Multimedia Tools and Applications (2021) 80:32335–32378
https://doi.org/10.1007/s11042-021-10966-1

A unified framework for semantic similarity computation of concepts

Yuncheng Jiang [1]

Received: 2 August 2020 / Revised: 5 January 2021 / Accepted: 14 April 2021 /
Published online: 29 July 2021

Abstract
Semantic similarity assessment between concepts is an important task in many language related applications. In the past, many approaches to assess similarity of concepts have been proposed by using one knowledge source. In this paper, some limitations of the existing similarity measures are identified. To tackle these problems, we propose an extensive study for semantic similarity of concepts from which a unified framework for semantic similarity computation is presented. Based on our framework, we give some generic and flexible approaches to semantic similarity measures resulting from instantiations of the framework. In particular, we obtain some new approaches to similarity measures that existing methods cannot deal with by introducing multiple knowledge sources. The evaluation based on eight benchmarks, three widely used benchmarks (i.e., M&C, R&G, and WordSim-353 benchmarks) and five benchmarks developed in ourselves (i.e, Jiang-1, Jiang-2, Jiang-3, Jiang-4, and Jiang-5 benchmarks), sustains the intuitions with respect to human judgements. Overall, some methods proposed in this paper have a good human correlation (Pearson correlation with human judgments and Spearman correlation with human judgments) and constitute some effective ways of determining semantic similarity between concepts.

Keywords Semantic similarity · Concept similarity · IC (information content)-based measures · Feature-based measures · Distance-based measures · Multiple knowledge sources

1 Introduction

Semantic similarity between concepts is becoming a common problem for many applications such as natural language processing, text categorization, text clustering, information retrieval, and word sense disambiguation [1, 9, 12, 36, 37, 55, 57]. However, making judgments about the semantic similarity of different concepts is a routine yet deceptively complex task. To

✉ Yuncheng Jiang
ycjiang@scnu.edu.cn; ycjiang21@qq.com

[1] School of Computer Science, South China Normal University, Guangzhou 510631, China

perform it, people need to draw on an immense amount of background knowledge about the concepts. Usually, these sources can be search engines [15], topical directories such as Open Directory Project [46], well-defined semantic networks such as WordNet [24, 43], more domain-dependent ontologies [67, 74] such as Gene Ontology [17] and biomedical ontologies MeSH or SNOMED CT [4, 69], Wikipedia [34, 37], or Linked Data [13, 56]. In fact, several works have been developed in the past years proposing semantic similarity measures. According to the concrete knowledge sources exploited and the way in which they are used, various similarity measures have been proposed [29, 71, 72]. Semantic similarity measures can be classified into four main categories: [37, 47, 51]: (1) distance-based models that are based on the structural representation of the underlying context; (2) feature-based models that define concepts or entities as sets of features; (3) statistical methods that consider statistics derived from the underlying context; and (4) hybrid models that comprise combinations of the three basic categories. Concretely, distance-based models, also referred to as edge-counting or path-based methods, define similarity as a function of distance between concepts [51, 62]. Feature-based methods assume that concepts can be represented as sets of features. They assess the similarity of concepts based on the commonalities among their feature sets: any increase in common features among concepts results in a higher similarity score and any decrease in shared features results in lower levels of similarity [51, 80]. Statistical similarity measures incorporate statistics derived from various aspects of the underlying domain into the similarity computation.

It is worth noting that all these measures mentioned above are some specific computation methods by using different knowledge sources such as WordNet [20], Wikipedia [48], or Linked Data [10] or different mathematical tools such as information content (IC) [63], pointwise mutual information (PMI) [14], or latent semantic analysis (LSA) [19]. Furthermore, for the same kind of knowledge source, different computation approaches for semantic similarity need different contents of the knowledge source. For example, in Wikipedia based similarity measures, IC-based measures need the category structure of Wikipedia, however, feature-based methods need the articles (e.g., the redirect pages and hyperlinks) of Wikipedia. In fact, we can propose some novel computation approaches for semantic similarity of concepts by exploiting new knowledge sources or mathematical tools. Clearly, there are some issues in existing researches. Firstly, there are lots of computation approaches of semantic similarity, however, there is not a unified framework for these methods. Therefore, in practical applications it is difficult for the users to choose which computation method for semantic similarity of concepts. Secondly, if two concepts A and B belong to two heterogeneous knowledge sources, the semantic similarity between A and B cannot be computed using existing methods. For example, if $A \in WordNet$, $A \notin DBpedia$, $B \in DBpedia$, and $B \notin WordNet$, existing approaches cannot compute the semantic similarity $sim(A, B)$. Of course, if two concepts A and B belong to two homogeneous knowledge sources such as two different domain ontologies built in the same language, the value of $sim(A, B)$ can be computed by using existing methods such as [8, 71].

To fill these gaps, this paper proposes an extensive study for semantic similarity of concepts from which a unified framework for semantic similarity computation is presented. It should be noted that Cross et al. [18] and Harispe et al. [32] have studied the unified framework issue for semantic similarity measures. However, their works are different from our research in this paper: Cross et al. [18] and Harispe et al. [32] present a framework for unifying ontology-based semantic similarity measures, and we will propose a unified framework for semantic similarity measures for multiple heterogeneous knowledge sources [68] such as WordNet [20], ontologies [77], Wikipedia [48], and Linked Data [10]. Based on our framework for semantic similarity of

concepts, we give some generic and flexible approaches to semantic similarity measures resulting from instantiations of the framework. The main contributions of this paper are as follows:

- The semantic representation and a unified framework for semantic similarity computation of concepts are presented.
- Some generic and flexible approaches to semantic similarity measures of concepts resulting from instantiations of the framework are provided.
- Several new approaches to semantic similarity computation of concepts that existing methods cannot measure are proposed.

It is worth mentioning that semantic similarity measures can also be used in multimedia system such as multimedia databases and retrieval, personalized electronic journals, multimedia encyclopedias, digital libraries, executive information systems, and multimedia documents. For example, in multimedia (e.g., image, audio or video) retrieval with text annotation, we may use the semantic similarity of text to assist multimedia retrieval, where the computation of semantic similarity of text can be implemented by exploiting semantic similarity measures of concepts. Another example, in digital libraries or multimedia documents, there are many image, audio, video, and text data. In a similar manner, we also can utilize the semantic similarity of text to assist the processing (e.g., retrieval, classification, recommendation, mining, and analysis) of digital libraries and multimedia documents. That is, semantic similarity measure of concepts is also relevant to multimedia system.

The rest of the paper is organized as follows. In the next section, we briefly present the related works on semantic similarity measures. Section 3 presents our unified framework for semantic similarity computation of concepts. This includes the semantic representation of concepts and a framework for semantic similarity computation. In Section 4, we investigate several similarity measures resulting from instantiations of the framework. Section 5 is devoted to presenting detail of experiments and evaluation of our approaches. Finally, in Section 6, we draw our conclusion and present some perspectives for future research.

2 Related work

As a fundamental concept in theories of perception, behavior, social bonding, learning, and judgment, the notion of similarity has been extensively studied for several decades. Many researchers have endeavored to understand and represent the way humans judge the similarity of two or more objects [12, 27, 51, 53, 76, 80]. Semantic similarity reflects the relationship between the meaning of two concepts (words, entities, or terms), sentences (or short texts) or documents (or texts) [21, 31, 54, 59]. The literature on semantic similarity measures is very extensive, thus, we only focus on the measures that are evaluated in this work, that is, this section takes an overview of the methods for semantic similarity measures for concepts.

As stated in Section 1, semantic similarity between concepts (semantic similarity for short) can be computed based on a set of factors derived typically from a knowledge representation model. Depending on the structure of the application context and its knowledge representation model, various similarity measures have been proposed and different families of methods can also be identified [51, 71]. These families are [37, 47, 51, 58]: (1) distance-based similarity measures; (2) feature-based similarity measures; (3) statistical similarity measures; and (4) hybrid similarity measures.

2.1 Distance-based similarity measures

Modern research in this area starts with the work presented by Rada et al. [62]. Concretely, Rada et al. propose to use the length of the shortest path between concepts as a measurement of distance. Formally, their definition of conceptual distance is as follows:

$$Dist(A, B) = \text{minimum number of edges separating } a \text{ and } b,$$

where A and B are the two concepts represented by the nodes a and b, respectively, in an is-a semantic net [24].

The distance measure is converted to a similarity measure by subtracting the path length from the maximum possible path length, which can be shown in the following equation:

$$Sim(A, B) = 2 \times Distance_{max} - Dist(A, B),$$

where $Distance_{max}$ is the maximum possible path length [24].

The work proposed by Rada et al. [62] opens up the family of edge-counting semantic measures and shows that conceptual distance (or similarity) between concepts in a semantic network is proportional to the length of the path that links them [38]. The ideas of Rada et al. are followed by other works such as Wu and Palmer [82], Leacock and Chodorow [39], Hirst and St-Onge [33], Li et al. [41], Pedersen et al. [56], and Garla and Brandt [25] which also propose similarity measures based on features derived from the length of shortest path between concepts. For example, the metric presented by Wu and Palmer [82] relies on the fact that in is-a hierarchies, concepts that are more distant from the root are more specific than the ones that are near the root. Formally, the conceptual similarity between concepts A and B is defined as follows:

$$Sim(A, B) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3},$$

where N_1 (N_2) is the number of edges on the path from A (B) to $LCS(A, B)$, N_3 is the number of edges on the path from $LCS(A, B)$ to root, and $LCS(A, B)$ means the least common subsumber (LCS) of concept A and concept B [24].

Leacock and Chodorow [39] propose a non-linear adaptation of Rada's distance to define the similarity measure:

$$Sim(A, B) = -\log\left(\frac{Dist(A, B)}{2 \times Max\_depth}\right),$$

where Max_depth is the longest of the shortest paths linking a concept to the concept which subsumes all the others [32]. It should be noted that the non-linear adaptation here means logarithmic function of Rada's distance, while the adaptation in [6] means runtime/semantic adaptation and management of software to support source-code semantic flexibility.

Garla and Brandt [25] give a proposal for the normalization of the metric of Leacock and Chodorow to the unit interval as follows [3]:

$$Sim(A, B) = 1 - \frac{\log(Dist(A, B))}{\log(2 \times Max\_depth)}.$$

Li et al. [41] introduce a family of ten different parametric similarity measures whose core idea is the breaking down of the overall similarity function into a combination of functions linearly or nonlinearly, where each base function relies on a different taxonomical feature such as the

length of the shortest path between concepts, and the depth of the lowest common ancestor [38]. One of the best measures among them is shown in the following equation:

$$Sim(A,B) = e^{-\alpha * Dist(A,B)} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}},$$

where $Dist(A, B)$ is the number of edges separating A and B, h is the depth of LCS of A and B, α and β are parameters scaling the contribution of $Dist(A, B)$ and h, $\alpha \geq 0$ and $\beta > 0$.

2.2 Feature-based similarity measures

Feature-based methods assume that concepts can be represented as sets of features. They assess the similarity of concepts based on the commonalities among their feature sets: any increase in common features among concepts results in a higher similarity score and any decrease in shared features results in lower levels of similarity [51, 80]. For discrete-valued vectors similarity measures are inspired by the comparison of sets and the cardinality of sets. Some common set-inspired similarity measures for discrete-valued vectors include [45]:

$$\text{Jaccard coefficient } Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|},$$

$$\text{Dice coefficient } Dice(A,B) = \frac{2 \times |A \cap B|}{|A| + |B|},$$

$$\text{Salton Cosine coefficient } SaltonCosine(A,B) = \frac{|A \cap B|}{|A| \times |B|},$$

where A and B denote the sets of features that correspond to concepts a and b.

The Tversky ration model [80] is defined by a weighted variant for the complement of the symmetric difference between the feature sets of two concepts and considers the distinctive characteristics of each concept (the features of one concept which are not part of the other):

$$Tversky(A,B) = \frac{|A \cap B|}{|A \cap B| + \alpha |A - B| + \beta |B - A|} \text{ for } \alpha, \beta > 0,$$

where α and β represent the relative contribution of unique features of A and B in the similarity value, respectively. The α and β parameters can be used to reflect the symmetric or asymmetric nature of a given context: if $\alpha = \beta$ then $Tversky(A, B) = Tversky(B, A)$ thus, the similarity comparison is symmetric, otherwise, it is asymmetric (i.e., $Tversky(A, B) \neq Tversky(B, A)$) [51].

With a perspective from set theory, the meaning of the Tversky measure is clear and well-founded. However, the feature sets associated to each concept cannot be derived directly from an ontology, which is a serious drawback for its practical implementation [38]. With the aim of bridging the gap in the Tversky measure, Sanchez et al. [73] introduce a feature-based dissimilarity measure which is based on the use of the common ancestors between concepts as a measure of their degree of similarity:

$$Dis(A,B) = \log_2 \left(1 + \frac{|\varphi(A) - \varphi(B)| + |\varphi(B) - \varphi(A)|}{|\varphi(A) - \varphi(B)| + |\varphi(B) - \varphi(A)| + |\varphi(A) \cap \varphi(B)|} \right),$$

where $\varphi(C) = \{D \in AllCons | C \leq D\}$, $AllCons$ is the set of concepts of a given ontology, and $\leq$ is a binary relation (i.e., concept subsumption).

The definition of the set of features such as the set of synonyms (called synsets in WordNet), definitions (i.e., glosses, containing textual descriptions of word senses), and the set of subconcepts (or subclasses, subcategories) is crucial in feature-based measures.

The Rodriguez and Egenhofer measure [65] is computed as the weighted sum of similarities between synsets, features (e.g., meronyms, attributes, etc.) and semantic neighborhoods (those linked via semantic pointer) of two concepts A and B:

$$Sim(A,B) = w \cdot S_{synsets}(A,B) + u \cdot S_{features}(A,B) + v \cdot S_{neighborhoods}(A,B) \text{ for } w,u,v \geq 0.$$

Weights assigned to w, u, and v depend on the characteristics of the ontologies. Only common specification components can be used in a similarity assessment. Their respective weights add up to 1.0.

X-Similarity [58] relies on matching between synsets and term description sets. The term description sets contains words extracted by parsing term definitions ("glosses" in WordNet or "scope notes" in MeSH). Two terms are similar if their synsets or description sets or, the synsets of the terms in their neighborhood (e.g., more specific and more general terms) are lexically similar. The similarity function is expressed as follows:

$$Sim(A,B) = \begin{cases} 1, & \text{if } S_{synsets}(A,B) > 0 \\ \max\{S_{neighborhoods}(A,B), S_{descriptions}(A,B)\}, & \text{if } S_{synsets}(A,B) = 0 \end{cases}.$$

Jiang et al. [37] investigate some feature-based approaches to semantic similarity assessment of concepts using Wikipedia and give the following framework for feature-based similarity using the sets of all synonym sets, gloss sets, anchor sets, and category sets of Wikipedia concepts:

$$Sim(A,B) = S_{concepts}(S_{synonyms}(Synonyms_A, Synonyms_B), S_{glosses}(Glosses_A, Glosses_B), S_{anchors}(Anchors_A, Anchors_B), S_{categories}(Categories_A, Categories_B)).$$

2.3 Statistical similarity measures

Statistical similarity measures incorporate statistics derived from various aspects of the underlying domain into the similarity computation [51]. Several approaches use the popularity of terms in a document as a measure of their informativeness and use this as a basis for measuring the similarity [34, 38, 42, 51, 63, 64, 71, 72]. These approaches are also known as Information Content (IC)-based measures.

Resnik [63] proposes an IC-based method which is not sensitive to the problem of varying link distance. They assume that the information shared by two concepts is indicated by the IC of the concepts that subsume them in a net (e.g. WordNet) [24]:

$$Sim(A,B) = IC(LCS(A,B)),$$

where $IC(C) = -\log(p(C))$ and $p(C)$ is the probability of encountering an instance of concept C in a given corpus (e.g. Brown Corpus).

Resnik's metric has two problems: any pair of concepts (words) with the same LCS will have the same semantic similarity; similarity between the same concepts (words) is not equal to one [24]. To correct these problems, Lin [42], Jiang and Conrath [35] propose their methods. Jiang and Conrath represent their metric as follows [35, 38]:

$$Distance(A,B) = IC(A) + IC(B) - 2 \times IC(LCS(A,B)) \text{ and}$$
$$Sim(A,B) = 1 - \frac{Distance(A,B)}{2}.$$

Lin's similarity function [42] is expressed as follows:

$$Sim(A,B) = \frac{2 \times IC(LCS(A,B))}{IC(A) + IC(B)}.$$

Recently, there are many researches in IC-based semantic similarity measure [4, 34, 51]. For example, Jiang et al. [34] present several new methods to IC computation of a concept and similarity computation between two concepts drawn from Wikipedia category structure. Since Wikipedia category structure is a graph, naturally, the semantic similarity between concepts can be assessed by extending traditional information theoretic approaches (i.e., IC-based approaches).

All the IC-based similarity measures require an IC model. An IC model is a concept-valued function that assigns an IC value to each concept [38]. Except the corpus-based IC models [24, 35] [38, 42], some intrinsic IC models are developed. The pioneering work is the intrinsic IC model of Seco et al. [75]. Some new intrinsic IC models are also proposed [28, 49, 70, 72]. For example, in a recent work, Sanchez et al. [72] propose estimating the IC value of concept C as the ratio between the number of leaves on the taxonomical hierarchy under the concept C (as a measure of C's generality) and the number of taxonomical subsumers above C including itself (as a measure of C's concreteness). Formally,

$$IC(C) = -\log\left(\frac{\frac{|leaves(C)|}{|subsumers(C)|} + 1}{max\_leaves + 1}\right),$$

where $leaves(C)$ is the set of concepts found at the end of the taxonomical tree under concept C and $subsumers(C)$ is the complete set of taxonomical ancestors of C including itself. The ratio is normalized by the least informative concept (i.e., the root of the taxonomy), for which the number of leaves is the total amount of leaves in the taxonomy (max_leaves) and the number of subsumers including itself is 1. To produce values in the range [0, 1] (i.e., in the same range as the original probability) and avoid log(0) values, 1 is added to the numerator and denominator.

Other approaches such as pointwise mutual information (PMI) [14] and vector-based methods such as latent semantic analysis (LSA) [19] and explicit semantic analysis (ESA) [23] can be classified as statistical semantic similarity measures as they use functions of term frequency for computing the similarity [51].

2.4 Hybrid similarity measures

A number of approaches can be classified as hybrid methods: they are based on combinations of some of the above presented methods. For example, Pirro [60] presents a similarity metric combining the feature-based and information theoretic theories of similarity. In particular, the proposed metric exploits the notion of intrinsic IC which quantifies IC values by scrutinizing how concepts are arranged in an ontological structure. Meng et al. [50] introduce a variant of the Lin measure [42], concretely, the similarity measure of Meng et al. [50] is a hybrid measure

that combines the Lin IC-based measure with a power factor based on the shortest path length between concepts. In IS-A taxonomies, intrinsic IC (IIC) [75] incorporates the number of subclass of a concept for estimating the information content: the higher the number of subclass of a term, the lower its informativeness [51]. IIC has also been combined with feature-based [60] and edge counting methods [61, 78]. Gao et al. [24] propose an approach to calculate the semantic similarity between word pairs based on WordNet, specifically, they present an approach for semantic similarity measuring which is based on edge-counting and IC theory.

