Computing and Informatics, Vol. 42, 2023, 191–209, doi: 10.31577/cai_2023_1_191

SEMANTIC SIMILARITY IN A TAXONOMY BY REFINING THE RELATEDNESS OF CONCEPT INTENDED SENSES

Anna Formica, Francesco Taglino

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

Abstract. In this paper, we present an evolution of a novel approach for evaluating semantic similarity in a taxonomy, based on the well-known notion of information content. Such an approach takes into account not only the generic sense of a concept but also its intended sense in a given context. In this work semantic similarity is evaluated according to a refined relatedness measure between the generic sense and the intended sense of a concept, leading to higher correlation values with human judgment with respect to the original proposal.

Keywords: Semantic similarity, information content, taxonomy, semantic relatedness, concept sense

Mathematics Subject Classification 2010: 68T30

1 INTRODUCTION

The similarity measures based on the information content approach defined by Resnik [1] and Lin [2] have been extensively investigated in the literature and, in general, have shown higher correlation values with human judgment with respect to other proposals which do not originate from it [3, 4, 5, 6].

In [7] a novel approach has been presented that allows semantic similarity to be computed by taking into account not only the information contents of the concepts but also the *context*, i.e., the meanings of the concepts in the given application domain. As shown in [7], the context (or *perspective* [2]) is fundamental in evaluating semantic similarity, and different contexts can lead to different similarity degrees among the same concepts. The role of context is more evident if we focus on *siblings*, i.e., concepts of the taxonomy with the same parent, which share the same information content. Note that also the approach proposed by Lin is based on the notion of perspective, but it does not allow for evaluating similarity by addressing a single perspective at a time, and the 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, analogously to [7], we distinguish the notion of concept *generic sense*, i.e., the sense of the concept that is not related to any specific context, from the concept *intended sense*, i.e., the meaning of the concept in a specific context. In order to compute semantic similarity between concepts, the essential activity consists in evaluating the relatedness [8] between the generic sense and the intended sense of a given concept. With this regard, in this paper, we refine the original proposal of the authors by addressing the $ASRMP_m$ relatedness measure proposed in [9], which shows the best correlation with the human judgment concerning other methods defined in the literature [10]. The new experiment leads, for each method addressed in [7], to an average increment of the average correlation with the human judgment of about 0.04.

The paper is organized as follows. In Section 2 the problem is informally recalled, and in Section 3 the enriched similarity measure is given, with a subsection describing the $ASRMP_m$ measure. In Section 4 the new experiment is presented. The related work follows in Section 5, and Section 6 presents the conclusion.

2 SEMANTIC SIMILARITY IN A TAXONOMY

In this section, the informal presentation of the method given in [7] is recalled.

According to Resnik [1], the notion of semantic similarity between concepts organized according to a taxonomy relies on concept frequencies in text corpora, e.g., gross 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 where the 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).

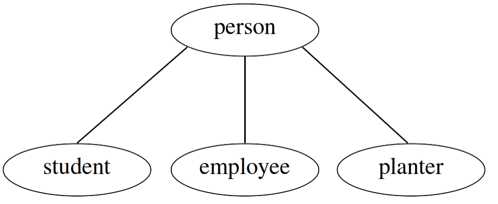

Figure 1. A simple taxonomy

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 for further semantic aspects of the concepts to be captured, in order to have different pairs of siblings with different similarity degrees.

In the approach proposed by Lin [2], 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 3 where the similarity measure of Lin is formally recalled in Equation (1)), i.e.:

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

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.

Assume now that siblings have the same frequencies, 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:

