sense in a given context. On the contrary, we use the concept intended senses to improve the computation of semantic similarity. Finally, senses are also addressed in [19], where concept similarity is computed between the most-related pairs among the concept corresponding meanings, but the intended senses of the compared concepts are not considered.

VI. CONCLUSION AND FUTURE WORK

In this work an enrichment of the information-theoretic definition of semantic similarity has been presented, for concepts organized according to a ISA taxonomy. The proposed measure is based on a novel approach that essentially addresses the context of the contrasted concepts, by associating them with their intended senses. In this way, concept similarity scores can be refined and made closer to the specificity of the given application domain, and the related purpose. This proposal has been applied to some among the most relevant state-of-the-art similarity measures, and the results show that it achieves high correlation values with human judgment in line with the results presented in the literature for the specific methods.

Regarding future work, this proposal can be placed within the more general framework of Linked Data [9]. The idea is to consider, for instance, the DBpedia knowledge graph and use the rdf triples as all possible senses of concepts. In particular a concept (node) of the rdf graph can be associated with as many senses as the number of rdf triples in which it appears as subject. Therefore, given an application domain, this graph could support the domain expert in selecting the intended sense of a concept according to the addressed domain.

Furthermore, besides an analysis about the impact of non-taxonomic relations on the proposed approach, we plan to refine this measure by defining the intended sense of a concept as a *set* of concepts, rather than a single one, in order to better characterize the concept meaning with respect to a given context. Furthermore, each concept belonging to such a set will be associated with a degree of accuracy representing how much it describes the related concept in the given domain. Therefore, a method for computing semantic similarity between sets of concepts will be investigated.

REFERENCES

[1] Abdelrahman, A.M.B., Kayed, A. A Survey on Semantic Similarity Measures between Concepts in Health Domain. American Journal of Computational Mathematics 5, 204-214 (2015).

[2] Adhikari, A., Singh, S., Mondal, D., Dutta, B., Dutta, A., A Novel Information Theoretic Framework for Finding Semantic Similarity in WordNet. CoRR, arXiv:1607.05422, abs/1607.05422 (2016).

[3] Adhikari, A., Dutta, B., Dutta, A., Mondal, D., Singh, S. An intrinsic information content-based semantic similarity measure considering the disjoint common subsumers of concepts of an ontology. J. Assoc. Inf. Sci. Technol. 69(8), 1023-1034 (2018).

[4] Ajumder, G.O.M., Akray, P.A.P, Elbukh, A.L.G. Measuring Semantic Textual Similarity of Sentences Using Modified Information Content and Lexical Taxonomy. Int. J. of Computational Linguistics and Applications 7(2), 65-85 (2016).

[5] AlMousa, M., Benlamri, R., Khoury, R. Exploiting non-taxonomic relations for measuring semantic similarity and relatedness in WordNet. Knowledge-Based Systems, 212, 106565 (2020).

[6] Aouicha, M.,B., Hadj Taieb, M., A., Ezzeddine, M. Derivation of "is a" taxonomy from Wikipedia Category Graph. Engineering Applications of Artificial Intelligence, 50, 265-286 (2016).

[7] Banu, A., Fatima, S.S., Khan, K. Information Content Based Semantic Similarity Measure for Concepts Subsumed By Multiple Concepts. Int. J. Web Appl. 7(3), 85-94 (2015).

[8] Batet, M., Sànchez, D. Leveraging synonymy and polysemy to improve semantic similarity assessments based on intrinsic information content. Artif. Intell. Rev. 53(3), 2023-2041 (2020).

[9] Bizer, C., Heath, T., Berners-Lee, T. Linked data - the story so far. Int. J. Semantic Web Inf. Syst. 5(3), 1-22 (2009).

[10] Cazzanti, L., Gupta, M.R. Information-theoretic and Set-theoretic Similarity. IEEE Int. Symp. on Information Theory, Seattle, WA, 1836-1840 (2006).

[11] Chandrasekaran, D., Mago, V. Evolution of Semantic Similarity - A Survey. ACM Computing Surveys, 54(2), Article 41 (2021).

[12] Elavarasi, S. A., Akilandeswari, J., Menaga K. A Survey on Semantic Similarity Measure. Int. J. of Research in Advent Technology 2(3) (2014).