3 A framework for semantic similarity computation

To compute the semantic similarity $sim(A, B)$ for two concepts A and B, we firstly need to get some related information such as synonyms or taxonomy structures of A and B from certain knowledge source such as WordNet [20] or domain ontologies [77]. For example, if users want to evaluate $Sim(A, B)$ using IC-based measures, the users must have a taxonomy structure T (or two homogeneous taxonomy structures T_1 and T_2) such that $A, B \in T$ (or $A \in T_1$ and $B \in T_2$). If A and B belong to two different heterogeneous knowledge sources such as $A \in WordNet$ and $B \in DBpedia$, $Sim(A, B)$ cannot be computed using existing IC-based methods. Similarly, to compute $Sim(A, B)$, distance-based measures or feature-based measures also need some related information of A and B. If these related information comes from different knowledge sources, existing distance-based or feature-based measures also cannot compute $Sim(A, B)$. On the other hand, when we compute $Sim(A, B)$, the more related information of A and B we get, the more accurate result of $Sim(A, B)$ can be obtained. Therefore, we need to get as much related information for A and B as possible from different knowledge sources in order to better computation of $Sim(A, B)$. For instance, we can get the synonyms or taxonomy structures of A (or B) via WordNet [20], domain ontologies [77], Wikipedia [34, 37], DBpedia [11] or YAGO [79]. Obviously, we have to integrate these related information of A and B that comes from different (heterogeneous) knowledge sources. To this end, we first present the notion of semantic representation of concepts in theory. We then give a framework for semantic similarity computation based on the semantic representation of concepts.

3.1 Semantic representation of concepts

How to represent a concept for semantic similarity computation? Because the semantic information of a concept may come from multiple knowledge sources, in particular, with the development of information technology, some new knowledge sources might be developed, we need a flexible way to represent the semantic information of a concept. Let us see an example.

> Example 1. Consider a concept $C_1 = Artificial\ Intelligence$. Clearly, from WordNet, Wikipedia and DBpedia we know that $C_1 \in WordNet$, $C_1 \in Wikipedia$, and $C_1 \in DBpedia$. From WordNet we know that the set of synonyms of C_1 is $synonyms(C_1) = \{AI\}$. From Wikipedia or DBpedia we have that the set of synonyms of C_1 is $synonyms(C_1) = \{AI, Machine\ Intelligence, Cognitive\ System, Computational\ Rationality, Soft\ AI, \ldots\}$. Similarly, from WordNet we also know that C_1 has a taxonomy structure (tree structure) $TS_{WordNet}(C_1)$ (see Fig. 1), and C_1 has a taxonomy structure (graph structure) $TS_{Wikipedia}(C_1)$ (see Fig. 2) or knowledge network (graph structure) $TS_{DBpedia}(C_1)$ (see Fig. 3) from Wikipedia or DBpedia, respectively. Of course, we also can get other semantic information such as glosses for C_1 from WordNet, Wikipedia, DBpedia, or YAGO.

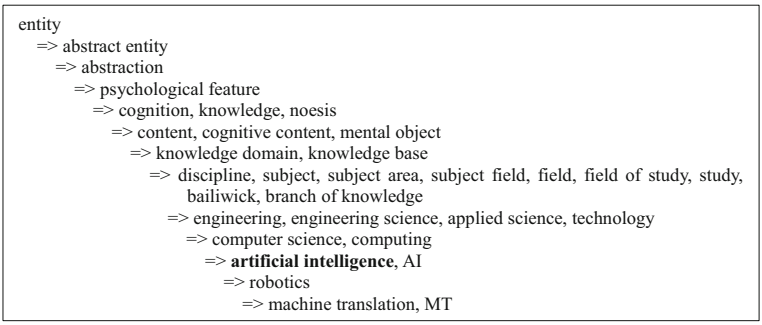

Fig. 1 Taxonomy structure of *Artificial Intelligence* in WordNet

Consider another concept C_2 = *Semantic Web*. From WordNet, Wikipedia and DBpedia we know that $C_2 \notin$ *WordNet*, $C_2 \in$ *Wikipedia*, and $C_2 \in$ *DBpedia*. Clearly, we cannot obtain the semantic information such as synonyms or taxonomy structure of C_2 from WordNet, however, the information can be obtained from Wikipedia or DBpedia.

Now we propose the definition of semantic representation of concepts.

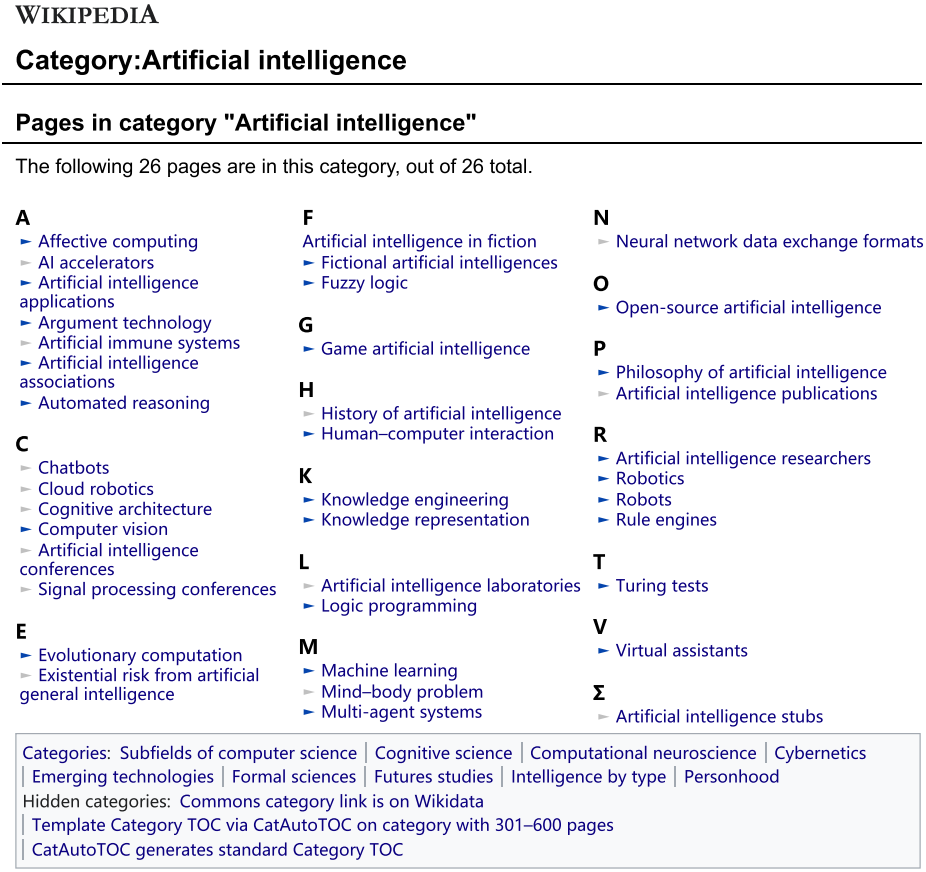

Fig. 2 Taxonomy structure of *Artificial Intelligence* in Wikipedia

About: 人工智能

人工智能（英語：Artificial Intelligence，缩写为AI）亦稱智械、機器智能，指由人製造出來的機器所表現出來的智能。通常人工智能是指通过普通電腦程式來呈現人類智能的技術。該詞也指出研究這樣的智能系統是否能夠實現，以及如何實現。同时，通過醫學、神經科學、機器人學及統計學等的進步，常態預測則認為人類的無數職業也逐漸被其取代。 人工智能於一般教材中的定义领域是“智慧主体（intelligent agent）的研究与设计”，智慧主体指一个可以观察周遭环境并作出行动以达致目标的系统。约翰·麦卡锡于1955年的定义是「制造智能机器的科学与工程」。安德烈亚斯·卡普兰（Andreas Kaplan）和迈克尔·海恩莱因（Michael Haenlein）将人工智能定义为“系统正确解释外部数据，从这些数据中学习，并利用这些知识通过灵活适应实现特定目标和任务的能力”。 人工智能的研究是高度技术性和专业的，各分支领域都是深入且各不相通的，因而涉及範圍極廣。人工智能的研究可以分为几个技术问题。其分支领域主要集中在解决具体问题，其中之一是，如何使用各种不同的工具完成特定的应用程序。

Property	Value
rdfs:label	• 人工智能 (zh) • Artificial Intelligence (en)
dct:subject	• dbc:Emerging_technologies • dbc:Artificial_intelligence • dbc:Computational_fields_of_study • dbc:Computational_neuroscience • dbc:Formal_sciences • dbc:Cybernetics • dbc:Unsolved_problems_in_computer_science
owl:sameAs	• freebase:人工智能 • dbpedia-fr:人工智能 • wikidata:人工智能 • http://mn.dbpedia.org/resource/Хиймэл_оюун • http://tt.dbpedia.org/resource/Ясалма_интеллект • http://hi.dbpedia.org/resource/कृत्रिम_बुद्धि • http://lt.dbpedia.org/resource/Dirbtinis_intelektas • http://or.dbpedia.org/resource/ଆର୍ଟିଫିସିଆଲ_ଇଣ୍ଟେଲିଜେନ୍ସ • http://ia.dbpedia.org/resource/Intelligentia_artificial • .....

Fig. 3 Knowledge network of *Artificial Intelligence* in DBpedia

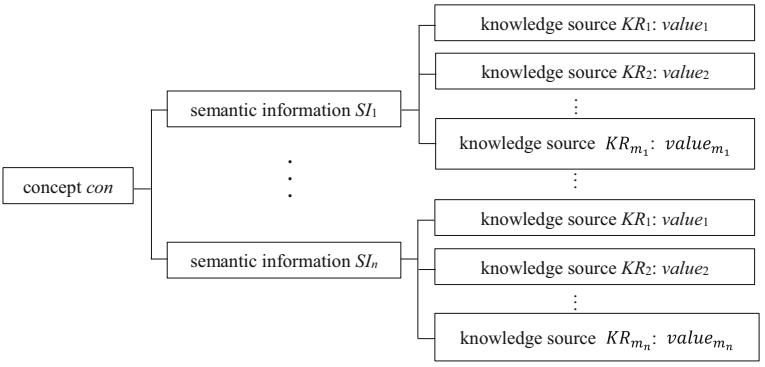

Fig. 4 Semantic representation of concepts

Definition 1. Let *con* be a concept. The semantic representation of concept *con* is defined as follows:

$$con = \langle SI_1(con), SI_2(con), \ldots, SI_n(con)\rangle,$$

where the *i*th semantic information $SI_i(con)$ of *con* $(1 \le i \le n)$ is as below:

$$SIi(con) = \langle\langle KS_{i_1} : value_{i_1}\rangle, \langle KS_{i_2} : value_{i_2}\rangle, \ldots, \langle KS_{i_m} : value_{i_m}\rangle\rangle,$$

where KS_{i_j} $(1 \le j \le m)$ means the *j*th knowledge source of $SI_i(con)$, and $value_{i_j}$ is the value of $SI_i(con)$ from KS_{i_j} in $\langle KS_{i_j}: value_{i_j}\rangle$.

The semantic representation of concept *con* can be shown in Fig. 4.

To understand Definition 1, let us see a simple example.

Example 2. From Example 1 we have the following:

$$Artificial\ Intelligence = \langle glosses(Artificial\ Intelligence), synonyms(Artificial\ Intelligence), \ldots, taxonomy\ structure(Artificial\ Intelligence)\rangle,$$

where *glosses*, *synonyms*, …, and *taxonomy structure* represent the titles of all semantic information of *Artificial Intelligence*, and.

glosses(*Artificial Intelligence*) = ⟨⟨*WordNet*: *the branch of computer science that deal with writing computer programs that can solve problems creatively*, …⟩,
…,
⟨*Wikipedia*: *Artificial intelligence* (*AI*), *sometimes called machine intelligence*, *is intelligence demonstrated by machines*, …⟩⟩,
synonyms(*Artificial Intelligence*) = ⟨⟨*WordNet*: {*AI*}⟩, …, ⟨*Wikipedia*: {*AI*, *Machine Intelligence*, *Cognitive System*, *Computational Rationality*, *Soft AI*, …}⟩⟩,
taxonomy structure(*Artificial Intelligence*) = ⟨⟨*WordNet*: $TS_{WordNet}$⟩, …, ⟨*Wikipedia*: $TS_{Wikipedia}$⟩⟩.

Remark 1. The semantic representation of concepts in Definition 1 is a flexible representation mechanism. On one hand, we don't fix the numbers and kinds of semantic information of a concept, that is, users may add different semantic information such as *hyponym* (or *sub-concept*), *hypernym* (or *super-concept*), *category*, *path*, or *seealso* to semantic representation of concepts.

On the other hand, for any semantic information of concepts, we may obtain its value from multiple knowledge sources such as *WordNet*, domain ontologies (e.g., *MeSH* [44] or *SNOMED CT* [40]), *Wikipedia*, *DBpedia*, or *YAGO*. It is worth noting that the types of the values of different semantic information may be different, for instance, the types of the values of synonyms, glosses, or taxonomy structure are set, string, or tree (graph), respectively. Clearly, for some semantic information, its values from multiple knowledge sources can be integrated (merged). For example, the values of synonyms from different knowledge sources can be combined by using operation union in set theory, and the values of glosses from multiple knowledge sources may be merged by using operation concatenation of string. We

call such semantic information as operable (denoted by $\oplus$, see Definition 2). Of course, some semantic information such as taxonomy structure is inoperable.

For the sake of convenience, we use string to represent the types of values of all semantic information. Our notation for the encoding of the value v of semantic information into its representation as a string is $\langle v \rangle$ such as $\langle TS_{WordNet} \rangle$ and $\langle TS_{Wikipedia} \rangle$.

Definition 2. Let $\langle SI_1(con), SI_2(con), \ldots, SI_n(con) \rangle$ be the semantic representation of a concept *con*, where $SI_i(con)=\langle \langle KS_{i_1}: value_{i_1} \rangle, \langle KS_{i_2}: value_{i_2} \rangle, \ldots, \langle KS_{i_m}: value_{i_m} \rangle \rangle$. If $SI_i(con)$ is operable, its values $value_{i_1}$, $value_{i_2}$, …, and $value_{i_m}$ from KS_{i_1}, KS_{i_2}, and KS_{i_m} respectively can be merged by the following operator: $value_i=value_{i_1} \oplus value_{i_2} \oplus \ldots \oplus value_{i_m}$, where $\oplus$ denotes integration (or combination) operator of multiple values of same type such as $\cup$ for sets and + for strings.

$SI_i(con)$ is extended as follows:

$SI_i(con)=\langle \langle KS_{i_1}: value_{i_1} \rangle, \langle KS_{i_2}: value_{i_2} \rangle, \ldots, \langle KS_{i_m}: value_{i_m} \rangle, \langle KS_{i_1}, KS_{i_2}, \ldots, KS_{i_m}: value_i \rangle \rangle$.

In fact, for any $\{\langle KS_{i_s}: value_{i_s} \rangle, \ldots, \langle KS_{i_t}: value_{i_t} \rangle\} \subseteq \{\langle KS_{i_1}: value_{i_1} \rangle, \langle KS_{i_2}: value_{i_2} \rangle, \ldots, \langle KS_{i_m}: value_{i_m} \rangle\}$, we may have the following:

$value_i' = value_{i_s} \oplus \ldots \oplus value_{i_t}$,

$SI_i(con)$ can be extended as $SI_i(con)=\langle \langle KS_{i_1}: value_{i_1} \rangle, \ldots, \langle KS_{i_m}: value_{i_m} \rangle, \langle KS_{i_s}, \ldots, KS_{i_t}: value_i' \rangle \rangle$.

Example 3. From Example 2 we know that the glosses of *Artificial Intelligence* can be merged as follows:

$\langle$*WordNet*, …, *Wikipedia*: "*the branch of computer science that deal with writing computer programs that can solve problems creatively*, …"+…+"*Artificial intelligence* (*AI*), *sometimes called machine intelligence*, *is intelligence demonstrated by machines*, …"$\rangle$.

3.2 A framework for semantic similarity computation

Given two concepts A and B, we firstly need to obtain their semantic information in order to compute semantic similarity between them. Clearly, we can get their semantic information from the semantic representation $\langle SI_1(A), SI_2(A), \ldots, SI_n(A) \rangle$ and $\langle SI_1(B), SI_2(B), \ldots, SI_n(B) \rangle$ of A and B, respectively. Because there are lots of semantic information in A and B, we can design different similarity computation methods by using different semantic information. For example, feature-based measures need some features such as glooses, synonyms, hyponyms (sub-concepts), hypernyms (super-concepts), or categories, but IC-based measures need certain taxonomy structure (tree structure or graph structure). To unify these similarity measures (e.g., distance-based, feature-based, or IC-based measures) between two concepts, we need a framework for these semantic similarity measures.

Definition 3. Let $A = \langle SI_1(A), SI_2(A), \ldots, SI_n(A) \rangle$ and $B = \langle SI_1(B), SI_2(B), \ldots, SI_n(B) \rangle$ be semantic representation of two concepts, where $SI_i(A)=\langle \langle KS_{i_1}: value_{i_1} \rangle, \langle KS_{i_2}: value_{i_2} \rangle, \ldots, \langle KS_{i_m}: value_{i_m} \rangle \rangle$ and $SI_i(B)= \langle \langle KS_{i_1}: value_{i_1}' \rangle, \langle KS_{i_2}: value_{i_2}' \rangle, \ldots, \langle KS_{i_m}: value_{i_m}' \rangle \rangle$. The semantic similarity between A and B, denoted as $Sim(A, B)$, is the function $Sim: CON \times CON \rightarrow [0, 1]$, and is defined as follows:

$$Sim(A,B) = Sim_{concepts}(Sim_{SI_1}(ESetSI_1, ESetSI_1{}'), Sim_{SI_2}(ESetSI_2, ESetSI_2{}'), \ldots, Sim_{SI_n}(ESetSI_n, ESetSI_n{}'))$$

where (1) $Sim_{SI_i}(ESetSI_i, ESetSI_i')$ $(1 \le i \le n)$ is the similarity measure of semantic information $SI_i(A)$ and $SI_i(B)$, concretely, Sim_{SI_i} is the function Sim_{SI_i}: $SetSI_i \times SetSI_i' \to [a_i, b_i]$, where a_i, $b_i \in \mathbf{R}^+ \cup \{0\}$, $a_i \le b_i$, $\mathbf{R}^+ \cup \{0\}$ denotes the set of non-negative real numbers.

(2) $Sim_{concepts}$ is the function $Sim_{concepts}$: $[a_1, b_1] \times \ldots \times [a_n, b_n] \to [0, 1]$.