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

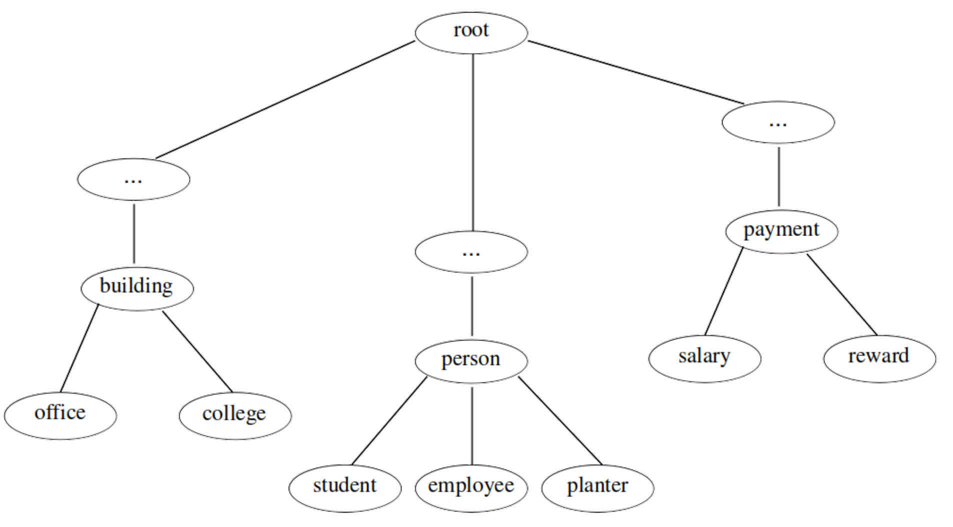

Figure 2. A simple taxonomy including concept senses

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 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:

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

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 in 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:

$$\mathit{sim}(\mathit{employee}, \mathit{student}) > \mathit{sim}(\mathit{employee}, \mathit{planter}).$$

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:

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

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

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

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:

$$\mathit{sim}(\mathit{employee}, \mathit{student}) < \mathit{sim}(\mathit{employee}, \mathit{planter}).$$

In the next section the similarity measure relying on this novel approach is formally recalled [7].

3 THE ENRICHED SEMANTIC SIMILARITY MEASURE

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

$$p : C \to [0, 1]$$

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. 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. Hence, 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 [1], $sim_R(c_i, c_j)$, 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 [1]) 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 [2] this notion has been 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:

$$sim_L(c_i, c_j) = \frac{2 \times IC(lcs(c_i, c_j))}{IC(c_i) + IC(c_j)} = \frac{2 \times sim_R(c_i, c_j)}{IC(c_i) + IC(c_j)}, \tag{1}$$

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 4, do not consider the semantic similarity of the meanings of concepts according to a given context. In this direction, in [7] an enrichment of the information content-based methods has been proposed, by characterizing the meanings of the compared concepts with respect to a given application domain.

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:

$$sim^{D_k}(c_i, c_j) = sim(c_i, c_j) \times (1 - \omega_k) + sim(\mathcal{S}_{D_k}(c_i), \mathcal{S}_{D_k}(c_j))) \times \omega_k, \qquad (2)$$

where *sim* is a similarity measure (e.g., sim_R or sim_L), 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.:

$$\mathcal{S}_{D_k} : C \to C$$

and

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

Note that the weight ω_k, depending on D_k, allows a balance between the roles of the generic senses and the intended senses of the concepts, according to the relevance they have in the domain D_k.

3.1 Evaluating the Relatedness of Concept Intended Senses

$ASRMP_m$ [9] is a family of semantic relatedness measures, originating from a previous proposal of the authors referred to as *Weighted Semantic Relatedness Measure* (*WSRM*). Such measures have been conceived in order to:

1. have a formal semantics,
2. have reasonable computational costs,
3. be transitive.

This family relies on the assumption that the more edges between nodes, the stronger their relatedness. Let us start by recalling the *WSRM* measure.

Consider a graph $\mathcal{G}$, and the set N of nodes of such a graph. Given two nodes n_i, $n_j \in N$, $WSRM(n_i, n_j)$ between n_i, and n_j is defined by the following Equation (3):

$$WSRM(n_i, n_j) = \frac{|\{p|(n_i, p, n_j) \in \mathcal{G}\}|}{\sum_{n' \in N} |\{p'|(n_i, p', n') \in \mathcal{G}\}|}, \qquad (3)$$

where for any set S, $|S|$ is the cardinality of S. According to the mentioned paper, the authors propose three different strategies in order to compute semantic relatedness between two nodes of the graph.

The first measure, referred to as $ASRMP^a_m$, considers all the paths between the compared nodes of length *equal* to m. In particular, given the nodes n_i, n_j, $ASRMP^a_m(n_i, n_j)$ is defined as shown in Equation (4):

$$ASRMP^a_m(n_i, n_j) = \oplus_{p \in n_i \rightsquigarrow n_j, |p|=m} \otimes^m_{k=1} WSRM(n_k, n_{k+1}), \qquad (4)$$

where:

- $n_i \rightsquigarrow n_j$ is the set of the directed paths between n_i and n_j with length equal to m,
- n_k is the k^{th} node of the path p (therefore $n_1 = n_i$, and $n_{m+1} = n_j$),
- $\otimes$ and $\oplus$ are the *t-norm* and the related *s-norm* aggregators, respectively, the former for the edges of a given path, and the latter for different paths of length m.