[13] Formica, A. Concept Similarity in Formal Concept Analysis: an Information Content Approach. Knowledge-Based Systems 21(1), 80-87 (2008).

[14] Formica, A., Missikoff M., Pourabbas E., Taglino F. Semantic search for matching user requests with profiled enterprises. Computers in Industry 64(3), 191-202 (2013).

[15] Formica, A. Similarity reasoning in formal concept analysis: from one- to many-valued contexts. Knowledge and Information Systems 60(2), 715-739 (2019).

[16] Formica, A., Mazzei, M., Pourabbas E., Rafanelli, M. Approximate Query Answering Based on Topological Neighborhood and Semantic Similarity in OpenStreetMap, IEEE Access 8, 87011-87030 (2020).

[17] Francis, W.N., Kucera, H. Frequency Analysis of English Usage: Lexicon and Grammar. Houghton Mifflin, Boston (1982).

[18] Gabsi, I., Kammoun, H., Brahmi, S., Amous, I. MeSH-based disambiguation method using an intrinsic information content measure of semantic similarity. Procedia Computer Science 112, 564-573 (2017).

[19] Gao, J. B., Zhang, B. W., Chen, X. H. A WordNet-based semantic similarity measurement combining edge-counting and information content theory. Eng. Appl. Artif. Intell. 39, 80-88 (2015).

[20] Hadj Taieb, M.A., Aouicha, M.B., Hamadou, A.B. *Computing semantic relatedness using Wikipedia features*. Knowledge-Based Systems, 50, 260-278 (2013).

[21] Hadj Taieb, M.A., Zesch, T. Aouicha, M.B. *A survey of semantic relatedness evaluation datasets and procedures*. Artif. Intell. Rev. 53, 4407-4448 (2020).

[22] Hussain, M.J., Wasti, S.H., Huang, G., Wei, L., Jiang, Y., Tang, Y. An approach for measuring semantic similarity between Wikipedia concepts using multiple inheritances. Information Processing & Management, 57(3), 102188 (2020).

[23] Jeong, S., Yim, J.H., Lee, H.J., Sohn, M.M. Semantic Similarity Calculation Method using Information Contents-based Edge Weighting. J. Internet Serv. Inf. Secur. 7(1), 40-53 (2017).

[24] Jia, Z., Lu, X., Duan, H., Li., H. *Using the distance between sets of hierarchical taxonomic clinical concepts to measure patient similarity*. BMC Medical Informatics and Decision Making, 19(91), 1-11 (2019).

[25] Jiang, Y., Bai, W., Zhang, X., Hu, J. Wikipedia-based information content and semantic similarity computation. Inf. Process. Manag. 53(1), 248-265 (2017).

[26] Jiang, J.J., Conrath, D.W. *Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy*. In Proc. of Inter. Conf. Research on Computational Linguistics (ROCLING X), Taiwan (1997).

[27] Jiang, Y., Zhang, X., Tang, Y., Nie, R. *Feature-based approaches to semantic similarity assessment of concepts using Wikipedia*. Information Processing & Management, 51(3), 215-234 (2015).

[28] Kacmajor, M., Kelleher, J.D. Capturing and measuring thematic relatedness. Lang. Resources & Evaluation 54, 645-682 (2020).

[29] Lastra-Dìaz,.J.J., Garcìa-Serrano A. A new family of information content models with an experimental survey on WordNet. Knowledge-Based Systems 89, 509-526 (2015).

[30] Li, Y., Bandar, A.Z., McLean, D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans. on Knowledge and Data Engineering, 15(4), 871-882 (2003).

[31] Li, F., Liao, L., Zhang, L., Zhu, X., Zhang, B., Wang, Z. An Efficient Approach for Measuring Semantic Similarity Combining WordNet and Wikipedia. IEEE Access 8, 184318-184338 (2020).

[32] Lin, D. An Information-Theoretic Definition of Similarity. In Proceedings of the Int. Conf. on Machine Learning, Madison, Wisconsin, USA, Morgan Kaufmann, 296-304 (1998).

[33] Meymandpour, R., Davis, J.G. A semantic similarity measure for linked data: An information content-based approach. Knowledge-Based Systems 109, 276-293 (2016).