(3) *CON* stands for the set of all concepts, $SetSI_i$ and $SetSI_i'$ denote the set of all values of semantic information $SI_i(A)$ and $SI_i(B)$ respectively, formally, $SetSI_i = \{\langle value_{i_1}\rangle \cup \langle value_{i_2}\rangle \cup \ldots \cup \langle value_{i_m}\rangle\}$ and $SetSI_i' = \{\langle value_{i_1}{}'\rangle \cup \langle value_{i_2}{}'\rangle \cup \ldots \cup \langle value_{i_m}{}'\rangle\}$, $ESetSI_i \in SetSI_i$, and $ESetSI_i' \in SetSI_i'$.

Example 4. Let A and B be two concepts, $A = \langle glosses(A), synonyms(A), taxonomy(A)\rangle$ and $B = \langle glosses(B), synonyms(B), taxonomy(B)\rangle$ be semantic representation of concepts A and B, where

$$\begin{aligned}
glosses(A) &= \left\langle \langle WordNet : g_{WordNet}(A)\rangle, \left\langle Wikipedia : g_{Wikipedia}(A)\right\rangle, \left\langle DBpedia : g_{DBpedia}(A)\right\rangle\right\rangle,\\
glosses(B) &= \left\langle \langle WordNet : g_{WordNet}(B)\rangle, \left\langle Wikipedia : g_{Wikipedia}(B)\right\rangle, \left\langle DBpedia : g_{DBpedia}(B)\right\rangle\right\rangle,\\
synonyms(A) &= \left\langle \langle WordNet : s_{WordNet}(A)\rangle, \left\langle Wikipedia : s_{Wikipedia}(A)\right\rangle, \left\langle DBpedia : s_{DBpedia}(A)\right\rangle\right\rangle,\\
synonyms(B) &= \left\langle \langle WordNet : s_{WordNet}(B)\rangle, \left\langle Wikipedia : s_{Wikipedia}(B)\right\rangle, \left\langle DBpedia : s_{DBpedia}(B)\right\rangle\right\rangle,\\
taxonomy(A) &= \left\langle \langle WordNet : t_{WordNet}(A)\rangle, \left\langle Wikipedia : t_{Wikipedia}(A)\right\rangle, \left\langle DBpedia : t_{DBpedia}(A)\right\rangle\right\rangle, and\\
taxonomy(B) &= \left\langle \langle WordNet : t_{WordNet}(B)\rangle, \left\langle Wikipedia : t_{Wikipedia}(B)\right\rangle, \left\langle DBpedia : t_{DBpedia}(B)\right\rangle\right\rangle
\end{aligned}$$

By Definition 3, we have the following:

$$\begin{aligned}
Sim(A,B) = Sim_{concepts}(&Sim_{glosses}\left(g_{WordNet}(A) \cup g_{Wikipedia}(A) \cup g_{DBpedia}(A), g_{WordNet}(B) \cup g_{Wikipedia}(B) \cup g_{DBpedia}(B)\right),\\
&Sim_{synonyms}\left(s_{WordNet}(A) \cup s_{Wikipedia}(A) \cup s_{DBpedia}(A), s_{WordNet}(B) \cup s_{Wikipedia}(B) \cup s_{DBpedia}(B)\right),\\
&Sim_{taxonomy}\left(t_{WordNet}(A) \cup t_{Wikipedia}(A) \cup t_{DBpedia}(A), t_{WordNet}(B) \cup t_{Wikipedia}(B) \cup t_{DBpedia}(B)\right)).
\end{aligned}$$

From Definition 3 and Example 4 we know that the framework for semantic similarity measures is very generic. For any similarity function Sim_{SI_i}: $SetSI_i \times SetSI_i' \to [a_i, b_i]$, there are many concrete implementation methods. Formally, for any $\{\langle value_{i_s}\rangle, \ldots, \langle value_{i_t}\rangle\} \subseteq \{\langle value_{i_1}\rangle, \langle value_{i_2}\rangle, \ldots, \langle value_{i_m}\rangle\}$, we can define a similarity function as follows from the perspective of knowledge sources:

$$Sim_{SI_i}: \{< value_{i_s} > \cup \ldots \cup < value_{it} >\} \times \{< value_{i_s}{}' > \cup \ldots \cup < value_{i_t}{}' >\} \to [a_i, b_i]$$

For example, in Example 4 part of the definitions of function $Sim_{glosses}$ can be defined as follows:

$$\begin{aligned}
Sim_{glosses}: g_{WordNet}(A) \times g_{WordNet}(B) &\to [a,b], g_{WordNet}(A)\\
&\times g_{Wikipedia}(B) \to [a,b], \text{ or } g_{WordNet}(A) \cup g_{Wikipedia}(A)\\
&\times g_{WordNet}(B) \cup g_{Wikipedia}(B) \to [a,b].
\end{aligned}$$

From the perspective of mathematical tools of semantic similarity measures, we may use different mathematical tools such as IC [63], PMI [14], LSA [19], ESA [23], or Jaccard and Dice coefficients [45] for $sim_{SI_i}(ESetSI_i, ESetSI_i')$ $(1 \le i \le n)$ in Definition 3. For instance, we can define $Sim_{glosses}$ and $Sim_{synonyms}$ using ESA, Jaccard or Dice coefficients, and define $Sim_{taxonomy}$ using IC.

Lastly, the function $Sim_{concepts}$ in Definition 3 is also very flexible. Generally speaking, we may implement $Sim_{concepts}$ by introducing some simple functions such as max, min, or average.

Now we give the implementation method of the framework for semantic similarity measures.

Algorithm 1. Implementation of the framework for similarity measures

Input: Two concepts A and B

Output: Semantic similarity $Sim(A, B)$ between A and B

(1) Specify the set $\{SI_1, SI_2, \ldots, SI_n\}$ of titles of semantic information of A and B.

(2) Specify the set $\{KS_1, KS_2, \ldots, KS_m\}$ of knowledge sources.

(3) For each $SI_i \in \{SI_1, SI_2, \ldots, SI_n\}$, obtain the value $SI_i(A)$ and $SI_i(B)$ of semantic information of A and B from all knowledge sources $\{KS_1, KS_2, \ldots, KS_m\}$, respectively:

$$SI_i(A)=\langle\langle KS_1: value_{i_1}\rangle, \langle KS_2: value_{i_2}\rangle, \ldots, \langle KS_m: value_{i_m}\rangle\rangle \text{ and}$$
$$SI_i(B)=\langle\langle KS_1: value_{i_1}'\rangle, \langle KS_2: value_{i_2}'\rangle, \ldots, \langle KS_m: value_{i_m}'\rangle\rangle.$$

(4) Let $SetSI_i(A)=\{\langle value_{i_1}\rangle \cup \langle value_{i_2}\rangle \cup \ldots \cup \langle value_{i_m}\rangle\}$ and $SetSI_i(B)=\{\langle value_{i_1}'\rangle \cup \langle value_{i_2}'\rangle \cup \ldots \cup \langle value_{i_m}'\rangle\}$.

(5) Determine the similarity function $Sim_{SI_i}: SSetSI_i(A) \times SSetSI_i(B) \rightarrow [a_i, b_i]$ using mathematical tools, where $SSetSI_i(A)=\{\langle value_{i_s}\rangle \cup \ldots \cup \langle value_{i_t}\rangle\} \subseteq SetSI_i(A)$, $SSetSI_i(B)=\{\langle value_{i_s}'\rangle \cup \ldots \cup \langle value_{i_t}'\rangle\} \subseteq SetSI_i(B)$.

(6) For each $SI_i \in \{SI_1, SI_2, \ldots, SI_n\}$, get the value $\alpha_i \in [a_i, b_i]$ of $Sim_{SI_i}(ESetSI_i(A), ESetSI_i(B))$, where $ESetSI_i(A) \in SSetSI_i(A)$ and $ESetSI_i(B) \in SSetSI_i(B)$.

(7) Determine the function $Sim_{concepts}([a_1, b_1], \ldots, [a_n, b_n])$.

(8) Get the value $Sim(A, B)$ of $Sim_{concepts}(\alpha_1, \ldots, \alpha_n)$.

(9) Return $Sim(A, B)$.

(10) End.

Remark 2. In Algorithm 1, the sets $\{SI_1, SI_2, \ldots, SI_n\}$ and $\{KS_1, KS_2, \ldots, KS_m\}$ can be specified by users or experts. The value of $SI_i(A)$ and $SI_i(B)$ may be obtained from knowledge sources automatically. In fact, we may obtain the values of $SI_i(A)$ and $SI_i(B)$ offline. If we cannot get $\langle KS: value_{i_j}\rangle$ (resp., $\langle KS: value_{i_j}'\rangle$) of $SI_i(A)$ (resp., $SI_i(B)$), we may assign $\langle KS: value_{i_j}\rangle=\phi$ (resp., $\langle KS: value_{i_j}'\rangle=\phi$). In Step (5) of Algorithm 1, we can assign lots of similarity functions for each $SI \in \{SI_1, SI_2, \ldots, SI_n\}$ in theory. However, we can selectively set up similarity functions according to the complementarity of knowledge sources in practical applications.

For example, let us consider knowledge sources $\{WordNet, Wikipedia, MeSH\}$. It is well known that WordNet is a large lexical database, Wikipedia is a free online encyclopedia, and

MeSH is a hierarchically-organized terminology for indexing and cataloging of biomedical information. Clearly, these are three complementary knowledge sources. If we only consider semantic information *glosses* and *taxonomy* (see Example 4), we may set up the following similarity functions:

$$\begin{aligned} Sim_{glosses} &: glosses_{WordNet}(A) \times glosses_{WordNet}(B) \rightarrow [a,b], Sim_{glosses} \\ &: glosses_{Wikipedia}(A) \times glosses_{Wikipedia}(B) \rightarrow [a,b], Sim_{glosses} \\ &: glosses_{WordNet}(A) \cup glosses_{Wikipedia}(A) \\ &\quad \times glosses_{WordNet}(B) \cup glosses_{Wikipedia}(B) \rightarrow [a,b], Sim_{taxonomy} \\ &: taxonomy_{WordNet}(A) \times taxonomy_{WordNet}(B) \rightarrow [a,b], Sim_{taxonomy} \\ &: taxonomy_{MeSH}(A) \times taxonomy_{MeSH}(B) \rightarrow [a,b], \text{and} thtrue in Sim_{taxonomy} \\ &: taxonomy_{Wikipedia}(A) \times taxonomy_{Wikipedia}(B) \rightarrow [a,b]. \end{aligned}$$

If $A \in Wikipedia$, $A \notin WordNet$, $A \notin MeSH$, $B \in MeSH$, $B \notin WordNet$, and $B \notin Wikipedia$, then we also can give the similarity functions as follows:

$$Sim_{taxonomy} : taxonomy_{Wikipedia}(A) \times taxonomy_{MeSH}(B) \rightarrow [a,b].$$

Obviously, all existing methods of similarity computation can be obtained by instantiating the framework (Definition 3), that is, all existing approaches to similarity measures (including distance-based measures, feature-based measures, statistical measures, and hybrid measures, see Section 2 for more details) can result from instantiations of the framework. Concretely, existing methods to similarity measures consider only one knowledge source such as WordNet, Wikipedia, domain ontology, or DBpedia, thus, in Step (5) of Algorithm 1 there is only one kind of similarity function for each $SI_i \in \{SI_1, SI_2, \ldots, SI_n\}$. Clearly, in addition to the existing similarity computation methods, we can get a lot of new similarity measure methods by instantiating the framework, in particular, we may obtain some new approaches to similarity measures that existing methods cannot deal with by introducing multiple knowledge sources.

4 Some approaches for measuring semantic similarity

In Section 3 our framework for semantic similarity of concepts is proposed. In this section we give some generic and flexible approaches to similarity measures by

instantiating the framework. As stated in Section 3, all existing approaches can result from instantiations of our framework, the instantiation method is as follows:

Algorithm 2. Obtain existing approaches by instantiating the framework

Input: Two concepts A and B

Output: Semantic similarity $Sim(A, B)$ between A and B

(1) Specify the set $\{SI_1, SI_2, \ldots, SI_n\}$ of titles of semantic information of A and B.

(2) Specify a knowledge source KS.

(3) For each $SI_i \in \{SI_1, SI_2, \ldots, SI_n\}$, obtain the value $SI_i(A)$ and $SI_i(B)$ of semantic information of A and B from KS: $SI_i(A)=\langle KS: value_i\rangle$ and $SI_i(B)=\langle KS: value_i'\rangle$.

(4) Determine the similarity function $Sim_{SI_i}: \{\langle value_i\rangle\}\times\{\langle value_i'\rangle\} \rightarrow [a_i, b_i]$ using mathematical tools.

(5) For each $SI_i \in \{SI_1, SI_2, \ldots, SI_n\}$, get the value $\alpha_i \in [a_i, b_i]$ of $Sim_{SI_i}(ESI_i(A), ESI_i(B))$, where $ESI_i(A) \in \{\langle value_i\rangle\}$ and $ESI_i(B) \in \{\langle value_i'\rangle\}$.

(6) Determine the function $Sim_{concepts}([a_1, b_1], \ldots, [a_n, b_n])$.

(7) Get the value $Sim(A, B)$ of $Sim_{concepts}(\alpha_1, \ldots, \alpha_n)$.

(8) Return $Sim(A, B)$.

(9) End.

In what follows, we present some new similarity measures that existing methods cannot deal with by instantiating the framework. Similarly to existing similarity measures, we also give three families of similarity measure methods: (1) IC-based similarity measures; (2) distance-based similarity measures; and (3) feature-based similarity measures. Based on these three similarity measure families, we will naturally get hybrid similarity measures.

4.1 IC-based measures under multiple knowledge sources

In the framework in Definition 3 or Algorithm 1, to implement IC-based similarity measures, we need one or multiple taxonomy structures (tree structures or graph structures). Suppose that A and B are two concepts, $KS_1, KS_2, \ldots, KS_m$ are knowledge sources, and $T_1, T_2, \ldots, T_m$ are taxonomy structures in $KS_1, KS_2, \ldots, KS_m$, respectively.

If there exists a taxonomy structure T_i $(1 \le i \le m)$ such that $A, B \in T_i$, it is easy to get the LCS (least common subsumber) for A and B in T_i. Furthermore, we can compute $Sim(A, B)$ by using IC-based similarity measure methods (see Section 2.3). However, if there does not exist any taxonomy structure T_i $(1 \le i \le m)$ such that $A, B \in T_i$, that is, for any taxonomy structure T_i $(1 \le i \le m)$, either $A \in T_i$, $B \notin T_i$, or $A \notin T_i$, $B \in T_i$, how should we compute $Sim(A, B)$ by using IC-based measures at this time (or how to find the LCS for A and B by using $KS_1, \ldots, KS_m$)? To solve this problem, we propose some new IC-based similarity measures for concepts.

Without loss of generality, suppose that all knowledge sources that we consider are the set $AllKS = \{KS_1, KS_2, \ldots, KS_m\}$, and there exist some knowledge sources $KSA = \{KS_k, KS_{k+1}, \ldots, KS_l\} \subseteq AllKS$ and $KSB = \{KS_s, KS_{s+1}, \ldots, KS_t\} \subseteq AllKS$ such that for any $KS_i \in KSA$ and $KS_j \in KSB$ we have the following:

$A \in T_i$, $A \notin T_j$, $B \in T_j$, $B \notin T_i$, where $T_1, T_2, \ldots, T_m$ are taxonomy structures of $KS_1, KS_2, \ldots, KS_m$, respectively.

Obviously, there is no LCS for A and B in T_i (or T_j), thus, we cannot compute $Sim(A, B)$ only by considering T_i (or T_j). Now we give some methods for $Sim(A, B)$ by considering both T_i and T_j.

Definition 4. Let T be a taxonomy structure and concept subsumption ($<_T$) be a binary relation $<_T$: $CON \times CON$, being CON the set of all concepts, where $A <_T C$ means that A is a subconcept of C or C is a parent concept of A in T. $A <_T C$ iff $C >_T A$, that is, $A >_T C$ means that A is a parent concept of C or C is a subconcept of A in T. $A \leq_T C$ iff $A <_T C$ or $A = C$ (i.e., A and C are two identical concepts). $A \geq_T C$ iff $A >_T C$ or $A = C$. We define the set of *subconcepts*, *superconcepts*, *hyponyms*, and *hypernyms* of a concept $A \in CON$ w.r.t T as follows:

$$subconcepts(A, T) = \{C \in CON \mid C <_T A\};$$

$$superconcepts(A, T) = \{C \in CON \mid C >_T A\};$$

$$hyponyms(A, T) = \{C \in CON \mid \exists C_1, C_2, \ldots, C_{n-1}, C_n \in CON \wedge n \geq 2 \wedge C_1 = A \wedge C_n = C \wedge C_1 >_T C_2 \wedge \ldots \wedge C_{n-1} >_T C_n \wedge C_1 \neq C_2 \neq \ldots \neq C_{n-1} \neq C_n\};$$

$$hypernyms(A, T) = \{C \in CON \mid \exists C_1, C_2, \ldots, C_{n-1}, C_n \in CON \wedge n \geq 2 \wedge C_1 = A \wedge C_n = C \wedge C_1 <_T C_2 \wedge \ldots \wedge C_{n-1} <_T C_n \wedge C_1 \neq C_2 \neq \ldots \neq C_{n-1} \neq C_n\}.$$

Clearly, we have that $subconcepts(A, T) \subseteq hyponyms(A, T)$ and $superconcepts(A, T) \subseteq hypernyms(A, T)$.

Definition 5. Let $A, B \in CON$ be two different concepts (i.e., $A \neq B$) and T be a taxonomy structure. The set of walks between A and B w.r.t. T can be defined as follows:

$$walks(A, B, T) = \{\langle C_1, C_2, \ldots, C_n\rangle \mid C_1, C_2, \ldots, C_n \in CON \wedge C_1 = A \wedge C_n = B \wedge (\forall 1 \leq i < n, C_i \in superconcepts(C_{i+1}, T)) \wedge C_1 \neq C_2 \neq \ldots \neq C_{n-1} \neq C_n\}.$$

Definition 6. Let T_i and T_j be two taxonomy structures, $A, B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The set of *common ancestors* of A and B w.r.t. T_i and T_j is defined as follows:

$$CommonAnc(A, B, T_i, T_j) = \{C \in CON \mid C \in hypernyms(A, T_i) \wedge C \in hypernyms(B, T_j)\}.$$

Definition 7. Let T_i and T_j be two taxonomy structures, $A, B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The set of *GCS* (Good Common Subsumer) of A and B w.r.t. T_i and T_j can be defined as follows:

$$GCS(A, B, T_i, T_j) = \{C \in CON \mid C \in CommonAnc(A, B, T_i, T_j) \wedge p_1 \in walks(C, A, T_i), p_2 \in walks(C, B, T_j), |p_1| + |p_2| = \min_{D \in CommonAnc(A,B,T_i,T_j),\ p' \in walks(D,A,T_i),\ p'' \in walks(D,B,T_j)} \{|p'| + |p''|\}\},$$

where $|p|$ is the length of walk p, i.e., if $p = \langle c_1, c_2, \ldots, c_{n+1}\rangle$, then $|p| = |\langle c_1, c_2, \ldots, c_{n+1}\rangle| = n$.