Among the different aggregators defined in the literature, the fuzzy logic operator *t-norm* selected by the authors for $\otimes$ (and the corresponding *s-norm* for $\oplus$) is the Hamacher operator, which ensures transitivity and is defined as follows:

$$T_{H,0}(x, y) = \frac{xy}{x + y - xy}. \tag{5}$$

The second measure proposed by the authors is $ASRMP^b_m$ that, with respect to the previous one, aggregates all the paths of length *less than or equal to* m, as defined in the following Equation (6).

$$ASRMP^b_m(n_i, n_j) = \oplus_{p \in n_i \rightsquigarrow n_j, |p| \leq m} \otimes_{k=1}^{|p|} WSRM(n_k, n_{k+1}), \tag{6}$$

where $n_i \rightsquigarrow n_j$ is the set of the directed paths between n_i and n_j of length less than or equal to m. However, the authors state that direct links should represent stronger relations, whereas indirect ones should account for weaker relations and, therefore, the longer the path, the weaker the relation. For this reason, they propose the following third measure where paths are weighted on the basis of their length l, $l = 1, \ldots, m$, as shown in Equation (7):

$$ASRMP^c_m(n_i, n_j) = \sum_{l=1}^{m} \sum_{p \in n_i \rightsquigarrow n_j, |p| = l} z_l \otimes_{k=1}^{|p|} WSRM(n_k, n_{k+1}), \tag{7}$$

where the directed paths between n_i and n_j with length equal to l are addressed, and z_l is a length-dependent weight, approximately corresponding to the percentage of paths of length l.

In order to achieve symmetry, the three strategies above are reformulated according to the $\psi^x_m(n_i, n_j)$ relatedness measure, $x \in \{a, b, c\}$, defined in the following Equation (8):

$$\psi^x_m(n_i, n_j) = \frac{1}{2}(ASRMP^x_m(n_i, n_j) + ASRMP^x_m(n_j, n_i)). \tag{8}$$

Among these measures the authors state that, according to their experiments, the $ASRMP^a_m$ provides the best performances, and this is the strategy we have adopted in order to evaluate the weight ω_k of Equation (2), in the experiment presented in the next section.

4 EXPERIMENTAL RESULTS

As mentioned in [7], the measure addressed in this paper relies on a novel approach for which the experimentation requires, besides the dataset composed of a set of pairs of concepts, further pairs of concepts representing the concept senses. Therefore, in order to compare the new experimental results with the ones of the original proposal, the Miller & Charles (M & C) dataset [11] has been addressed and, for each pair of concepts of this dataset, all the pairs of concepts of the same dataset have been considered as possible contexts. Furthermore, the same six information content-based approaches discussed in [7] have been analyzed and, in particular, besides Resnik (sim_R) and Lin (sim_L), also Jiang and Conrath ($sim_{J\,\&\,C}$) [12], Pirrò ($sim_{P\,\&\,S}$) [13], Adhikari et al. (sim_A) [3], and the measure proposed by Adhikari et al. with the information content model computed as Meng ($sim_{A\,\&\,M}$) [14]. The Wu and Palmer method ($sim_{W\,\&\,P}$) [15] has also been addressed, as representative of the edge-counting approach [16].

Consider the 28 pairs of concepts of the M & C dataset, and the same dataset in order to associate each pair with 28 possible application domains D_k, $k = 1, \ldots, 28$, in the following referred to as contexts (therefore we have $28 \times 28 = 784$ similarity scores). For instance, for the pair of concepts (*coast*, *shore*), the 28 contexts are:

$$\begin{aligned}
\mathcal{S}_{D_1}(\textit{coast}) &= \textit{car},\\
\mathcal{S}_{D_1}(\textit{shore}) &= \textit{automobile},\\
\mathcal{S}_{D_2}(\textit{coast}) &= \textit{gem},\\
\mathcal{S}_{D_2}(\textit{shore}) &= \textit{gewel},\\
&\vdots\\
\mathcal{S}_{D_{28}}(\textit{coast}) &= \textit{rooster},\\
\mathcal{S}_{D_{28}}(\textit{shore}) &= \textit{voyage}.
\end{aligned}$$

As mentioned in Section 3, 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 (see Equation (2)). In the experiment presented in [7], 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 used the method proposed in [17]. Indeed, in [10] an extensive experiment has been presented in order to compare the methods for evaluating concept relatedness in knowledge graphs, showing that the $ASRMP_m$ approach [9] (recalled in Subsection 3.1) provides the best correlation with human judgment with respect to others methods, including [17]. For this reason, in this experiment, given a pair of concepts c_i, c_j and a context D_k, we assume that ω_k is defined as follows:

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

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 [9].