[34] Miller, G.A., Charles, W.G. Contextual correlates of semantic similarity. Language and Cognitive Processes 6(1), 1-28 (1991).

[35] Pirrò, G. *A Semantic Similarity Metric Combining Features and Intrinsic Information Content*. Data Knowl. Eng, 68(11), 1289-1308 (2009).

[36] Rada, R., Mili, H., Bichnell, E., Blettner, M. *Development and application of a metric on semantic nets*. IEEE Trans. Syst. Man. Cybern. 9, 17-30 (1989).

[37] Resnik, P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In Proc. of the Int. Joint Conf. on Artificial Intelligence, Montreal, Quebec, Canada, August 20-25, Morgan Kaufmann, 448-453 (1995).

[38] Resnik, P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. J. Artif. Intell. Res. 11, 95-130 (1999).

[39] Ross, S. *A First Course in Probability*. Macmillan (1976).

[40] Schwering, A. Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. Transactions in GIS, 12(1), 5-29 (2008).

[41] Schuhmacher, M., Ponzetto, S.P. Knowledge-based Graph Document Modeling, Proc. of the 7th ACM Int. Conf. on Web Search and Data Mining, (WSDM), New York, USA, 543-552 (2014).

[42] Tversky, A. *Features of similarity*. Psychological Review 84, 327-352 (1977).

[43] Wang, F., Wang, N., Cai S., and Zhang, W. *A Similarity Measure in Formal Concept Analysis Containing General Semantic Information and Domain Information*. IEEE Access 8, 75303-75312 (2020).

[44] Weller-Fahy, D.J., Borghetti, B.J., Sodemann, A.A. A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection. IEEE Communication Surveys & Tutorials 17(1),70-91 (2015).

[45] Wood, D., Lanthaler, M., Cyganiak, R. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014 (2014).

[46] Wu, Z., Palmer, M. *Verb semantics and lexical selection*.In Proc. of the 32nd Annual Meeting of the Associations for Computational Linguistics, Las Cruces, New Mexico, 133-138 (1994).

[47] Zhang, P., Huang, X., Wang, Y., Jiang, C., He, S., Wang, H. Semantic Similarity Computing Model Based on Multi Model Fine-Grained Nonlinear Fusion. IEEE Access 9, 8433-8443 (2021).

[48] Zhang, X., Sun, S., Zhang, K. An information Content-Based Approach for Measuring Concept Semantic Similarity in WordNet. Wirel. Pers. Commun. 103(1), 117-132 (2018).

[49] Zhu, X., Yang, X., Huang, Y., Guo, Q., Zhang, B. Measuring similarity and relatedness using multiple semantic relations in WordNet. Knowledge and Information Systems, 62, 1539-1569 (2020).

ANNA FORMICA received the full-honors degree in Mathematics from the University of Rome "La Sapienza" in 1989. Currently, she is a senior researcher at the "Istituto di Analisi dei Sistemi ed Informatica" (IASI) "Antonio Ruberti" of the Italian National Research Council (Consiglio Nazionale delle Ricerche - CNR) in Rome, where she manages the "Software and Knowledge-based Systems" (SaKS) group. She serves as referee of several international journals and conferences and she took part in various research projects of the European Framework Programs and bilateral projects with international institutions. Her current research interests are: Semantic Web, Similarity Reasoning, formal specification and validation of Domain Ontologies, Fuzzy Formal Concept Analysis, Geographical Information Systems, and e-Learning.

FRANCESCO TAGLINO was graduated in Information Science at the University of Rome "La Sapienza" in 1999. Since 2009, he is a permanent researcher at the "Istituto di Analisi dei Sistemi ed Informatica" (IASI) "Antonio Ruberti" of the Italian National Research Council (Consiglio Nazionale delle Ricerche - CNR) in Rome, as member of the "Software and Knowledge-based Systems" (SaKS) group. His main research interests are on knowledge representation and reasoning and semantic technologies, and in particular on ontology engineering, semantic similarity and relatedness, as well as quantum computing. He participated in several national and international projects, mainly in the context of enterprise interoperability, and serves as referee in several international journals and conferences.