Based on the GCS for two concepts in two taxonomy structures (Definition 7), we can present some new IC-based measures under multiple knowledge sources by extending traditional IC-based similarity measures (see Section 2.3) [35, 41, 42, 63, 72]. To compute semantic similarity of two concepts A and B using IC-based measures, we firstly need to give some approaches to IC computation for concepts.

Definition 8. Let $A \in CON$ be a concept and T be a taxonomy structure. The first IC of A w.r.t. T is defined as follows:

$$IC_{fir}(A,T) = 1 - \frac{\log\left(hyponyms(A,T)+1\right)}{\log\left(|CON_T|\right)},$$

where CON_T denotes the set of all concepts in T.

In fact, $IC_{fir}(A, T)$ is an extension of the IC model of Seco et al. [75].

Definition 9. Let $A \in CON$ be a concept and T be a taxonomy structure. The depth $depth(A, T)$ of A in T is defined as follows:

$$depth(A,T) = \max\{|p \mid p \in walks(root(T),A,T)\}, \text{where } root(T) \text{ is the root of } T$$

Definition 10. Let $A \in CON$ be a concept and T be a taxonomy structure. The set of *leaves* of A in T is defined as follows:

$$leaves(A,T) = \{C \in CON \mid C \in hyponyms(A,T) \wedge hyponyms(C,T) = \phi\}$$

Furthermore, we define the following:

$$\begin{aligned} maxleaves(T) &= leaves(root(T),T) and\ maxdepth(T) \\ &= \max|p\|p \in walks(root(T),A,T), A \in maxleaves(T). \end{aligned}$$

By extending the IC definitions of Zhou et al. [83] and Sanchez et al. [72], we can propose the following approaches to IC computation.

Definition 11. Let $A \in CON$ be a concept and T be a taxonomy structure. The second and third ICs of A w.r.t. T are defined as follows:

$$IC_{sec}(A,T) = \gamma\left(1 - \frac{\log\left(hyponyms(A,T)+1\right)}{\log\left(|CON_T|\right)}\right) + (1-\gamma)\left(\frac{\log\left(depth(A,T)+1\right)}{\log\left(maxdepth(T)\right)}\right),$$

$$IC_{thi}(A,T) = -\log\left(\frac{\frac{|leaves(A,T)|}{|hypernyms(A,T)\cup\{A\}|}+1}{|maxleaves(T)|+1}\right)$$

where γ is a tuning factor that adjusts the weight of the two features involved in the IC computation. We use $\gamma = 0.5$ in default.

Now we propose some new approaches to semantic similarity measures for concepts under multiple knowledge sources by using GCS (Definition 7) and IC (Definitions 8 and 11). It is worth noting that we can obtain lots of new IC-based measures by extending traditional IC-based similarity measures. In this paper we only extend some classical IC-based measures.

Definition 12. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The IC-based semantic similarity $SimIC1_{ord}$ between A and B w.r.t. T_i and T_j can be defined as:

$$SimIC1_{ord}(A,B,Ti,Tj) = \max_{C\in GCS(A,B,T_i,T_j)} \max IC_{ord}(C,T_i), IC_{ord}(C,T_j),$$

where $IC_{ord} = IC_{fir}$, IC_{sec}, or IC_{thi}. For example, if $IC_{ord} = IC_{fir}$, $SimIC1_{ord}$ means $SimIC1_{fir}$.

Clearly, $SimIC_{ord}$ is an extension of Resnik's metric [63].

By extending the Lin's metric [42], we can present another similarity measure for concepts.

Definition 13. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The IC-based semantic similarity $SimIC2_{ord}$ between A and B w.r.t. T_i and T_j can be defined as:

$$SimIC2_{ord}(A,B,T_i,T_j) = \max_{C\in GCS(A,B,T_i,T_j)} \left\{ \frac{2\times \max\{IC_{ord}(C,T_i), IC_{ord}(C,T_j)\}}{IC_{ord}(A,T_i) + IC_{ord}(B,T_j)} \right\},$$

where $IC_{ord} = IC_{fir}$, IC_{sec}, or IC_{thi}.

Obviously, we also can define a kind of similarity measure $SimIC3_{ord}$ by extending the Jiang and Conrath's metric [35].

Definition 14. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The IC-based semantic similarity $SimIC3_{ord}$ between A and B w.r.t. T_i and T_j can be defined as:

$$SimIC3_{ord}(A,B,T_i,T_j) = 1 - \frac{Distance(A,B,T_i,T_j)}{2},$$

where $Distance(A, B, T_i, T_j) =$.

$$IC_{ord}(A,T_i) + IC_{ord}(B,T_j) - 2\times \max_{C\in GCS(A,B,T_i,T_j)} \{\max\{IC_{ord}(C,T_i), IC_{ord}(C,T_j)\}\},$$
$$IC_{ord} = IC_{fir}, IC_{sec}, \text{or } IC_{thi}.$$

From Definitions 12-14 we know that $SimIC1_{ord}$, $SimIC2_{ord}$, and $SimIC3_{ord}$ are based on two knowledge sources. In fact, we need multiple knowledge sources in practical applications in order to obtain better results. Therefore, we have to give some similarity measures for multiple knowledge sources.

Definition 15. Let $AllTS = \{T_1, T_2, \ldots, T_m\}$ be all taxonomy structures, $TSA = \{T_k, T_{k+1}, \ldots, T_l\} \subseteq AllTS$ and $TSB = \{T_s, T_{s+1}, \ldots, T_t\} \subseteq AllTS$. For any $T_i \in TSA$ and $T_j \in TSB$, we have that $A \in T_i$, $A \notin T_j$, $B \in T_j$, $B \notin T_i$. The IC-based semantic similarity measures $SimIC1M_{ord}$, $SimIC2M_{ord}$, and $SimIC3M_{ord}$ between A and B w.r.t. multiple taxonomy structures TSA and TSB can be defined as:

$$SimIC1M_{ord}(A,B,TSA,TSB) = \max_{i=k}^{l} \max_{j=s}^{t} \{SimIC1_{ord}(A,B,T_i,T_j)\},$$
$$SimIC2M_{ord}(A,B,TSA,TSB) = \max_{i=k}^{l} \max_{j=s}^{t} \{SimIC2_{ord}(A,B,T_i,T_j)\}, \text{and.}$$
$$SimIC3M_{ord}(A,B,TSA,TSB) = \max_{i=k}^{l} \max_{j=s}^{t} \{SimIC3_{ord}(A,B,T_i,T_j)\}.$$

The IC-based semantic similarity measure $SimIC$ between A and B w.r.t. TSA and TSB and all baseline measures can be defined as:

$$SimIC(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimIC1_{ord}(A,B,T_i,T_j), SimIC2_{ord}(A,B,T_i,T_j), SimIC3_{ord}(A,B,T_i,T_j)\}.$$

In order to compare the values of different similarities $SimIC1_{ord}$, $SimIC2_{ord}$, and $SimIC3_{ord}$, we normalize the value of each similarity.

Remark 3. In Definition 15, $SimIC1M_{ord}$, $SimIC2M_{ord}$, and $SimIC3M_{ord}$ are extensions of $SimIC1_{ord}$, $SimIC2_{ord}$, and $SimIC3_{ord}$, respectively. That is, $SimIC1_{ord}$, $SimIC2_{ord}$, and $SimIC3_{ord}$ are based on two taxonomy structures, and $SimIC1M_{ord}$, $SimIC2M_{ord}$, and $SimIC3M_{ord}$ are based on multiple taxonomy structures.

On the other hand, if we give some new IC computation approaches (e.g., IC_{fou}), $SimIC1_{ord}$, $SimIC2_{ord}$, and $SimIC3_{ord}$ can be expanded accordingly (e.g., $SimIC1_{fou}$, $SimIC2_{fou}$, $SimIC3_{fou}$). Furthermore, $SimIC1M_{ord}$, $SimIC2M_{ord}$, and $SimIC3M_{ord}$ also can be expanded accordingly (e.g., $SimIC1M_{fou}$, $SimIC2M_{fou}$, $SimIC3M_{fou}$). Obviously, if we consider other baseline measures, we also can obtain some new similarity measures such as $SimIC4_{ord}$ and $SimIC4M_{ord}$ by instantiating our framework.

The similarity measure $SimIC$ can be based on multiple taxonomy structures and baseline measures, clearly, it is easy to extend $SimIC$ when we add new similarity measures for two or multiple taxonomy structures. For example, if a new measure $SimIC4_{ord}$ is provided, $SimIC$ can be expanded as follows:

$$SimIC(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimIC1_{ord}(A,B,T_i,T_j) \mid SimIC2_{ord}(A,B,T_i,T_j), SimIC3_{ord}(A,B,T_i,T_j), SimIC4_{ord}(A,B,T_i,T_j)\}.$$

Lastly, it is worth noting that the condition of Definition 15 can be relaxed as follows:

Let $AllTS = \{T_1, T_2, \ldots, T_m\}$ be all taxonomy structures, $TSA = \{T_k, T_{k+1}, \ldots, T_l\} \subseteq AllTS$ and $TSB = \{T_s, T_{s+1}, \ldots, T_t\} \subseteq AllTS$. For any $T_i \in TSA$ and $T_j \in TSB$, we have that $A \in T_i$ and $B \in T_j$.

If $TSA \cap TSB \neq \phi$, traditional IC-based measures under one taxonomy structure are included in this framework of Definition 15. For example, if $T_u \in TSA \cap TSB$, $SimICN_{ord}(A, B, T_u, T_u)$ ($N = 1, 2, 3$) is based on one taxonomy structure.

The relationships among all definitions of IC-based measures under multiple knowledge sources are shown as Fig. 5.

4.2 Distance-based measures under multiple knowledge sources

Similarly to IC-based measures under multiple knowledge sources (see Section 4.1), in the framework in Definition 3 or Algorithm 1, we also need one or multiple taxonomy structures (tree structures or graph structures) in order to implement distance-based similarity measures. Assume that A and B are two concepts, KS_1, KS_2, …, KS_m are knowledge sources, and T_1, T_2, …, T_m are taxonomy structures in KS_1, KS_2, …, KS_m, respectively. Clearly, if there exists a taxonomy structure T_i ($1 \le i \le m$) such that A, $B \in T_i$, it is easy to compute $Sim(A, B)$ by using distance-based similarity measures (see Section 2.1). However, if there does not exist any taxonomy structure T_i ($1 \le i \le m$) such that A, $B \in T_i$, we need some new distance-based similarity measures.

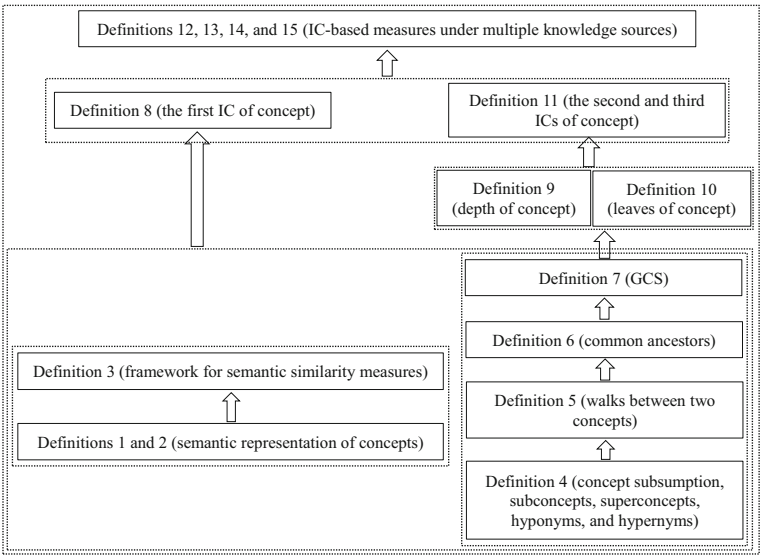

Fig. 5 The relationships among all definitions of IC-based measures

Let all knowledge sources be the set $AllKS = \{KS_1, KS_2, \ldots, KS_m\}$. Suppose that $KSA = \{KS_k, KS_{k+1}, \ldots, KS_l\} \subseteq AllKS$ and $KSB = \{KS_s, KS_{s+1}, \ldots, KS_t\} \subseteq AllKS$, and for any $KS_i \in KSA$ and $KS_j \in KSB$ we have that $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_j$. Obviously, there is no a path between A and B in T_i (or T_j), thus, we cannot compute $Sim(A, B)$ only by considering T_i (or T_j). Now we give some methods for $sim(A, B)$ by considering both T_i and T_j.

Assume that part of T_i and T_j are shown as Fig. 6.

Obviously, if there exists a concept C, such that $C \in T_i$, $C \in T_j$, C is a super-concept of A in T_i, and C is also a super-concept of B in T_j (see Fig. 6), we can find a path between A and B in T_i and T_j, formally, the path is made up of two paths $A \rightarrow C$ (the bold solid line in T_i) and $B \rightarrow C$ (the bold solid line in T_j), that is, there are four edges between A and B in this path.

Similarly, if there exists a concept D, such that $D \in T_i$, $D \in T_j$, D is a sub-concept of A in T_i, and D is also a sub-concept of B in T_j (see Fig. 6), we also may find another path between A and B in T_i and T_j, formally, the path is made up of two paths $A \rightarrow D$ (the bold dotted line in T_i) and $B \rightarrow D$ (the bold dotted line in T_j), that is, there are five edges between A and B in this new path.

Furthermore, we can compute $Sim(A, B)$ by making use of these paths. It's obvious that we meet a problem here: How to obtain the common super-concept or common sub-concept that we need such as C and D in Fig. 6? Because there may be multiple common super-concepts or common sub-concepts, for example, both C and E are super-concepts of A (resp., B) in T_i (resp., T_j), and both D and F are sub-concepts of A (or B) in T_i (or T_j) in Fig. 6. Clearly, we need to find the shortest path between concepts A and B in two taxonomy structures. To get the shortest path, we firstly introduce some notions.

Definition 16. Let T be a taxonomy structure (directed graph) and concept reachability ($\rightarrow_T$) be a binary relation $\rightarrow_T$: $CON \times CON$, being CON the set of all concepts, where $A \rightarrow_T C$ means that there is an edge e from A to C, that is, e is associated with the ordered pair (A, C) in T. $A \leftarrow_T C$ iff $C \rightarrow_T A$, that is, $A \leftarrow_T C$ means that there is an edge which is

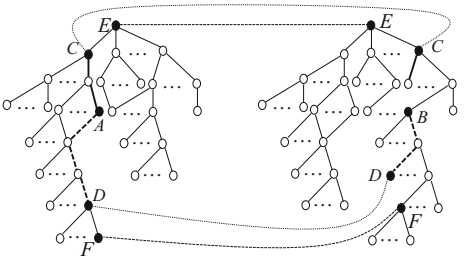

Fig. 6 Taxonomy structures T_i and T_j

Taxonomy structure T_i Taxonomy structure T_j

associated with the ordered pair (C, A) in T. We define the set of *relatedconcepts* of a concept $A \in CON$ w.r.t T as follows:

$$relatedconcepts(A, T) = \{C \in CON \mid \exists C_1, C_2, \ldots, C_{n-1}, C_n \in CON \wedge n \geq 2 \wedge C_1 = A \wedge C_n = C \wedge ((C_1 \rightarrow_T C_2 \wedge \ldots \wedge C_{n-1} \rightarrow_T C_n) \vee (C_1 \leftarrow_T C_2 \wedge \ldots \wedge C_{n-1} \leftarrow_T C_n)) \wedge C_1 \neq C_2 \neq \ldots \neq C_{n-1} \neq C_n\}.$$

Definition 17. Let T_i and T_j be two taxonomy structures, $A, B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The set of *common concepts* of A and B w.r.t. T_i and T_j is defined as follows, respectively:

$$CommonCon(A, B, T_i, T_j) = \{C \in CON \mid C \in relatedconcepts(A, T_i) \wedge C \in relatedconcepts(B, T_j)\}.$$

Definition 18. Let $A, B \in CON$ be two different concepts (i.e., $A \neq B$) and T be a taxonomy structure. The set of paths between A and B w.r.t. T can be defined as follows:

$$paths(A, B, T) = \{\langle C_1, C_2, \ldots, C_n \rangle \mid C_1, C_2, \ldots, C_n \in CON \wedge C_1 = A \wedge C_n = B \wedge ((\forall 1 \leq i < n, C_i \rightarrow_T C_{i+1}) \vee (\forall 1 \leq i < n, C_i \leftarrow_T C_{i+1})) \wedge C_1 \neq C_2 \neq \ldots \neq C_{n-1} \neq C_n\}.$$

Now we can give the shortest and longest paths between concepts A and B in two taxonomy structures.

Definition 19. Let T_i and T_j be two taxonomy structures, $A, B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The sets of the shortest paths *spaths* and the longest paths *lpaths* between A and B w.r.t. T_i and T_j can be defined as follows:

$$spaths(A, B, T_i, T_j) = \{\langle C_1, C_2, \ldots, C_n \rangle \mid C_1, C_2, \ldots, C_n \in CON \wedge C_1 = A \wedge C_n = B \wedge \exists C \in CommonCon(A, B, T_i, T_j), p_1 \in paths(C_1, C, T_i), p_2 \in paths(C_n, C, T_j), |p_1| + |p_2| = \min_{D \in CommonCon(A,B,T_i,T_j), p' \in paths(A,D,T_i), p'' \in paths(B,D,T_j)} \{|p'| + |p''|\}\},$$

$$lpaths(A, B, T_i, T_j) = \{\langle C_1, C_2, \ldots, C_n \rangle \mid C_1, C_2, \ldots, C_n \in CON \wedge C_1 = A \wedge C_n = B \wedge \exists C \in CommonCon(A, B, T_i, T_j), p_1 \in paths(C_1, C, T_i), p_2 \in paths(C_n, C, T_j), |p_1| + |p_2| =$$

$\{|p'| + |p''|\}\}$. $\max_{D \in CommonCon(A,B,T_i,T_j), p' \in paths(A,D,T_i), p'' \in paths(B,D,T_j)}$

Furthermore, we define the longest paths w.r.t. T_i and T_j as follows:
$maxdistance(T_i, T_j) = \max\{|p| \mid p \in lpaths(A, B, T_i, T_j), \forall A \in T_i, \forall B \in T_j\}$.
where $|p|$ is the length of path p, i.e., if $p = \langle c_1, c_2, \ldots, c_{n+1}\rangle$, then $|p| = |\langle c_1, c_2, \ldots, c_{n+1}\rangle| = n$.

Based on the shortest path between two concepts in two taxonomy structures (Definition 19), we can present some new distance-based measures under multiple knowledge sources by extending traditional distance-based similarity measures (see Section 2.1) [25, 39, 41, 62, 82].

Definition 20. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The distance-based semantic similarity *SimDis1* between A and B w.r.t. T_i and T_j can be defined as:

$$SimDis1(A, B, T_i, T_j) = 2 \times maxdistance(T_i, T_j) - |p|,$$

where $p \in spaths(A, B, T_i, T_j)$.

Clearly, *SimDis1* is an extension of the metric of Rada et al. [62].