It is important to recall that in order to compute the 28 tables, 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 the 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*, therefore when paired for instance with *implement*, it is disambiguated by using the sense *machine*.

Furthermore, in the experiment, in associating a given pair of concepts with a pair of possible concept senses, in some cases the weight ω_k, for a given context D_k, is null. In addition, 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 (*chord*, *smile*), or (*noon*, *string*). Therefore, in order to analyze significant contexts, a threshold for *HJ* has been introduced, in this case equal to 0.5 (on a scale from 0 to 4).

In Table 1, for reader's convenience, the average correlations for all the 28 pairs according to the experimental results presented in [7] are given, where the relatedness degrees have been computed by leveraging the semantic relatedness measure presented in [17]. In Table 2 the corresponding values obtained by relying on the semantic relatedness approach proposed in [9] are shown, with an average increment for each method of the average correlation with the human judgment of about 0.04. These relatedness measures are based on different weighting methods, however, it is worth mentioning that one of the key differences between them is the following: according to [17], semantic relatedness is computed by considering all the *undirected paths* connecting the compared entities, whereas according to [9], it is evaluated by addressing all the *directed paths* linking the compared resources (see Equation (4). The experimental results of this work show that the employment of the $ASRMP_m$ strategy has a direct impact on the increment of the semantic similarity correlation values, and the combination of the measure defined in Equation (2) with the approach proposed in [9] provides the best strategy in order to evaluate seman-

$concept_1$, $concept_2$	sim_R	$sim_{W\,\&\,P}$	sim_L	$sim_{J\,\&\,C}$	$sim_{P\,\&\,S}$	sim_A	$sim_{A\,\&\,M}$
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

Table 1. Average Pearson's correlations in the 28 contexts according to the experiment presented in [7]

tic similarity in a taxonomy by addressing the concept intended senses in a given context.

The data concerning the new experiment are available at [18], where also the Spearman's correlations are provided. In Table 3 the average correlations for all 28 pairs according to Spearman are also shown, which do not differ significantly from the ones obtained according to the original experiment.

$concept_1$, $concept_2$	sim_R	$sim_{W\,\&\,P}$	sim_L	$sim_{J\,\&\,C}$	$sim_{P\,\&\,S}$	sim_A	$sim_{A\,\&\,M}$
car, automobile	0.98	0.97	0.98	0.98	0.98	0.98	0.98
gem, jewel	0.99	1.00	1.00	1.00	1.00	1.00	1.00
journey, voyage	0.98	0.98	0.98	0.99	0.99	0.98	0.98
boy, lad	0.95	0.94	0.96	0.89	0.96	0.96	0.96
coast, shore	0.92	0.92	0.94	0.88	0.92	0.95	0.95
asylum, madhouse	0.95	0.86	0.95	0.99	0.98	0.94	0.93
magician, wizard	0.98	0.95	0.99	0.98	0.99	0.99	0.99
midday, noon	0.99	0.99	0.99	1.00	0.99	0.99	0.99
furnace, stove	0.54	0.56	0.47	0.29	0.44	0.48	0.47
food, fruit	0.43	−0.12	0.59	0.93	0.83	0.46	0.45
bird, cock	0.91	0.68	0.92	0.94	0.94	0.92	0.92
bird, crane	0.69	0.59	0.75	0.90	0.87	0.71	0.70
tool, implement	0.92	0.90	0.94	0.88	0.95	0.95	0.94
brother, monk	0.33	0.21	0.32	0.41	0.81	0.87	0.88
crane, implement	0.91	0.89	0.94	0.95	0.97	0.94	0.94
lad, brother	0.87	0.86	0.88	0.83	0.86	0.88	0.88
journey, car	0.99	0.99	0.99	0.99	0.99	0.99	0.99
monk, oracle	0.79	0.41	0.71	0.43	0.71	0.79	0.78
food, rooster	0.83	0.27	0.86	0.94	0.90	0.83	0.83
coast, hill	0.64	0.71	0.71	0.49	0.62	0.75	0.79
forest, graveyard	0.88	0.79	0.90	0.92	0.91	0.87	0.87
monk, slave	0.66	0.28	0.61	0.29	0.68	0.79	0.78
coast, forest	0.80	0.81	0.84	0.81	0.84	0.85	0.85
lad, wizard	0.92	0.82	0.93	0.85	0.92	0.94	0.93
chord, smile	0.95	0.99	0.99	0.96	0.99	1.00	1.00
glass, magician	0.90	0.88	0.91	0.90	0.92	0.90	0.89
noon, string	0.99	1.00	1.00	1.00	1.00	1.00	1.00
rooster, voyage	0.96	0.95	0.97	0.98	0.98	0.97	0.96
Avg Correl.	0.85	0.75	0.86	0.84	0.89	0.88	0.88