Definition 21. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The distance-based semantic similarity *SimDis2* between A and B w.r.t. T_i and T_j can be defined as:

$$SimDis2(A, B, Ti, Tj) = \frac{2 \times N_3(A, B, T_i, T_j)}{N_1(A, T_i) + N_2(B, T_j) + 2 \times N_3(A, B, T_i, T_j)},$$

where

$N_1(A, T_i)$ $\max\{|p| \mid p \in walks(C, A, T_i), C \in GCS(A, B, T_i, T_j)\}$,
$N_2(B, T_j)$ $\max\{|p| \mid p \in walks(C, B, T_j), C \in GCS(A, B, T_i, T_j)\}$,
$N_3(A, B, T_i, T_j)$ $\max\{|p| \mid p \in walks(root(T_i), C, T_i) \vee p \in walks(root(T_j), C, T_j), C \in GCS(A, B, T_i, T_j)\}$.

Similarly to the Wu and Palmer' metric [82], *SimDis2* (Definition 21) is based on is-a hierarchies, where *walks* and *GCS* are defined in Section 4.1 (Definitions 5 and 7). Obviously, *SimDis2* is an extension of the Wu and Palmer' metric [82].

We can define the following similarity measure *SimDis3* by extending the Leacock and Chodorow's metric [39].

Definition 22. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The distance-based semantic similarity *SimDis3* between A and B w.r.t. T_i and T_j can be defined as:

$$SimDis3(A, B, Ti, Tj) = -\log\left(\frac{|p|}{2 \times \max\{|p_1|, |p_2|\}}\right),$$

where $p \in spaths(A, B, T_i, T_j)$, $p_1 \in lrpaths(T_i)$, $p_2 \in lrpaths(T_j)$,

$$lrpaths(Ti) = \left\{ p | p \in paths(root(Ti), C, Ti), C \in Ti, |p| \max_{D \in T_i} \left\{ |p'| |p' \in paths\left(root(T_i), D, T_i\right) \right\} \right\},$$

$$lrpaths(Tj) = \left\{ p | p \in paths(root(Tj), C, Tj), C \in Tj, |p| = \max_{D \in T_j} \left\{ |p'| |p' \in paths\left(root(T_j), D, T_j\right) \right\} \right\}.$$

Similarly to the metric of Garla and Brandt [25], we also can normalize *SimDis3* to the unit interval as follows.

Definition 23. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The distance-based semantic similarity *SimDis4* between A and B w.r.t. T_i and T_j can be defined as:

$$SimDis4\left(A, B, T_i, T_j\right) = 1 - \frac{\log\left(|p|\right)}{\log\left(2 \times \max|p_1|, |p_2|\right),}$$

where $p \in spaths(A, B, T_i, T_j)$, $p_1 \in lrpaths(T_i)$, and $p_2 \in lrpaths(T_j)$.

Obviously, we can define a kind of similarity measure *SimDis5* by extending the metric of Li et al. [41].

Definition 24. Let T_i and T_j be two taxonomy structures, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in T_i$, $A \notin T_j$, $B \in T_j$, and $B \notin T_i$. The distance-based semantic similarity *SimDis5* between A and B w.r.t. T_i and T_j can be defined as:

$$SimDis5\left(A, B, T_i, T_j\right) = e^{-\alpha \times |p|} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}},$$

where $p \in spaths(A, B, T_i, T_j)$, $h = \max\{|p| \mid p \in walks(root(T_i), C, T_i) \vee p \in walks(root(T_j), C, T_j), C \in GCS(A, B, T_i, T_j)\}$, $\alpha \geq 0$, and $\beta > 0$. In our experiments, we use the same optimal parameters as in [41], i.e., $\alpha = 0.2$ and $\beta = 0.6$.

In Definitions 20-24, *SimDis1*, *SimDis2*, *SimDis3*, *SimDis4*, and *SimDis5* are based on two knowledge sources. We may give the following similarity measures for multiple knowledge sources.

Definition 25. Let $AllTS = \{T_1, T_2, \ldots, T_m\}$ be all taxonomy structures, $TSA = \{T_k, T_{k+1}, \ldots, T_l\} \subseteq AllTS$ and $TSB = \{T_s, T_{s+1}, \ldots, T_t\} \subseteq AllTS$. For any $T_i \in TSA$ and $T_j \in TSB$, we have that $A \in T_i$, $A \notin T_j$, $B \in T_j$, $B \notin T_i$. The distance-based semantic similarity measures

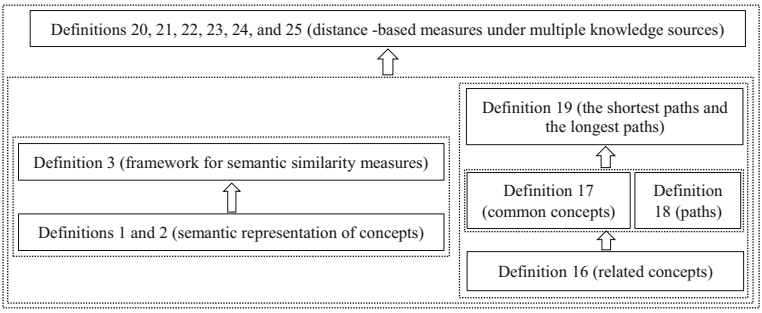

Fig. 7 The relationships among all definitions of distance-based measures

SimDis1M, *SimDis2M*, *SimDis3M*, *SimDis4M*, and *SimDis5M* between A and B w.r.t. multiple taxonomy structures *TSA* and *TSB* can be defined as:

$$\begin{aligned}
&SimDis1M(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimDis1(A,B,T_i,T_j)\},\\
&SimDis2M(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimDis2(A,B,T_i,T_j)\},\\
&SimDis3M(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimDis3(A,B,T_i,T_j)\},\\
&SimDis4M(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimDis4(A,B,T_i,T_j)\} \text{ and}\\
&SimDis5M(A,B,TSA,TSB) = \max_{i=k}^{l}\max_{j=s}^{t}\{SimDis5(A,B,T_i,T_j)\}.
\end{aligned}$$

The distance-based semantic similarity measure *SimDis* between A and B w.r.t. *TSA* and *TSB* and all baseline measures can be defined as:

$$\begin{aligned}
&SimDis(A,B,TSA,TSB)\\
&= \max_{i=k}^{l}\max_{j=s}^{t}\{SimDis1(A,B,T_i,T_j), SimDis2(A,B,T_i,T_j), SimDis3(A,B,T_i,T_j),\\
&\qquad SimDis4(A,B,T_i,T_j), SimDis5(A,B,T_i,T_j)\}.
\end{aligned}$$

In order to compare the values of different similarities *SimDis*1, *SimDis*2, *SimDis*3, *SimDis*4, and *SimDis*5, we also normalize the value of each similarity. Similar to Definition 15 (see Remark 3), the distance-based measure *SimDis* is also a generic and flexible approach.

The relationships among all definitions of distance-based measures under multiple knowledge sources are shown as Fig. 7.

4.3 Feature-based measures under multiple knowledge sources

Unlike IC-based or distance-based similarity measures, feature-based measures assess similarity between concepts as a function of their properties (i.e., features). Therefore, in the framework in Definition 3 or Algorithm 1, for each concept we need one or multiple knowledge sources in order to get its properties (i.e., features). Assume that A and B are two concepts, KS_1, KS_2, …, and KS_m are knowledge sources. Clearly, if there exists a knowledge source KS_i ($1 \le i \le m$) such that the features of A and B can be obtained from KS_i, it is easy to compute $Sim(A, B)$ by using traditional feature-based similarity measures (see Section 2.2). However, if there does not exist any knowledge source KS_i ($1 \le i \le m$) that can provide the features of A and B at the same time, we need some new feature-based similarity measures.

Let all knowledge sources be the set $AllKS = \{KS_1, KS_2, \ldots, KS_m\}$. Suppose that $KSA = \{KS_k, KS_{k+1}, \ldots, KS_l\} \subseteq AllKS$ and $KSB = \{KS_s, KS_{s+1}, \ldots, KS_t\} \subseteq AllKS$, and for any $KS_i \in KSA$ and $KS_j \in KSB$ we have that $A \in KS_i$, $A \notin KS_j$, $B \in KS_j$, and $B \notin KS_i$. Obviously, we cannot compute $Sim(A, B)$ only by considering KS_i (or KS_j). Now we give some methods for $sim(A, B)$ by considering both KS_i and KS_j.

Definition 26. Let KS_i and KS_j be two knowledge sources, A, $B \in CON$ be two different concepts (i.e., $A \neq B$), $A \in KS_i$, $A \notin KS_j$, $B \in KS_j$, and $B \notin KS_i$. Assume that all features that we consider are $\{fea_1, fea_2, \ldots, fea_n\}$, i.e., the semantic representation of A and B is as follows:

$$A = \{fea_1(A), fea_2(A), \ldots, fea_n(A)\} \text{ and } B = \{fea_1(B), fea_2(B), \ldots, fea_n(B)\},$$

where the value of $fea_u(A)$ (resp., $fea_u(B)$) $(1 \le u \le n)$ that comes from KS_i (resp., KS_j) is follows:

$fea_u(A) = \langle KS_i\text{: } value_{i_u} \rangle$ (resp., $fea_u(B) = \langle KS_j\text{: } value_{j_u} \rangle$).

The feature-based semantic similarity framework *SimFea* between A and B w.r.t. KS_i and KS_j can be defined as:

$$SimFea(A, B, KS_i, KS_j) = \max Sim_1\left(value_{i_1}, value_{j_1}\right), \ldots, Sim_n\left(value_{i_n}, value_{j_n}\right)$$

In this paper, we only consider four kinds of features, i.e., glooses, synonyms, hyponyms (or sub-concepts), and hypernyms (or super-concepts). Thus, *SimFea* is instantiated as follows:

$$\begin{aligned} &SimFea(A, B, KS_i, KS_j) \\ &= \max Sim_{glooses}\left(glooses_i(A), glooses_j(B)\right), Sim_{synonyms}\left(synonyms_i(A), synonyms_j(B)\right), \\ &Sim_{hyponyms}\left(hyponyms_i(A), hyponyms_j(B)\right), Sim_{hypernyms}\left(hypernyms_i(A), hypernyms_j(B)\right), \end{aligned}$$

where $A = \{\langle KS_i\text{: } glooses_i(A)\rangle, \langle KS_i\text{: } synonyms_i(A)\rangle, \langle KS_i\text{: } hyponyms_i(A)\rangle, \langle KS_i\text{: } hypernyms_i(A)\rangle\}$ and $B = \{\langle KS_j\text{: } glooses_j(B)\rangle, \langle KS_j\text{: } synonyms_j(B)\rangle, \langle KS_j\text{: } hyponyms_j(B)\rangle, \langle KS_j\text{: } hypernyms_j(B)\rangle\}$.

$Sim_{glooses}$, $Sim_{hyponyms}$, and $Sim_{hypernyms}$ are defined using Jaccard index, Sorensen coefficient, and Symmetric difference. $Sim_{synonyms}$ is defined as follows:

$$Sim_{synonyms}\left(synonyms_i(A), synonyms_j(B)\right) = \begin{array}{ll} 1, & \text{if } synonyms_i(A) \cap synonyms_j(B) \neq \varnothing \\ 0, & \text{if } synonyms_i(A) \cap synonyms_j(B) = \varnothing \end{array}$$

Therefore, we can define the following three kinds of feature-based semantic similarity between A and B w.r.t. KS_i and KS_j:

$$\begin{aligned} SimFea1(A, B, KS_i, KS_j) &= \max\{Sim_{synonyms}\left(synonyms_i(A), synonyms_j(B)\right) | Jaccard\left(glooses_i(A), glooses_j(B)\right), \\ &\quad Jaccard\left(hyponyms_i(A), hyponyms_j(B)\right), Jaccard\left(hypernyms_i(A), hypernyms_j(B)\right)\}, \\ SimFea2(A, B, KS_i, KS_j) &= \max\{Sim_{synonyms}\left(synonyms_i(A), synonyms_j(B)\right) | \\ &\quad Dice\left(glooses_i(A), glooses_j(B)\right), Dice\left(hyponyms_i(A), hyponyms_j(B)\right), \\ &\quad Dice\left(hypernyms_i(A), hypernyms_j(B)\right)\}, \end{aligned}$$

$$\begin{aligned} &SimFea3(A, B, KS_i, KS_j) \\ &\quad = \max\{Sim_{synonyms}\left(synonyms_i(A), synonyms_j(B)\right) | SaltonCosine\left(glooses_i(A), glooses_j(B)\right), \\ &\qquad SaltonCosine\left(hyponyms_i(A), hyponyms_j(B)\right), SaltonCosine\left(hypernyms_i(A), hypernyms_j(B)\right)\}. \end{aligned}$$

where $synonyms_i(A)$, $synonyms_j(B)$, $hyponyms_i(A)$, $hyponyms_j(B)$, $hypernyms_i(A)$, and $hypernyms_j(B)$ are some sets of concepts (or terms), and $glooses_i(A)$ and $glooses_j(B)$ are concept sets that contain words extracted by parsing glosses of A and B, respectively.

In Definition 26, *SimFea1*, *SimFea2*, and *SimFea3* are based on two knowledge sources. Now we give some similarity measures for multiple knowledge sources.

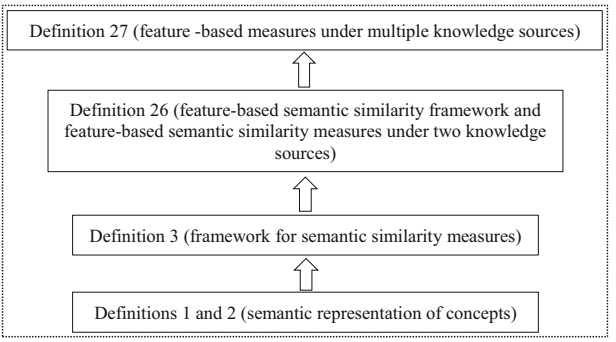

Fig. 8 The relationships among all definitions of feature-based measures

Definition 27. Let $AllKS = \{KS_1, KS_2, \ldots, KS_m\}$ be all knowledge sources, $KSA = \{KS_k, KS_{k+1}, \ldots, KS_l\} \subseteq AllKS$ and $KSB = \{KS_s, KS_{s+1}, \ldots, KS_t\} \subseteq AllKS$. For any $KS_i \in KSA$ and $KS_j \in KSB$, we have that $A \in KS_i$, $A \notin KS_j$, $B \in KS_j$, $B \notin KS_i$. The feature-based semantic similarity measures *SimFea1M*, *SimFea2M*, *SimFea3M*, *SimFea4M*, *SimFea5M*, and *SimFea6M* between A and B w.r.t. multiple knowledge sources KSA and KSB can be defined as:

$$\begin{aligned}
SimFea1M(A,B,KSA,KSB) &= \max_{i=k}^{l}\max_{j=s}^{t}\{SimFea1(A,B,KS_i,KS_j)\}\\
SimFea2M(A,B,KSA,KSB) &= \max_{i=k}^{l}\max_{j=s}^{t}\{SimFea2(A,B,KS_i,KS_j)\}\\
SimFea3M(A,B,KSA,KSB) &= \max_{i=k}^{l}\max_{j=s}^{t}\{SimFea3(A,B,KS_i,KS_j)\}\\
SimFea4M(A,B,KSA,KSB) &= \max_{i=k}^{l}\max_{j=s}^{t}\{SimFea4(A,B,KS_i,KS_j)\}\\
SimFea5M(A,B,KSA,KSB) &= \max_{i=k}^{l}\max_{j=s}^{t}\{SimFea5(A,B,KS_i,KS_j)\}\\
SimFea6M(A,B,KSA,KSB) &= \max_{i=k}^{l}\max_{j=s}^{t}\{SimFea6(A,B,KS_i,KS_j)\}
\end{aligned}$$

where $SimFea4(A, B, KS_i, KS_j) = \max\{Sim_{synonyms}(synonyms(A), synonyms(B)), Jaccard(glooses(A), glooses(B)), Jaccard(hyponyms(A), hyponyms(B)), Jaccard(hypernyms(A), hypernyms(B))\}$,

$SimFea5(A, B, KS_i, KS_j) = \max\{Sim_{synonyms}(synonyms(A), synonyms(B)), Dice(glooses(A), glooses(B)), Dice(hyponyms(A), hyponyms(B)), Dice(hypernyms(A), hypernyms(B))\}$,

$SimFea6(A, B, KS_i, KS_j) = \max\{Sim_{synonyms}(synonyms(A), synonyms(B)), SaltonCosine(glooses(A), glooses(B)), SaltonCosine(hyponyms(A), hyponyms(B)), SaltonCosine(hypernyms(A), hypernyms(B))\}$,

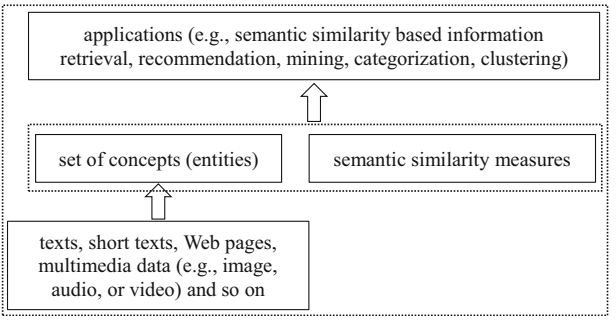

Fig. 9 Application architecture of semantic similarity measures

$glooses(A) = glooses_k(A) \cup \ldots \cup glooses_l(A)$,
$glooses(B) = glooses_s(B) \cup \ldots \cup glooses_t(B)$,
$synonyms(A) = synonyms_k(A) \cup \ldots \cup synonyms_l(A)$,
$synonyms(B) = synonyms_s(B) \cup \ldots \cup synonyms_t(B)$,
$hyponyms(A) = hyponyms_k(A) \cup \ldots \cup hyponyms_l(A)$,
$hyponyms(B) = hyponyms_s(B) \cup \ldots \cup hyponyms_t(B)$,
$hypernyms(A) = hypernyms_k(A) \cup \ldots \cup hypernyms_l(A)$,
$hypernyms(B) = hypernyms_s(B) \cup \ldots \cup hypernyms_t(B)$.

The feature-based semantic similarity measure *SimFea* between *A* and *B* w.r.t. *KSA* and *KSB* and baseline measures can be defined as:

SimFea(*A*, *B*, *KSA*, *KSB*)=

$$\max_{i=k}^{l} \max_{j=s}^{t} \left\{ \begin{array}{l} SimFea1\left(A, B, KS_i, KS_j\right), SimFea2\left(A, B, KS_i, KS_j\right), \\ SimFea3\left(A, B, KS_i, KS_j\right), SimFea4\left(A, B, KS_i, KS_j\right), \\ SimFea5\left(A, B, KS_i, KS_j\right), SimFea6\left(A, B, KS_i, KS_j\right) \end{array} \right\}.$$

In order to compare the values of different similarities *SimFea*1, *SimFea*2, *SimFea*3, *SimFea*4, *SimFea*5, and *SimFea*6, we also normalize the value of each similarity. Similarly to Definitions 15 and 25, the feature-based measure *SimFea* is also a generic and flexible approach.