Table 2. Average Pearson's correlations in the 28 contexts according to the new experiment

5 RELATED WORK

Within the semantic similarity measures [5], below we restrict our attention to the methods based on the information content (IC) approach, which has been employed in different research areas, such as Natural Language Processing [19], Semantic Web [20, 6, 21], Formal Concept Analysis [22, 23, 24], Geographical Information Systems [25, 26, 27], and different application domains, such as health [28], network security [29], and e-learning [30], to mention a few examples. The IC approach, although recognized as "the state of the art on semantic similarity" [3, 4], has shown some limitations. In particular, one objection to the early IC-based mea-

$concept_1$, $concept_2$	sim_R	$sim_{W\&P}$	sim_L	$sim_{J\&C}$	$sim_{P\&S}$	sim_A	$sim_{A\&M}$
car, automobile	0.91	0.95	0.96	0.96	0.96	0.97	0.97
gem, jewel	0.96	0.93	0.95	0.95	0.98	0.98	0.98
journey, voyage	0.86	0.98	0.99	0.99	0.99	0.99	0.99
boy, lad	0.73	0.75	0.91	0.71	0.92	0.89	0.89
coast, shore	0.99	0.86	0.98	0.99	0.98	0.99	0.94
asylum, madhouse	1.00	0.83	1.00	0.83	1.00	1.00	1.00
magician, wizard	1.00	0.97	0.97	0.97	1.00	0.88	0.88
midday, noon	1.00	0.99	0.99	1.00	1.00	1.00	1.00
furnace, stove	0.52	0.20	0.40	0.32	0.43	0.52	0.52
food, fruit	0.64	0.40	0.70	0.74	0.71	0.63	0.60
bird, cock	0.82	0.78	0.83	0.85	0.86	0.83	0.83
bird, crane	0.76	0.79	0.81	0.81	0.79	0.79	0.78
tool, implement	0.89	0.76	0.85	0.72	0.81	0.82	0.83
brother, monk	0.61	0.32	0.56	0.70	0.89	0.39	0.39
crane, implement	0.64	0.56	0.64	0.85	0.82	0.74	0.75
lad, brother	0.77	0.55	0.81	0.63	0.61	0.66	0.66
journey, car	0.56	0.49	0.55	0.75	0.74	0.47	0.54
monk, oracle	0.82	0.52	0.71	0.65	0.72	0.77	0.77
food, rooster	0.75	0.58	0.73	0.76	0.85	0.74	0.74
coast, hill	0.58	0.73	0.58	0.40	0.33	0.58	0.57
forest, graveyard	0.66	0.56	0.66	0.67	0.72	0.62	0.58
monk, slave	0.65	0.52	0.52	0.51	0.67	0.62	0.63
coast, forest	0.71	0.74	0.74	0.84	0.86	0.73	0.73
lad, wizard	0.47	0.50	0.48	0.45	0.67	0.47	0.47
chord, smile	1.00	1.00	1.00	0.99	0.99	1.00	1.00
glass, magician	0.82	0.86	0.77	0.81	0.85	0.70	0.64
noon, string	0.95	0.94	0.96	0.84	0.87	0.86	0.91
rooster, voyage	0.89	0.89	0.90	0.99	0.98	0.92	0.88
Avg Correl.	0.78	0.71	0.78	0.77	0.82	0.77	0.77

Table 3. Average Spearman's correlations in the 28 contexts according to the new experiment

sures relies on the use of large-scale corpora [3, 4, 31]. In fact, evaluating the IC on the basis of statistical information taken from textual corpora requires a considerable amount of manual effort at the 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 [32]. 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] (sim_A in our experiment), arguing that relying only on the maximum among the ICs of the least common subsumers leads to ignoring 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 [14], which shows slightly better performances with respect to the other measure ($sim_{A\,\&\,M}$ 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. Analogously, in [13] ($sim_{P\,\&\,S}$ in our experiment), an IIC approach for semantic similarity has been proposed by relying on the Tversky contrast model [33], that shows a high correlation with human judgment with respect to the state-of-the-art.

With regard to the works of Resnik [1] and Lin [2], it is worth mentioning that according to the former, concept similarity in a taxonomy is computed by considering only concept commonalities, therefore it shows some limitations since pairs of concepts having the same least common subsumers have the same similarity degrees. According to [34], the latter can be re-conducted to the well-known Tversky linear contrast model of similarity mentioned above, which addresses both concept commonalities and differences. In particular, also in [2] the importance of observing an object from different perspectives is emphasized, however, as mentioned in the Introduction, 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 [12], in the late 1990s, a proposal combining the IC with the edge-counting approach has been presented ($sim_{J\,\&\,C}$ in our experiment), showing better performances with respect to the mentioned methods.

It is important to note that, with respect to the existing literature, in this paper we do not present a new IC (or IIC) computing model, and our proposal is independent of it. In fact, although the IIC approaches show 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.

The notion of sense has been addressed by Resnik in [1], 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 [35] 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. Senses are also addressed in [36], where concept similarity is computed between the most-related pairs among the concept's corresponding meanings, but the intended senses of the compared concepts are not considered. Finally, the semantic similarity measure proposed in [31] 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's intended senses.

6 CONCLUSION AND FUTURE WORK

In this work, the novel approach for evaluating semantic similarity in a taxonomy presented in [7] has been refined. In particular, in order to evaluate the relatedness of the generic sense of a concept with its intended sense, the $ASRMP_m$ measure [9] has been selected, and the experimental results, when compared to the ones presented in the original proposal, show for each method an average increment of the average correlation with the human judgment of about 0.04.

As a future work, we plan to refine this approach by defining the intended sense of a concept as a *set* of concepts, rather than a single one and, for example, to rely on the *SemSim* semantic similarity method [6] in order to perform the disambiguation step by comparing the sets of concept senses.

REFERENCES

[1] Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: Mellish, C. S. (Ed.): Proceedings of the 14$^{\text{th}}$ International Joint Conference on Artificial Intelligence – Volume 1 (IJCAI '95). 1995, pp. 448–453, doi: 10.48550/arXiv.cmp-lg/9511007.

[2] Lin, D.: An Information-Theoretic Definition of Similarity. In: Shavlik, J. W. (Ed.): Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98). 1998, pp. 296–304.

[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. Journal of the Association for Information Science and Technology, Vol. 69, 2018, No. 8, pp. 1023–1034, doi: 10.1002/asi.24021.

[4] Batet, M.—Sánchez, D.: Leveraging Synonymy and Polysemy to Improve Semantic Similarity Assessments Based on Intrinsic Information Content. Artificial Intelligence Review, Vol. 53, 2020, No. 3, pp. 2023–2041, doi: 10.1007/s10462-019-09725-4.

[5] Chandrasekaran, D.—Mago, V.: Evolution of Semantic Similarity – A Survey. ACM Computing Surveys, Vol. 54, 2021, No. 2, Art. No. 41, doi: 10.1145/3440755.

[6] Formica, A.—Missikoff, M.—Pourabbas, E.—Taglino, F.: Semantic Search for Matching User Requests with Profiled Enterprises. Computers in Industry, Vol. 64, 2013, No. 3, pp. 191–202, doi: 10.1016/j.compind.2012.09.007.

[7] Formica, A.—Taglino, F.: An Enriched Information-Theoretic Definition of Semantic Similarity in a Taxonomy. IEEE Access, Vol. 9, 2021, pp. 100583–100593, doi: 10.1109/ACCESS.2021.3096598.

[8] Hadj Taieb, M. A.—Zesch, T.—Ben Aouicha, M.: A Survey of Semantic Relatedness Evaluation Datasets and Procedures. Artificial Intelligence Review, Vol. 53, 2020, No. 6, pp. 4407–4448, doi: 10.1007/s10462-019-09796-3.