The relationships among all definitions of feature-based measures under multiple knowledge sources are shown as Fig. 8.

Until now some generic and flexible approaches (including IC-based measures, distance-based measures, and feature-based measures) to similarity measures of concepts have been presented. As stated in Section 1, semantic similarity between concepts can be applied to many fields such as multimedia databases, multimedia encyclopedias, digital libraries, and multimedia documents. The application architecture is as follows (Fig. 9):

5 Experiments and evaluation

In this section we discuss the evaluation problem of our similarity measures (see Section 4). Section 5.1 introduces some experimental datasets and evaluation metrics. Section 5.2 gives our experimental results. Lastly, in Section 5.3, we discuss and analyze the experimental results.

5.1 Experimental datasets and evaluation metrics

We collect several publicly available gold standard benchmarks for evaluating concept semantic similarity, which are conventionally most common-used and some recently most updated benchmarks. The descriptions of these benchmarks used in the experiments are listed below.

(1) WS353 [22] benchmark contains 353 word pairs and 13 to 16 human subjects were asked to assign a numerical similarity score between 0.0 to 10.0 (0 means totally unrelated and 10 means very closely related). In fact, this benchmark measures general

relatedness rather than similarity because it considers other semantic relations (e.g., antonyms are considered as similar).

(2) WordSim-353 [2] benchmark is a subset of WS353. WS353 is divided into two subsets. The first one concerns about relatedness while the second subset focuses on similarity. We only use the second one named WordSim-353 in our experiments. It contains 203 pairs of words and it has been identified by the authors to be suitable for evaluating semantic similarity specially.

(3) R&G [66] benchmark is the first and most used benchmark containing human assessment of word similarity. The benchmark resulted from the experiment conducted in 1965 where a group of 51 students (all native English speakers) assessed the similarity of 65 pairs of words selected from ordinary English nouns. Those 51 subjects were requested to judge the similarity of meaning for two given words on a scale from 0.0 (completely dissimilar) to 4.0 (highly synonymous). It focuses on semantic similarity and ignores any other possible semantic relationships between the words.

(4) M&C [52] benchmark contains 30 word pairs. It replicated the R&G experiment again in 1991 by taking a subset of 30 noun pairs. The similarity between words was judged by 38 human subjects.

(5) Jiang-1 [37] and Jiang-2 [34] benchmarks contain 30 pairs of real-world Wikipedia concepts, respectively. The similarity between each concept pair is assessed by 10 students and 10 teachers in a scale between 0 (semantically unrelated) and 4 (highly synonymous). After a normalization process, a final set of 30 concept pairs is rated with the average of the similarity values provided by the students and the teachers. Thus, these two benchmarks are created and can be used to evaluate the accuracy of our approaches so that we use them in this work.

Each benchmark described above contains a list of triples comprising two words and a similarity score denoting word similarity judged by human. Concretely, we select 203 word pairs from WordSim-353, 65 word pairs from R&G, 30 word pairs from M&C, 30 word pairs from Jiang-1, and 30 word pairs from Jiang-2 in our experiments.

It is well known that an objective evaluation of the accuracy of semantic similarity functions is difficult because the notion of similarity is subjective. Generally, similarity measures are evaluated by means of standard benchmarks of word pairs whose similarity has been assessed by a group of human experts [37]. However, in this paper we evaluate our new approaches to measure similarity under multiple knowledge sources that existing similarity computation methods cannot deal with (traditional methods are generally based on one knowledge source). In particular, for any word pairs (or concept pairs) (A, B), A and B belong to different knowledge sources (A and B belong to the same knowledge source in traditional methods). Therefore, comparison of the proposed methods with standard benchmarks imposes some challenges and requires some modifications and adjustments in order to make such comparison meaningful. The comparative experiments have been group into three parts.

Firstly, we evaluate our methods over 5 benchmarks, namely M&C, R&G, WordSim-353, Jiang-1, and Jiang-2 and two kinds of knowledge sources, namely Wikipedia[1] and WordNet.[2]

[1] https://en.wikipedia.org/
[2] http://wordnet.princeton.edu/

Table 1 Our benchmark Jiang-3

$Concept_1$	$Concept_2$	Similarity
Categorization	Migraine with aura	1.07
Categorization	Migraine without aura	1.07
Folk	Vitamin A deficiency	1.40
Folk	Intellectual disability	1.67
Folk	Histrionic personality disorder	1.63
Folk	Borderline personality disorder	1.60
Gender	Tracheomalacia	0.70
Gender	Retinopathy of prematurity	0.67
Gender	Pericardial effusion	0.63
Gender	Hypertrichosis	1.13
Gender	Venous insufficiency	0.57
Gender	Retinal vasculitis	0.57
Gender	Retinoschisis	0.60
Immortality	Thyroid nodule	0.53
Maxillaria	Facial paralysis	1.53
Paranormal	Otitis media with effusion	1.23
Paranormal	Nevus of Ota	1.20
Rescue	Histrionic personality disorder	0.73
Rescue	Intellectual disability	0.70
Video	Thyroid nodule	0.37
Gender	Tricuspid atresia	0.57
Gender	Varicose veins	0.53
Gender	Tricuspid valve prolapse	0.53
Gender	Corneal neovascularization	0.53
Gender	Hydrops fetalis	0.97
Gender	Cholesteatoma	0.77
Gender	Budd-Chiari syndrome	0.77
Priacanthidae	Vitamin E deficiency	0.73
Gender	Bladder exstrophy	0.73
Priacanthidae	Histrionic personality disorder	0.20
Gender	Wolff-Parkinson-White syndrome	0.67
Priacanthidae	Vitamin A deficiency	1.47
Gender	Cutis laxa	0.53
Gender	Keratoconjunctivitis sicca	0.57
Gender	Thyroid nodule	0.60
Gender	Hypotrichosis	0.57
Gender	Fibrosarcoma	0.53
Gender	Retinal vein occlusion	0.53
Gender	Aniridia	0.53
Ignorance	Craniopharyngioma	0.30
Prevention	Migraine without aura	0.63
Reasoning	Chordoma	0.03
Theme	Ganglioneuroma	0.07
Protectionism	Hepatoblastoma	0.00
Corruption	Burkitt lymphoma	0.17
Form	Pilomatrixoma	0.57
Gender	Budd-Chiari syndrome	0.53
Minuartia	Vitamin D deficiency	0.40
Paranormal	Vitamin D deficiency	1.20
Gender	Chordoma	0.53

To evaluate our methods objectively, for any concept pair (A, B), we require that the value of A comes from Wikipedia and the value of B comes from WordNet.

Secondly, we develop a benchmark Jiang-3 and then use it to evaluate the accuracy of our proposals. For comparison purposes, we select 30 pairs of real-world concepts extracted from

Table 2 Our benchmark Jiang-4

$Concept_1$	$Concept_2$	Similarity
Rete testis adenocarcinoma	Carcinoid tumor	1.43
Orthostatic proteinuria	Renal insufficiency	1.87
Orthostatic proteinuria	Hyperoxaluria	1.63
Vesiculobullous skin disease	Erythema	1.77
Vesiculobullous skin disease	Pruritus	1.20
Vesiculobullous skin disease	Skin ulcer	2.17
Large intestine adenocarcinoma	Carcinoid tumor	1.73
Angiodysplasia of intestine	Intestinal fistula	1.67
Angiodysplasia of intestine	Intestinal polyposis	2.30
Renal hypertension	Renal insufficiency	2.00
Achenbach syndrome	Erythema	1.13
Achenbach syndrome	Pruritus	0.90
Intestinal tuberculosis	Intestinal fistula	1.87
Intestinal tuberculosis	Intestinal polyposis	2.13
Nervous system disease	Vertigo	2.47
Nervous system disease	Cerebral hemorrhage	2.60
Nervous system disease	Brain abscess	2.53
Skin atrophy	Necrolytic migratory erythema	1.97
Skin atrophy	Erythema	1.70
Skin atrophy	Skin ulcer	2.03
Exanthem	Necrolytic migratory erythema	3.30
Exanthem	Erythema	3.27
Intestinal disaccharidase deficiency	Intestinal fistula	1.20
Sitosterolemia	Intestinal polyposis	0.97
Urethra adenocarcinoma	Carcinoid tumor	1.47
Prostate adenocarcinoma	Carcinoid tumor	1.43
Pleural disease	Hemothorax	2.40
Pleural disease	Pleural effusion	2.77
Primary hyperoxaluria	Renal insufficiency	1.60
Primary hyperoxaluria	Hyperoxaluria	3.50
Hemangioma of subcutaneous tissue	Pruritus	1.60
Hemangioma of subcutaneous tissue	Skin ulcer	2.43
Osteochondrodysplasia	Osteochondroma	2.60
Bile duct adenoma	APUdoma	2.77
Kidney cancer	Renal insufficiency	2.83
Kidney cancer	Hyperoxaluria	1.83
Skin disease	Hyperhidrosis	3.07
Skin disease	Albinism	2.73
Skin disease	Erythema nodosum	3.10
Skin disease	Pressure ulcer	3.07
Skin abnormality	Pallor	3.10
Bowel dysfunction	Enterocolitis	3.07
Familial juvenile hyperuricemic nephropathy	Renal insufficiency	2.33
Bowel dysfunction	Intestinal fistula	2.80
Basophilic carcinoma	Carcinoid tumor	1.97
Bartholins gland adenocarcinoma	Carcinoid tumor	1.93
Sweat gland disease	Skin ulcer	1.13
Sebaceous gland disease	Erythema	1.13
Inflammatory bowel disease	Intestinal fistula	2.10
Mucocele of appendix	Intestinal polyposis	2.27

Table 3 Our benchmark Jiang-5

$Concept_1$	$Concept_2$	Similarity
Regulation of gene expression, epigenetic	Gene expression regulation	3.70
Regulation of gene expression, epigenetic	Chromatin assembly and disassembly	3.33
Regulation of gene expression, epigenetic	Ectopic gene expression	3.50
Regulation of gene expression, epigenetic	Gene amplification	2.97
Regulation of gene expression, epigenetic	Transcriptional activation	3.23
Biological process	Life cycle stages	3.07
Biological process	Pathologic processes	2.37
Biological process	Action potentials	1.87
DNA polymerase complex	Multifunctional enzymes	2.97
DNA polymerase complex	DNA restriction-modification enzymes	3.03
DNA polymerase complex	Deubiquitinating enzymes	2.23
DNA polymerase complex	Recombinases	2.90
DNA polymerase complex	Holoenzymes	2.90
Chromosome	Karyotype	3.00
Chromosome	Chromosomes insect	3.43
Chromosome	Karyotyping	1.60
Chromosome	Cells	2.60
Chromosome	Chromosome structures	3.33
Cellular component	Ribosomes	2.83
Cellular component	Axons	2.70
Cell	Bone marrow cells	3.00
Cell	Neurons	2.90
Cell	Connective tissue cells	2.87
Cell killing	Excitatory postsynaptic potentials	1.67
Cell killing	Gene expression regulation	2.50
Biological process	Biological science disciplines	2.83
Biological process	Suicide attempted	1.47
Biological process	Helping behavior	1.77
Biological process	Vasoplegia	2.47
Biological process	Biological phenomena	3.43
Biological process	Environment and public health	1.43
Biological process	Myocardial contraction	2.13
Biological process	Breast milk expression	2.00
Chromatin remodeling	Cell cycle checkpoints	2.20
DNA polymerase complex	Isomerases	2.63
DNA polymerase complex	Oxidoreductases	2.67
DNA polymerase complex	Lyases	2.70
DNA polymerase complex	Ligases	3.00
Cellular process	Cell cycle checkpoints	3.17
Neuron part	Lewy bodies	2.63
Neuron part	Membranes	2.60
Neuron part	Synaptic vesicles	2.90
Ion channel activity	Cell cycle checkpoints	1.23
Chromosome	Membranes	1.97
Cellular component	Adherens junctions	2.97
Cellular component	Inclusion bodies	2.93
Cellular component	Cilia	3.00
Cellular component	Cytoplasmic vesicles	3.03
Cellular component	Organelles	3.27
Cell	Myeloid cells	2.80

some widely used knowledge sources, i.e., Wikipedia, WordNet, Medical Subject Headings (MeSH),[3] Disease Ontology (DO)[4] and Human Phenotype Ontology (HPO).[5] Our benchmark Jiang-3 is shown in Table 1. The similarity between each concept pair is assessed by 10 students and 10 teachers in biomedical fields in a scale between 0 (semantically unrelated) and 4 (highly synonymous), respectively. After a normalization process, a final set of 30 concept pairs is rated with the average of the similarity values provided by the students and the teachers. To evaluate our methods objectively, for any concept pair (A, B), we require that $A \in Wikipedia$, $A \in WordNet$, $A \notin MeSH$, $A \notin DO$, $A \notin HPO$, $B \in MeSH$, $B \in DO$, $B \in HPO$, $B \notin Wikipedia$, and $B \notin WordNet$.

Lastly, in our benchmark Jiang-3 there are five kinds of knowledge sources, i.e., Wikipedia, WordNet, MeSH, DO, and HPO. Clearly, Wikipedia and WordNet are two kinds of general-purpose knowledge sources, but MeSH, DO, and HPO are three kinds of domain dependent knowledge sources (biomedical ontologies). To evaluate the accuracy of our proposals in another setting, we build another two benchmarks Jiang-4 and Jiang-5 by using knowledge sources MeSH, DO, HPO, Gene Ontology (GO),[6] and Ontology for Biomedical Investigations (OBI).[7] In our benchmark Jiang-4 there are 30 pairs of real-world concepts extracted from three kinds of knowledge sources, i.e., MeSH, DO, and HPO. Jiang-4 is shown in Table 2. For any concept pair (A, B), we require that $A \in MeSH$, $A \in HPO$, $A \notin DO$, $B \in DO$, $B \notin MeSH$, and $B \notin HPO$. In our benchmark Jiang-5 there are 30 pairs of real-world concepts extracted from three kinds of knowledge sources, i.e., MeSH, GO, and OBI. Jiang-5 is shown in Table 3. For any concept pair (C, D), we require that $C \in MeSH$, $C \notin GO$, $C \notin OBI$, $D \in GO$, $D \in OBI$, and $D \notin MeSH$.

Different knowledge sources have different semantic information such as concept taxonomies and distributions of instances over concepts. We apply different combinations of knowledge sources to different benchmarks in this work and express the semantics of concepts through integrating different semantic information. To further illustrate it, we describe the relations among seven knowledge sources considered in our experiments and eight benchmarks in Fig. 10. The mark "1" on the arrow from knowledge source to benchmark represents the first concept in each pair of benchmark is computed in corresponding knowledge source. Similarly, the mark "2" represents the second concept in each pair of benchmark is computed in corresponding knowledge source. For example, the first concept in each pair of Jiang-3 benchmark is computed on WordNet and Wikipedia, and the second concept is computed on HPO, DO, and MeSH.

The knowledge sources WordNet and Wikipedia are used in measuring semantic similarities of concept pairs in M&C, R&G, WordSim-353, Jiang-1, Jiang-2 and Jiang-3 benchmarks. The WordNet organizes the lexical information in meanings (senses) and synsets (set of synonym words in a specific context) [5]. Each synset has a gloss that defines the concept. Hypernymy is a relation that organizes noun synsets into a lexical inheritance taxonomy. In this taxonomy, a subordinate term inherits the basic features from the superordinate term and adds its distinctive features to form its own meaning. The Wikipedia is a free, online multilingual knowledge source that is collaboratively maintained by volunteers and known to have a good coverage capacity [30]. At the bottom of each page in Wikipedia, all assigned categories are listed with links to the category page. These categories are connected to form the

[3] https://www.nlm.nih.gov/mesh/
[4] http://www.disease-ontology.org/
[5] http://human-phenotype-ontology.github.io/
[6] http://www.geneontology.org/
[7] http://obi-ontology.org/

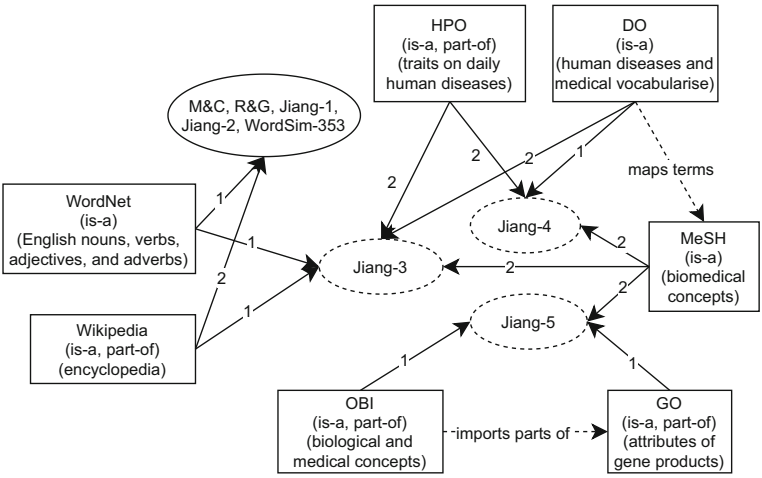

Fig. 10 The relation among seven knowledge sources WordNet, Wikipedia, OBI, GO, MeSH, DO, and HPO and eight benchmarks M&C, R&G, WordSim353, Jiang-1, Jiang-2, Jiang-3, Jiang-4, and Jiang-5 (the connections from knowledge sources to benchmarks show the components of each benchmarks)

Wikipedia Category Graph (WCG). Wikipedia categories and their relations do not have explicit semantics like WordNet. The Wikipedia categorization system does not form a taxonomy like the WordNet "is-a" taxonomy with a fully subsumption hierarchy, but only through a thematically organized thesaurus. For example, *Computer systems* is categorized in the upper category of *Technology systems* (is-a) and *Computer hardware* (has-part).

The knowledge source MeSH is used in measuring semantic similarities of concept pairs in Jiang-3, Jiang-4, and Jiang-5 benchmarks. The MeSH organizes biomedical concepts in a meaningful way with explicit semantic relations. It consists of single- and multi-word terms that are used to index and catalog the medical literature [16]. Among the relations [5], we use the MeSH "is-a" taxonomy. The knowledge sources DO and HPO are used in measuring semantic similarities of concept pairs in Jiang-3 and Jiang-4 benchmarks. The DO has been developed as a standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts. Also, the DO semantically integrates disease and medical vocabularies through extensive cross mapping terms to the MeSH thesaurus. The HPO is devising a system or a domain for the traits of phonomes and their effects on daily encountered human diseases [81]. The aim is to provide a well-structured vocabulary for these traits so that they can be easily studied and searched in the field of medical science to bring awareness about the traits and how they can damage a person's health and body organs. The HPO currently contains over 13,000 different terms of traits and characteristics, and over 156,000 annotations to hereditary diseases. Each term describes a phenotypic abnormality such as *Atrial septal defect*.