[9] El Vaigh, C. B.—Goasdoué, F.—Gravier, G.—Sébillot, P.: A Novel Path-Based Entity Relatedness Measure for Efficient Collective Entity Linking. In: Pan, J. Z., Tamma, V., d'Amato, C., Janowicz, K., Fu, B., Polleres, A., Seneviratne, O., Kagal, L. (Eds.): The Semantic Web – ISWC 2020. Springer, Cham, Lecture Notes in Computer Science, Vol. 12506, 2020, pp. 164–182, doi: 10.1007/978-3-030-62419-4_10.

[10] Formica, A.—Taglino, F.: Semantic Relatedness in DBPedia: A Comparative and Experimental Assessment. Information Sciences, Vol. 621, 2023, pp. 474–505, doi: https://doi.org/10.1016/j.ins.2022.11.025.

[11] Miller, G. A.—Charles, W. G.: Contextual Correlates of Semantic Similarity. Language and Cognitive Processes, Vol. 6, 1991, No. 1, pp. 1–28, doi: 10.1080/01690969108406936.

[12] Jiang, J. J.—Conrath, D. W.: Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In: Chen, K. J., Huang, C. R., Sproat, R. (Eds.): Proceedings of the 10th Research on Computational Linguistics International Conference (ROCLING 1997). The Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Taipei, Taiwan, 1997, pp. 19–33, doi: 10.48550/arXiv.cmp-lg/9709008.

[13] Pirró, G.: A Semantic Similarity Metric Combining Features and Intrinsic Information Content. Data and Knowledge Engineering, Vol. 68, 2009, No. 11, pp. 1289–1308, doi: 10.1016/j.datak.2009.06.008.

[14] Adhikari, A.—Singh, S.—Mondal, D.—Dutta, B.—Dutta, A.: A Novel Information Theoretic Framework for Finding Semantic Similarity in WordNet. 2016, doi: 10.48550/arXiv.1607.05422.

[15] Wu, Z.—Palmer, M.: Verbs Semantics and Lexical Selection. Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics (ACL '94), 1994, pp. 133–138, doi: 10.3115/981732.981751.

[16] Rada, R.—Mili, H.—Bicknell, E.—Blettner, M.: Development and Application of a Metric on Semantic Nets. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, 1989, No. 1, pp. 17–30, doi: 10.1109/21.24528.

[17] Schuhmacher, M.—Ponzetto, S. P.: Knowledge-Based Graph Document Modeling. Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14), 2014, pp. 543–552, doi: 10.1145/2556195.2556250.

[18] Formica, A.—Taglino, F.: Semantic Similarity with Concept Senses. 2021, doi: 10.17632/ck365nbpk9.2 (V2).

[19] Majumder, G.—Pakray, P.—Gelbukh, A. F.: Measuring Semantic Textual Similarity of Sentences Using Modified Information Content and Lexical Taxonomy. International Journal of Computational Linguistics and Applications, Vol. 7, 2016, No. 2, pp. 65–85.

[20] Formica, A.—Missikoff, M.—Pourabbas, E.—Taglino, F.: Weighted Ontology for Semantic Search. In: Meersman, R., Tari, Z. (Eds.): On the Move to Meaningful Internet Systems: OTM 2008. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 5332, 2008, pp. 1289–1303, doi: 10.1007/978-3-540-88873-4_26.

[21] Meymandpour, R.—Davis, J. G.: A Semantic Similarity Measure for Linked Data: An Information Content-Based Approach. Knowledge-Based Systems, Vol. 109, 2016, pp. 276–293, doi: 10.1016/j.knosys.2016.07.012.

[22] Formica, A.: Similarity Reasoning in Formal Concept Analysis: From One- to Many-Valued Contexts. Knowledge and Information Systems, Vol. 60, 2019, No. 2, pp. 715–739, doi: 10.1007/s10115-018-1252-4.

[23] Formica, A.: Concept Similarity in Formal Concept Analysis with Many-Valued Contexts. Computing and Informatics, Vol. 40, 2021, No. 3, pp. 469–488, doi: 10.31577/cai_2021_3_469.

[24] Wang, F.—Wang, N.—Cai, S.—Zhang, W.: A Similarity Measure in Formal Concept Analysis Containing General Semantic Information and Domain Information. IEEE Access, Vol. 8, 2020, pp. 75303–75312, doi: 10.1109/ACCESS.2020.2988689.