The knowledge sources GO and OBI are used in measuring semantic similarities of concept pairs in Jiang-5 benchmark. The GO provides an ontology to describe attributes of gene products in three non-overlapping domains of molecular biology [26]. It includes several of the world's major repositories for plant, animal and microbial genomes. Within each ontology, terms have free text definitions and stable unique identifiers. The vocabularies are structured in

a classification that supports "is-a" and "part-of" relationships. The OBI is an ontology that provides terms with precisely defined meanings to describe all aspects of how investigations in the biological and medical domains are conducted [7]. It imports parts of other biomedical ontologies such as GO, Chemical Entities of Biological Interest (ChEBI) and Phenotype Attribute and Trait Ontology (PATO) without altering their meanings. OBI is being used in a wide range of projects covering genomics, multi-omics, immunology, and catalogs of services.

The accuracy of the similarity computation is quantified by computing the correlation between the human judgements and the results provided by the computerized measures. This enables an objective evaluation of the different similarity computation methods. The correlation between two variables is the degree to which there is a relationship between them. Correlation is usually expressed as a coefficient which measures the strength of a relationship between the variables. Our experiments will use two measures of correlation: Pearson (Pearson correlation coefficient) and Spearman (Spearman correlation coefficient). Pearson reflects the linear correlation between measuring result with human judgments. Spearman is another metric and compares the correlation between measuring result with human judgments based on the ranking strategy.

5.2 Experimental results

For environment of our evaluation, the version of Wikipedia is released on April 20, 2018, the version of WordNet is 3.1, the version of MeSH is released in 2018, the version of GO is released on May 31, 2018, the version of DO is released on May 15, 2018, the version of HPO is released on March 9, 2018, and the version of OBI is released on April 29, 2016. At the same time, we use JWPL (Java Wikipedia Library), Java with JDK1.8 and MySQL to implement our algorithms to measure similarity by the formulas given in Section 4. According

Table 4 Numbers of the concepts in the intersections of seven considered knowledge sources

Type	Numbers
Set(*GO*)∩*Set*(*DO*)	0
Set(*OBI*)∩*Set*(*DO*)	2
Set(*OBI*)∩*Set*(*HPO*)	2
Set(*GO*)∩*Set*(*HPO*)	4
Set(*GO*)∩*Set*(*OBI*)	149
Set(*MeSH*)∩*Set*(*HPO*)	1685
Set(*OBI*)∩*Set*(*GO*)∩*Set*(*Mesh*)	10
Set(*HPO*)∩*Set*(*DO*)∩*Set*(*Mesh*)	412
Set(*DO*)∩*Set*(*OBI*)∩*Set*(*Mesh*)∩*Set*(*HPO*)	0
Set(*GO*)∩*Set*(*OBI*)∩*Set*(*Mesh*)∩*Set*(*HPO*)	0
Set(*HPO*)∩*Set*(*DO*)∩*Set*(*MeSH*)∩!*Set*(*Wikipedia*)∩!*Set*(*WordNet*)	162
!*Set*(*HPO*)∩!*Set*(*DO*)∩!*Set*(*MeSH*)∩*Set*(*Wikipedia*)∩*Set*(*WordNet*)	5553
!*Set*(*GO*)∩! *Set*(*OBI*)∩*Set*(*MeSH*)	28,578
Set(*GO*)∩*Set*(*OBI*)∩!*Set*(*MeSH*)	140
!*Set*(*DO*)∩*Set*(*HPO*)∩*Set*(*MeSH*)	367
Set(*DO*)∩!*Set*(*HPO*)∩!*Set*(*Mesh*)	9432

Note. Suppose that *A* and *B* are two knowledge sources. *Set*(*A*) represents the set of concepts of *A*. *Set*(*A*)∩*Set*(*B*) is a set of concepts and contains the concepts appear in both *Set*(*A*) and *Set*(*B*). *Set*(*A*)∩!*Set*(*B*) is a set of concepts and contains the concepts appear in *Set*(*A*), and don't appear in *Set*(*B*). In addition, *MeSH*, *HPO*, *DO*, *GO*, *OBI*, *Wikipedia*, and *WordNet* are seven knowledge sources (see Section 5.1 for more details)

to our statistics, there are 47,204 concepts in GO, 11192 concepts in DO, 3337 concepts in OBI, 13544 concepts in HPO, 28938 concepts in MeSH, 1,679,499 concepts in Wikipedia and 147,479 concepts in WordNet.

Tables 4 and 5 are some related results about the developments of Jiang-3, Jiang-4 and Jiang-5 benchmarks. To evaluate the similarity of the concepts come from different knowledge sources, the common concepts are the key factor in both path-based and IC-based approaches proposed in this paper. In fact, common concepts are the elements of the intersections of the corresponding concept sets of different knowledge sources. In this case, we list the numbers of the elements in the intersections of different concept sets of seven knowledge sources MeSH, DO, HPO, OBI, GO, Wikipedia and WordNet in Table 4. We take different combinations of knowledge sources to build benchmarks Jiang-3, Jiang-4 and Jiang-5. We list the numbers of the concept pairs that are generated by different combinations in Table 5. According to the numbers of the pairs that have common ancestors or children and the numbers of the pairs that perform well on all three types approaches proposed in this paper, we adopt the last three division schemes and extract 50 concept pairs from each scheme to generate Jiang-3, Jiang-4 and Jiang-5 benchmarks, respectively.

The second (M&C), third (R&G), fourth (WordSim-353), fifth (Jiang-1), sixth (Jiang-2), seventh (Jiang-3), eighth (Jiang-4), and ninth (Jiang-5) columns in Table 6 show the Pearson correlation coefficients of the different measures with human judgments.

The second (M&C), third (R&G), fourth (WordSim-353), fifth (Jiang-1), sixth (Jiang-2), seventh (Jiang-3), eighth (Jiang-4), and ninth (Jiang-5) columns in Table 7 show the Spearman correlation coefficients of the different measures with human judgments.

5.3 Discussion and analysis

Now we analyze and discuss the experimental results (see Tables 6 and 7) from four different aspects: (1) the influence of knowledge sources, (2) the influence of benchmarks, (3) the differences among three kinds of measures: IC-based measures, Distance-based measures, and

Table 5 The details of the concept pairs in different combinations of knowledge sources

Division scheme	Common concept pairs	Final concept pairs	Flag	Benchmark name
(*MeSH*, *HPO*, *DO*)/(*GO*, *OBI*)	44	44	×	–
(*GO*, *OBI*)/(*MeSH*, *HPO*)	126	50	×	–
(*MeSH*, *HPO*, *DO*)/(*Wikipedia*, *WordNet*)	4270	50	√	Jiang-3
DO/(*MeSH*, *HPO*)	52,871	50	√	Jiang-4
(*GO*, *OBI*)/*MeSH*	7592	50	√	Jiang-5

Note. Take the third row for example, "(*MeSH*, *HPO*, *DO*)/(*Wikipedia*, *WordNet*)" means the first concept in a concept pair comes from the intersection of *Set*(*MeSH*), *Set*(*HPO*), and *Set*(*DO*) while the second comes from the intersection of *Set*(*Wikipedia*) and *Set*(*WordNet*). "4270" means the number of concept pairs that have a common concept (see Definition 17) in multiple taxonomy structures is 4270. "50" is the number of concept pairs we exploit in constructing the benchmark. The flag "√" means this combination is appropriate to construct the benchmark and "Jiang-3" is the name for the benchmark. In addition, in the first and second rows, the flag "×" means the combination is not appropriate to construct our benchmark and the flag "-" means that we don't construct the corresponding benchmarks. Finally, *MeSH*, *HPO*, *DO*, *GO*, *OBI*, *Wikipedia*, and *WordNet* are seven knowledge sources (see Section 5.1 for more details)

Feature-based measures, (4) the performances of three most generic and flexible measures: *SimIC*, *SimDis*, and *SimFea*.

5.3.1 Influence of knowledge sources

The results in Tables 6 and 7 show that the most results on both Pearson correlation and Spearman correlation coefficients on the benchmarks M&C, R&G, Jiang-1, and Jiang-3 are better than those on benchmarks WordSim-353, Jiang-2, Jiang-4, and Jiang-5. It indicates that domain-independent knowledge sources like Wikipedia and WordNet perform better in measuring similarities among both general and special concepts. The reason is that the semantic information of the concepts in M&C, R&G, Jiang-1, and Jiang-3 is computed on Wikipedia and WordNet, but Jiang-4 and Jiang-5 are computed on five biomedical knowledge sources. Furthermore, they are biomedical ontologies and the expressions of the same word are often different from encyclopedia. For example, the glosses of the same concept in HPO and WordNet varies from each other and semantic information in WordNet contains more features.

GO, DO, OBI, HPO, and MeSH are all the domain-specific ontologies which express the concepts professionally, but Wikipedia and WordNet express the concepts more general. So this is a problem which the features of the same concept from different knowledge sources are different and even some features are empty. And we use our methods to compute semantic similarity between concepts based on the features so that it causes the differences in our results.

Table 6 Results on Pearson correlation with human judgments of similarity measures

Measure	M&C	R&G	WordSim-353	Jiang-1	Jiang-2	Jiang-3	Jiang-4	Jiang-5
$SimIC1M_{fir}$	**0.797**	**0.670**	0.451	0.413	0.001	**0.702**	0.056	0.437
$SimIC2M_{thi}$	**0.685**	**0.541**	0.175	0.157	0.054	−0.542	0.129	0.475
$SimIC3M_{fir}$	**0.566**	0.383	0.166	**0.538**	0.350	**0.663**	−0.245	0.187
$SimIC3M_{sec}$	**0.805**	**0.740**	0.421	**0.723**	0.186	0.452	−0.171	0.208
SimDis2M	**0.813**	**0.632**	0.492	0.196	0.106	0.044	−0.151	**0.503**
SimDis3M	**0.681**	0.372	0.102	0.067	−0.054	0.008	**0.572**	0.381
SimDis4M	**0.842**	**0.720**	0.378	**0.605**	−0.054	0.008	**0.571**	0.380
SimDis5M	**0.822**	**0.738**	0.442	**0.699**	0.039	−0.059	**0.567**	0.416
SimDis	**0.710**	**0.585**	0.388	0.169	−0.058	0.057	0.394	0.274
SimFea1M	**0.811**	**0.750**	0.427	**0.722**	0.215	0.006	0.287	0.251
SimFea2M	**0.811**	**0.752**	0.428	**0.723**	0.215	0.013	0.287	0.245
SimFea3M	−0.114	0.201	0.426	0.008	0.211	0.007	0.287	0.251
SimFea4M	0.245	0.221	0.327	0.156	0.138	**0.683**	0.168	0.292
SimFea5M	0.244	0.222	0.328	0.159	0.138	**0.682**	0.149	0.316
SimFea6M	0.167	0.270	0.425	0.436	0.220	**0.683**	0.168	0.292
SimFea	**0.682**	0.025	0.328	−0.175	0.127	**0.682**	0.238	0.320
$SimIC1M_{thi}$	−0.749	−0.657	−0.288	−0.519	0.138	**0.702**	−0.066	0.480
$SimIC1M_{sec}$	0.054	−0.181	−0.015	0.309	−0.208	0.000	0.056	−0.285
$SimIC2M_{fir}$	−0.309	−0.320	−0.044	−0.325	−0.388	−0.512	0.344	−0.170
$SimIC2M_{sec}$	−0.674	−0.741	−0.265	−0.144	−0.224	−0.283	0.281	−0.307
$SimIC3M_{thi}$	−0.697	−0.631	−0.164	−0.092	0.085	−0.103	−0.428	−0.216
SimIC	0.121	0.134	0.132	0.161	−0.133	−0.068	0.125	−0.101
SimDis1M	−0.532	−0.530	−0.322	−0.526	−0.109	0.057	**0.583**	0.274

Note. From left to right: measure approach, correlation for M&C benchmark, correlation for R&G benchmark, correlation for WordSim-353 benchmark, correlation for Jiang-1 benchmark, correlation for Jiang-2 benchmark, correlation for Jiang-3 benchmark, correlation for Jiang-4 benchmark, and correlation for Jiang-5 benchmark

Table 7 Results on Spearman correlation with human judgments of similarity measures

Measure	M&C	R&G	WordSim-353	Jiang-1	Jiang-2	Jiang-3	Jiang-4	Jiang-5
$SimIC1M_{fir}$	**0.516**	0.081	−0.389	0.098	0.091	**0.987**	0.324	0.103
$SimIC2M_{thi}$	**0.644**	0.304	−0.397	0.030	−0.088	**0.763**	0.442	0.304
SimDis2M	**0.700**	0.355	−0.105	0.349	0.245	**0.543**	0.096	0.168
SimDis3M	0.468	0.039	−0.202	0.350	0.166	**0.961**	0.488	0.341
SimDis4M	**0.607**	0.157	−0.160	**0.512**	0.166	**0.961**	0.488	0.341
SimDis5M	**0.689**	0.144	−0.134	**0.525**	0.166	**0.765**	0.473	0.470
SimDis	**0.653**	0.416	0.214	0.050	−0.032	**0.703**	0.392	0.182
SimFea1M	**0.757**	**0.733**	0.073	0.343	0.081	0.167	0.267	0.066
SimFea2M	**0.757**	**0.733**	0.073	0.343	0.081	0.167	0.267	0.066
SimFea3M	0.259	−0.055	−0.122	−0.474	−0.234	0.167	0.267	0.066
SimFea4M	**0.636**	**0.576**	0.354	0.058	0.189	**0.671**	0.203	0.073
SimFea5M	**0.636**	**0.576**	0.354	0.058	0.189	**0.513**	0.203	0.073
SimFea6M	0.382	0.205	0.132	−0.222	0.437	**0.649**	0.203	0.073
SimFea	0.017	0.136	0.354	−0.073	0.221	0.473	0.209	0.092
$SimIC1M_{sec}$	−0.679	−0.839	−0.411	−0.195	−0.662	**0.986**	0.324	−0.329
$SimIC1M_{thi}$	−0.036	−0.448	−0.604	−0.513	−0.386	**0.987**	0.318	0.047
$SimIC2M_{fir}$	0.126	0.227	−0.031	−0.075	−0.179	−0.283	**0.512**	−0.389
$SimIC2M_{sec}$	0.092	0.225	0.146	−0.010	−0.245	0.008	0.216	−0.686
$SimIC3M_{fir}$	0.419	−0.171	−0.312	−0.378	−0.601	**0.987**	−0.082	−0.149
$SimIC3M_{sec}$	−0.036	−0.450	−0.489	−0.538	−0.512	**0.987**	0.116	−0.038
$SimIC3M_{thi}$	−0.252	−0.594	−0.476	0.001	−0.078	**0.916**	0.038	−0.319
SimIC	−0.850	−0.711	−0.558	−0.751	−0.548	**0.987**	0.009	−0.111
SimDis1M	0.211	−0.241	−0.223	0.071	0.166	**0.703**	0.476	0.182

Note. From left to right: measure approach, correlation for M&C benchmark, correlation for R&G benchmark, correlation for WordSim-353 benchmark, correlation for Jiang-1 benchmark, correlation for Jiang-2 benchmark, correlation for Jiang-3 benchmark, correlation for Jiang-4 benchmark, and correlation for Jiang-5 benchmark

5.3.2 Influence of benchmarks

Eight benchmarks are computed in our experiments. For the first five benchmarks, Tables 6 and 7 show that the results of both Pearson and Spearman correlation coefficients on M&C, R&G, and Jiang-1 are relatively better than WordSim-353 and Jiang-2. For all the concept pairs in these five benchmarks, we measure the semantic information of one concept of each pair on WordNet and the other of each pair on Wikipedia. M&C is a subset of R&G with the relabeled human judgments. All the concepts in these three benchmarks (M&C, R&G, and Jiang-1) are ordinary English nouns so they are fully described in both lexical databases like WordNet and encyclopedia like Wikipedia. The characteristics of both benchmarks (M&C, R&G, and Jiang-1) and knowledge sources (WordNet and Wikipedia) make the results on M&C, R&G, and Jiang-1 relatively good. Jiang-2 contains pairs of real-world Wikipedia concepts and over half of them don't appear in WordNet taxonomy structure. The correlation coefficients on Jiang-2 benchmark don't exceed 0.5 in Tables 6 and 7. WordSim-353 is a dataset for measuring semantic relatedness between words (concepts) so the correlation coefficients of semantic similarity task on it is not good by using both WordNet and Wikipedia.

The results of Pearson correlation coefficients on Jiang-3 are a little better compared with Jiang-4 and Jiang-5, and the Spearman correlation coefficients are much better than both Jiang-4 and Jiang-5. The best Spearman correlation coefficient is higher than 0.98 on Jiang-3 but lower than 0.52 on both Jiang-4 and Jiang-5. The first reason may be that the diversity of the concept pairs in Jiang-3 are much higher.

Since Jiang-3 involves both the biomedical ontologies and common knowledge sources, which can provide with more semantic information. The second reason may be caused by the small different integrality of the semantic information of the concepts in Jinag-4 and Jiang-5. The information contained in Jiang-4 and Jiang-5 are much more professional but in Jiang-3 are much more extensive.

5.3.3 Influence of measures

Three kinds of measures, i.e., IC-based measures, distance-based measures, and feature-based measures, are proposed in this paper. For the IC-based measures, it is obvious that the measures $SimIC1M_{fir}$, $SimIC2M_{thi}$, $SimIC3M_{fir}$, and $SimIC3M_{sec}$ perform well on Pearson results while other six measures ($SimIC1M_{thi}$, $SimIC1M_{sec}$, $SimIC2M_{fir}$, $SimIC2M_{sec}$, $SimIC3M_{thi}$, and $SimIC$) don't. Meanwhile, the measures $SimIC1M_{fir}$ and $SimIC2M_{thi}$ perform relatively better on Spearman results than other eight measures ($SimIC1M_{sec}$, $SimIC1M_{thi}$, $SimIC2M_{fir}$, $SimIC2M_{sec}$, $SimIC3M_{fir}$, $SimIC3M_{sec}$, $SimIC3M_{thi}$, and $SimIC$). These four measures ($SimIC1M_{fir}$, $SimIC2M_{thi}$, $SimIC3M_{fir}$, and $SimIC3M_{sec}$) involve three approaches of IC computation and three IC-based similarity measurement methods introduced in Section 4.1, which illustrates that IC-based measures are all feasible if they are adopted appropriately. The measure $SimIC3M_{sec}$ outperforms other measures with Pearson correlation coefficients 0.805, 0.740 and 0.723 on M&C, R&G, and Jiang-1, respectively. For Spearman results, the measure $SimIC2M_{thi}$ outperforms other measures with 0.644 on M&C and 0.763 Jiang-3. These confirm the statistical similarity measures like IC-based measures are effective on multiple heterogeneous taxonomy structures.