[25] Formica, A.—Mazzei, M.—Pourabbas, E.—Rafanelli, M.: Approximate Query Answering Based on Topological Neighborhood and Semantic Similarity in OpenStreetMap. IEEE Access, Vol. 8, 2020, pp. 87011–87030, doi: 10.1109/ACCESS.2020.2992202.

[26] Formica, A.—Pourabbas, E.: Content Based Similarity of Geographic Classes Organized as Partition Hierarchies. Knowledge and Information Systems, Vol. 20, 2009, No. 2, pp. 221–241, doi: 10.1007/s10115-008-0177-8.

[27] Schwering, A.: Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. Transactions in GIS, Vol. 12, 2008, No. 1, pp. 5–29, doi: 10.1111/j.1467-9671.2008.01084.x.

[28] Abdelrahman, A. M. B.—Kayed, A.: A Survey on Semantic Similarity Measures Between Concepts in Health Domain. American Journal of Computational Mathematics, Vol. 5, 2015, No. 2, pp. 204–214, doi: 10.4236/ajcm.2015.52017.

[29] Weller-Fahy, D. J.—Borghetti, B. J.—Sodemann, A. A.: A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection. IEEE Communications Surveys and Tutorials, Vol. 17, 2014, No. 1, pp. 70–91, doi: 10.1109/COMST.2014.2336610.

[30] Barbagallo, A.—Formica, A.: ELSE: An Ontology-Based System Integrating Semantic Search and e-Learning Technologies. Interactive Learning Environments, Vol. 25, 2017, No. 5, pp. 650–666, doi: 10.1080/10494820.2016.1172240.

[31] Jeong, S.—Yim, J. H.—Lee, H. J.—Sohn, M. M.: Semantic Similarity Calculation Method Using Information Contents-Based Edge Weighting. Journal of Internet

Services and Information Security, Vol. 7, 2017, No. 1, pp. 40–53, doi: 10.22667/JISIS.2017.02.31.040.

[32] Lastra-Díaz, J. J.—García-Serrano, A.: A New Family of Information Content Models with an Experimental Survey on WordNet. Knowledge-Based Systems, Vol. 89, 2015, pp. 509–526, doi: 10.1016/j.knosys.2015.08.019.

[33] Tversky, A.: Features of Similarity. Psychological Review, Vol. 84, 1977, No. 4, pp. 327–352, doi: 10.1037/0033-295X.84.4.327.

[34] Cazzanti, L.—Gupta, M. R.: Information-Theoretic and Set-Theoretic Similarity. 2006 IEEE International Symposium on Information Theory, 2006, pp. 1836–1840, doi: 10.1109/ISIT.2006.261752.

[35] Gabsi, I.—Kammoun, H.—Brahmi, S.—Amous, I.: MeSH-Based Disambiguation Method Using an Intrinsic Information Content Measure of Semantic Similarity. Procedia Computer Science, Vol. 112, 2017, pp. 564–573, doi: 10.1016/j.procs.2017.08.169.

[36] Gao, J. B.—Zhang, B. W.—Chen, X. H.: A WordNet-Based Semantic Similarity Measurement Combining Edge-Counting and Information Content Theory. Engineering Applications of Artificial Intelligence, Vol. 39, 2015, pp. 80–88, doi: 10.1016/j.engappai.2014.11.009.

Anna Formica received her degree (Hons.) in mathematics from the Sapienza University of Rome in 1989. She is currently Senior Researcher with the Istituto di Analisi dei Sistemi ed Informatica (IASI) "Antonio Ruberti" of the Italian National Research Council (Consiglio Nazionale delle Ricerche – CNR), Rome, where she leads the Software and Knowledge-based Systems (SaKS) Group. She took part in various research projects of the European framework programs and bilateral projects with international institutions. Her current research interests include semantic web, similarity reasoning, semantic relatedness of web resources, formal specification and validation of domain ontologies, fuzzy formal concept analysis, geographical information systems, and e-learning. She serves as a referee for several international journals and conferences.

Francesco Taglino graduated in information science at the Sapienza University of Rome 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, a member of the Software and Knowledge-based Systems (SaKS) group. His main research interests are knowledge representation and reasoning and semantic technologies, and 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 a referee in several international journals and conferences.