For the distance-based measures, all the measures obtain good correlation coefficients except *SimDis1M*. A major reason of poor result about *SimDis1M* is that the depths of knowledge sources are greatly different from each other. The same *spath* considered in Definition 19 represents different similarity values in different knowledge sources. Figure 11 shows an example to illustrate the different cases of the same *spath* of two concept pairs (A, B) and (A', B'). $spath(A, B, T_i, T_j) = 4$ and $spath(A', B', T_i, T_j) = 4$, but they are not equally similar since the maximum depths of T_i and T_j are 10 and 5 separately. So the lengths of the path in different taxonomy structures are of different semantic meanings. In contrast, the measure *SimDis5M* obtains Pearson correlation coefficients 0.822, 0.738, 0.442, 0.699, 0.567 and 0.416 respectively on M&C, R&G, WordSim-353, Jiang-1, Jiang-4 and Jiang-5 benchmarks. These

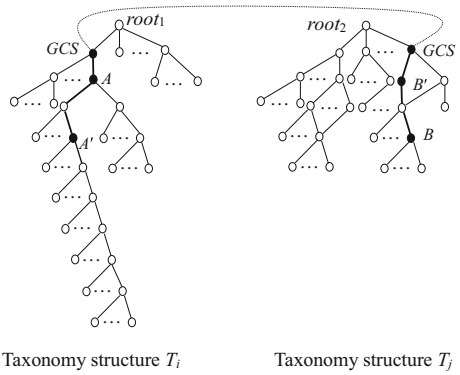

Fig. 11 Taxonomy structures T_i and T_j

show the feasibility of computing semantic similarity of concepts with the distances among them on multiple taxonomy structures.

Most of the correlation coefficients of feature-based measures listed in Tables 6 and 7 are positive. The measures *SimFea1M* and *SimFea2M* obtain nearly the same correlation coefficients on all benchmarks. It indicates the set operations *Jaccard* and *Dice* in the similarity computation of features influence the result slightly. For *SimFea3M*, both the Pearson and Spearman results are relatively lower than *SimFea1M* and *SimFea2M*. For measures *SimFea4M*, *SimFea5M*, and *SimFea6M*, they combine the features from multiple knowledge sources before computing similarities. However, comparing the performances of *SimFea1M* and *SimFea4M* on all benchmarks, we find there is a significant decrease from the former to the latter on M&C, R&G and Jiang-1, but an increase from the former to the latter on Jiang-3 and Jiang-5. Analogously, the decreases also appear in the measures pairs (*SimFea2M*, *SimFea5M*) and (*SimFea3M*, *SimFea6M*). This illustrates that computing similarity of aggregate features come from multiple knowledge sources doesn't always perform better than considering the feature in a separated knowledge source.

5.3.4 Our most generic and flexible approaches

The similarity computed in measure *SimIC* is completely depended on the maximum similarity of other nine IC-based measures ($SimIC1M_{fir}$, $SimIC1M_{sec}$, $SimIC1M_{thi}$, $SimIC2M_{fir}$, $SimIC2M_{sec}$, $SimIC2M_{thi}$, $SimIC3M_{fir}$, $SimIC3M_{sec}$, and $SimIC3M_{thi}$). To reduce the deviations of similarities among different IC-based measures, we normalize the similarities of each measure before computing *SimIC*. However, the results of correlation coefficients are not good in Tables 6 and 7 on all benchmarks, which means it is improper to compare different IC-based measures by similarities and set the value of *SimIC* to the maximum similarity. The major reason may be related to the incommensurable importance of different similarity values in different measures.

Similar to *SimIC* discussed above, the composite measures *SimDis* and *SimFea* also have lower correlation coefficients than separated measures such as *SimDis5M* and *SimFea2M*. This explicitly shows that the maximum methods generated by different semantic similarity computation measures can hardly improve the performance of similarity computation.

6 Conclusion

The final goal of computerized similarity measures is to accurately mimic human judgements about semantic similarity. At present similarity measures have been used for many different areas such as natural language processing, information retrieval, and word sense disambiguation. In this paper, some limitations of the existing similarity measures are identified (see Section 1). For example, there is not a unified framework for existing methods and existing approaches cannot compute similarity for two concepts that come from two different knowledge sources. To tackle these problems, this paper proposes an extensive study for semantic similarity of concepts from which a unified framework for semantic similarity computation is presented. Based on our framework, we give some generic and flexible approaches to semantic similarity measures resulting from instantiations of the framework. In particular, we obtain some new approaches to similarity measures that existing methods cannot deal

with by introducing multiple knowledge sources. The evaluation, based on three widely used benchmarks and five benchmarks developed in ourselves, sustains the intuitions with respect to human judgements. Some methods proposed in this paper have a good human correlation and constitute some effective ways of determining semantic similarity between concepts.

With the development of deep learning technology, in recent years semantic similarity measures can also be implemented by exploiting deep learning technologies such as long short-term memory (LSTM) deep learning methods and attention-based approaches combined with Word2Vec. As future works, we are planning to further explore semantic similarity computation by using deep learning technologies. In addition, we will theoretically and empirically investigate the unified framework issue of semantic relatedness for concepts. It is also desirable to apply our similarity measure approaches to text or short text search tasks (semantic search for texts or short texts).

Acknowledgements The authors would like to thank the anonymous referees for their valuable comments and suggestions which greatly improved the exposition of the paper. The works described in this paper are supported by The National Natural Science Foundation of China under Grant Nos. 61772210 and U1911201; Guangdong Province Universities Pearl River Scholar Funded Scheme (2018); The Project of Science and Technology in Guangzhou in China under Grant Nos. 201807010043 and 202007040006. Thanks for my students Rong Qu, Yongyi Fang, and Yudong Liu for their discussion, programming, and experiments.

References

1. Abid A, Rouached M, Messai N (2020) Semantic web service composition using semantic similarity measures and formal concept analysis. Multimed Tools Appl 79:6569–6597
2. Agirre E, Alfonseca E, Hall K, Kravalova J, Pasca M, Soroa A (2009) A study on similarity and relatedness using distributional and WordNet-based approaches. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Stroudsburg, pp 19–27
3. Alonso I, Contreras D (2016) Evaluation of semantic similarity metrics applied to the automatic retrieval of medical documents: an UMLS approach. Expert Syst Appl 44:386–399
4. Aouicha MB, Taieb MAH (2016) Computing semantic similarity between biomedical concepts using new information content approach. J Biomed Inform 59:258–275
5. Aouicha MB, Taieb MAH, Hamadou AB (2016) Taxonomy-based information content and wordnet-wiktionary-wikipedia glosses for semantic relatedness. Appl Intell 45(2):475–511
6. Baker T, Lamb D, Taleb-Bendiab A, Al-Jumeily D (2010) Facilitating semantic adaptation of web services at runtime using a meta-data layer. In: Proceedings of IEEE 2010 third international conference on Developments in eSystems Engineering (DESE 2010), IEEE, New York, pp 231–236
7. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, Clancy K, Courtot M, Derom D, Dumontier M, Fan L, Fostel J, Fragoso G, Gibson F, Gonzalez-Beltran A, Haendel MA, He Y, Heiskanen M, Hernandez-Boussard T, Jensen M, Lin Y, Lister AL, Lord P, Malone J, Manduchi E, McGee M, Morrison N, Overton JA, Parkinson H, Peters B, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Schober D, Smith B, Soldatova LN, Stoeckert CJ, Taylor CF, Torniai C, Turner JA, Vita R, Whetzel PL, Zheng J (2016) The ontology for biomedical investigations. PLoS One 11(4):e0154556
8. Batet M, Sanchez D, Valls A, Gibert K (2013) Semantic similarity estimation from multiple ontologies. Appl Intell 38(1):29–44
9. Bekhet S, Ahmed A (2020) Evaluation of similarity measures for video retrieval. Multimed Tools Appl 79: 6265–6278
10. Bizer C, Heath T, Berners-Lee T (2009) Linked data - the story so far. Int J Semant Web Inf Syst 5(3):1–22
11. Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S (2009) DBpedia - a crystallization point for the web of data. J Web Semant 7(3):154–165
12. Budanitsky A, Hirst G (2006) Evaluating WordNet-based measures of lexical semantic relatedness. Comput Linguist 32(1):13–47

13. Capuano A, Rinaldi AM, Russo C (2020) An ontology-driven multimedia focused crawler based on linked open data and deep learning techniques. Multimed Tools Appl 79:7577–7598
14. Church KW, Hanks P (1990) Word association norms, mutual information, and lexicography. Comput Linguist 16(1):22–29
15. Cilibrasi RL, Vitanyi PMB (2007) The Google similarity distance. IEEE Trans Knowl Data Eng 19(3):370–383
16. Coletti MH, Bleich HL (2001) Medical subject headings used to search the biomedical literature. J Am Med Inform Assoc 8(4):317–323
17. Couto FM, Silva MJ, Coutinho PM (2007) Measuring semantic similarity between gene ontology terms. Data Knowl Eng 61(1):137–152
18. Cross V, Yu X, Hu X (2013) Unifying ontological similarity measures: a theoretical and empirical investigation. Int J Approx Reason 54(7):861–875
19. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. J Am Soc Inf Sci 41(6):391–407
20. Fellbaum C (1998) WordNet: an electronic lexical database. Academic Press, Cambridge, MA
21. Ferreira R, Lins RD, Simske SJ, Freitas F, Riss M (2016) Assessing sentence similarity through lexical, syntactic and semantic analysis. Comput Speech Lang 39:1–28
22. Finkelstein L, Gabrilovich E, Matias Y, Rivlin E, Solan Z, Wolfman G, Ruppin E (2002) Placing search in context: the concept revisited. ACM Trans Inf Syst 20(1):116–131
23. Gabrilovich E, Markovitch S (2007) Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In: Proceedings of the 20th International Joint Conference on Artificial intelligence (IJCAI 2007). Morgan Kaufmann Publishers, San Francisco, CA, USA, pp 1606–1611
24. Gao JB, Zhang BW, Chen XH (2015) A WordNet-based semantic similarity measurement combining edge-counting and information content theory. Eng Appl Artif Intell 39:80–88
25. Garla VN, Brandt C (2012) Semantic similarity in the biomedical domain: an evaluation across knowledge sources. BioMed Central Bioinform 13(1):261–273
26. Gene Ontology Consortium (2004) The gene ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261
27. Goldstone RL (1994) The role of similarity in categorization: providing a groundwork. Cognition 52(2): 125–157
28. Hadj Taieb MA, Aouicha MB, Hamadou AB (2014) A new semantic relatedness measurement using WordNet features. Knowl Inf Syst 41(2):467–497
29. Hadj Taieb MA, Aouicha MB, Hamadou AB (2014) Ontology-based approach for measuring semantic similarity. Eng Appl Artif Intell 36:238–261
30. Halavais A, Lackaff D (2008) An analysis of topical coverage of Wikipedia. J Comput-Mediat Commun 13(2):429–440
31. Hamedani MR, Kim SW, Kim DJ (2016) SimCC: a novel method to consider both content and citations for computing similarity of scientific papers. Inf Sci 334-335:273–292
32. Harispe S, Sanchez D, Ranwez S, Janaqi S, Montmain J (2014) A framework for unifying ontology-based semantic similarity measures: a study in the biomedical domain. J Biomed Inform 48:38–53
33. Hirst G, St-Onge D (1998) Lexical chains as representations of context for the detection and correction of malapropisms. WordNet: An Electronic Lexical Database, The MIT Press, Cambridge, MA, pp 305–332
34. Jiang Y, Bai W, Zhang X, Hu J (2017) Wikipedia-based information content and semantic similarity computation. Inf Process Manag 53(1):248–265
35. Jiang JJ, Conrath DW (1997) Semantic similarity based on corpus statistics and lexical taxonomy. In: Proceedings of the 10th international conference on research on computational linguistics, The Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Taipei, pp 19–33
36. Jiang Y, Yang M, Qu R (2019) Semantic similarity measures for formal concept analysis using linked data and WordNet. Multimed Tools Appl 78:19807–19837
37. Jiang Y, Zhang X, Tang Y, Nie R (2015) Feature-based approaches to semantic similarity assessment of concepts using Wikipedia. Inf Process Manag 51(3):215–234
38. Lastra-Diaz JJ, Garcia-Serrano A (2015) A novel family of IC-based similarity measures with a detailed experimental survey on WordNet. Eng Appl Artif Intell 46:140–153
39. Leacock C, Chodorow M (1998) Combining local context and WordNet similarity for word sense identification. WordNet: An Electronic Lexical Database, The MIT Press, Cambridge, MA, pp 265–283
40. Lee D, Cornet R, Lau F, de Keizer N (2013) A survey of SNOMED CT implementations. J Biomed Inform 46(1):87–96
41. Li Y, Bandar ZA, McLean D (2003) An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans Knowl Data Eng 15(4):871–882

42. Lin D (1998) An information-theoretic definition of similarity. In: Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998). Morgan Kaufmann Publishers, San Francisco, CA, USA, pp 296–304
43. Liu H, Bao H, Xu D (2012) Concept vector for semantic similarity and relatedness based on WordNet structure. J Syst Softw 85(2):370–381
44. Liu YH, Wacholder N (2017) Evaluating the impact of MeSH (medical subject headings) terms on different types of searchers. Inf Process Manag 53(4):851–870
45. Maarek YS, Berry DM, Kaiser GE (1991) An information retrieval approach for automatically constructing software libraries. IEEE Trans Softw Eng 17(8):800–813
46. Maguitman AG, Menczer F, Erdinc F, Roinestad H, Vespignani A (2006) Algorithmic computation and approximation of semantic similarity. World Wide Web 9(4):431–456
47. Martinez-Gil J (2014) An overview of textual semantic similarity measures based on web intelligence. Artif Intell Rev 42(4):935–943
48. Medelyan O, Milne D, Legg C, Witten IH (2009) Mining meaning from Wikipedia. Int J Hum Comput Stud 67(9):716–754
49. Meng L, Gu J, Zhou Z (2012) A new model of information content based on concept's topology for measuring semantic similarity in WordNet. Int J Grid Distribute Comput 5(3):81–93
50. Meng L, Huang R, Gu J (2014) Measuring semantic similarity of word pairs using path and information content. Int J Future Generation Commun Netw 7(3):183–194
51. Meymandpour R, Davis JG (2016) A semantic similarity measure for linked data: an information content-based approach. Knowl-Based Syst 109:276–293
52. Miller GA, Charles WG (1991) Contextual correlates of semantic similarity. Lang Cogn Process 6(1):1–28
53. Nosofsky RM (1992) Similarity scaling and cognitive process models. Annu Rev Psychol 43(1):25–53
54. Oliva J, Serrano JI, del Castillo MD, Iglesias A (2011) SyMSS: a syntax-based measure for short-text semantic similarity. Data Knowl Eng 70(4):390–405
55. Ou W, Xuan R, Gou J, Zhou Q, Cao Y (2020) Semantic consistent adversarial cross-modal retrieval exploiting semantic similarity. Multimed Tools Appl 79:14733–14750
56. Pedersen T, Pakhomov SVS, Patwardhan S, Chute CG (2007) Measures of semantic similarity and relatedness in the biomedical domain. J Biomed Inform 40(3):288–299
57. Pellegrin L, Escalante HJ, Montes-y-Gomez M, Gonzalez FA (2019) Exploiting label semantic relatedness for unsupervised image annotation with large free vocabularies. Multimed Tools Appl 78:19641–19662
58. Petrakis EGM, Varelas G, Hliaoutakis A, Raftopoulou P (2006) X-similarity: computing semantic similarity between concepts from different ontologies. J Digit Inf Manag 4(4):233–237
59. Pilehvar MT, Navigli R (2015) From senses to texts: an all-in-one graph-based approach for measuring semantic similarity. Artif Intell 228:95–128
60. Pirro G (2009) A semantic similarity metric combining features and intrinsic information content. Data Knowl Eng 68(11):1289–1308
61. Ponzetto SP, Strube M (2007) Knowledge derived from Wikipedia for computing semantic relatedness. J Artif Intell Res 30:181–212
62. Rada R, Mili H, Bicknell M, Blettner E (1989) Development and application of a metric on semantic nets. IEEE Trans Syst Man Cybern 19(1):17–30
63. Resnik P (1995) Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of International Joint Conference for Artificial Intelligence (IJCAI 1995). Morgan Kaufmann Publishers, San Francisco, CA, USA, pp 448–453
64. Resnik P (1999) 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
65. Rodriguez MA, Egenhofer MJ (2003) Determining semantic similarity among entity classes from different ontologies. IEEE Trans Knowl Data Eng 15(2):442–456
66. Rubenstein H, Goodenough J (1965) Contextual correlates of synonymy. Commun ACM 8(10):627–633
67. Safyan M, Qayyum ZU, Sarwar S, Garcia-Castro R, Ahmed M (2019) Ontology-driven semantic unified modelling for concurrent activity recognition (OSCAR). Multimed Tools Appl 78:2073–2104
68. Samih H, Rady S, Gharib TF (2020) Enhancing image retrieval for complex queries using external knowledge sources. Multimed Tools Appl 79:27633–27657
69. Sanchez D, Batet M (2011) Semantic similarity estimation in the biomedical domain: an ontology-based information-theoretic perspective. J Biomed Inform 44(5):749–759
70. Sanchez D, Batet M (2012) A new model to compute the information content of concepts from taxonomic knowledge. Int J Semant Web Inf Syst 8(2):34–50
71. Sanchez D, Batet M (2013) A semantic similarity method based on information content exploiting multiple ontologies. Expert Syst Appl 40(4):1393–1399

72. Sanchez D, Batet M, Isern D (2011) Ontology-based information content computation. Knowl-Based Syst 24(2):297–303
73. Sanchez D, Batet M, Isern D, Valls A (2012) Ontology-based semantic similarity: a new feature-based approach. Expert Syst Appl 39(9):7718–7728
74. Sarwar S, Qayyum ZU, Garcia-Castro R, Safyan M, Munir RF (2019) Ontology based E-learning framework: a personalized, adaptive and context aware model. Multimed Tools Appl 78:34745–34771
75. Seco N, Veale T, Hayes J (2004) An intrinsic information content metric for semantic similarity in WordNet. In: Proceedings of the 16th European Conference on Artificial Intelligence (ECAI), IOS Press, Amsterdam, pp 1089–1094
76. Shepard RN (1962) The analysis of proximities: multidimensional scaling with an unknown distance function. I Psychometrika 27(2):125–140
77. Staab S, Studer R (2009) Handbook on Ontologies. Springer, Second Edition
78. Strube M, Ponzetto SP (2006) WikiRelate! Computing semantic relatedness using Wikipedia. In: Proceedings of the 21st national conference on artificial intelligence (AAAI 2006), AAAI Press, Cambridge, pp 1419-1424
79. Suchanek FM, Kasneci G, Weikum G (2008) YAGO: a large ontology from Wikipedia and WordNet. J Web Semant 6(3):203–217
80. Tversky A (1977) Features of similarity. Psychol Rev 84(4):327–352
81. Wolk K, Wolk A (2017) Machine enhanced translation of the human phenotype ontology project. Procedia Comput Sci 121:11–18
82. Wu Z, Palmer M (1994) Verb semantics and lexical selection. In: Proceedings of the 32nd annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Stroudsburg, PA, pp 133–138
83. Zhou Z, Wang Y, Gu J (2008) A new model of information content for semantic similarity in WordNet. In: Proceedings of second international conference on Future Generation Communication and Networking Symposia (FGCNS 2008), IEEE, New York, pp 85–89