Measuring Semantic Similarity Between Biomedical Concepts Within Multiple Ontologies

Hisham Al-Mubaid, *Member, IEEE*, and Hoa A. Nguyen, *Member, IEEE*

Abstract—Most of the intelligent knowledge-based applications contain components for measuring semantic similarity between terms. Many of the existing semantic similarity measures that use ontology structure as their primary source cannot measure semantic similarity between terms and concepts using multiple ontologies. This research explores a new way to measure semantic similarity between biomedical concepts using multiple ontologies. We propose a new ontology-structure-based technique for measuring semantic similarity in single ontology and across multiple ontologies in the biomedical domain within the framework of Unified Medical Language System (UMLS). The proposed measure is based on three features: 1) cross-modified path length between two concepts; 2) a new feature of common specificity of concepts in the ontology; and 3) local granularity of ontology clusters. The proposed technique was evaluated relative to human similarity scores and compared with other existing measures using two terminologies within UMLS framework: Medical Subject Headings and Systemized Nomenclature of Medicine Clinical Term. The experimental results validate the efficiency of the proposed technique in single and multiple ontologies, and demonstrate that our proposed measure achieves the best results of correlation with human scores in all experiments.

Index Terms—Biomedical information retrieval, biomedical ontology, biomedical terminology, semantic similarity, Unified Medical Language System (UMLS).

I. Introduction

SEMANTIC similarity techniques are interested in measuring the semantic similarity, or inversely, semantic distance between two concepts according to a given ontology. Semantic similarity techniques are becoming important components in most intelligent knowledge-based and semantic information retrieval (SIR) systems [1], [14], [19]. For example, in SIR, semantic similarity techniques play a crucial role in determining an optimal match between query terms and documents. Measures of semantic similarity are also used in broad applications such as sense disambiguation [1], [28], information extraction and retrieval [3], [19], classification and ranking, detection of redundancy, and detection and correction of malapropisms [1], [3], [14]–[16]. In bioinformatics, semantic similarity techniques can be of great benefit in integrating multiple information sources for knowledge discovery and information extraction applications [14].

Semantic similarity is the inverse of semantic distance, such that, if two concepts C_1 and C_2 belong to two different nodes $n1$ and $n2$ in a given ontology or terminology, then the distance between the nodes ($n1$ and $n2$) determines the similarity between these two concepts (i.e., C_1 and C_2). Each node, in an ontology (also called concept node), contains a set of synonymous terms, and hence, two terms are synonymous if they are in the same node. If the two terms are synonymous, then their semantic similarity is maximum, and one term can be represented by one or more concept nodes.

The semantic similarity measuring techniques can be roughly classified into two classes: 1) structure-based measures are the measures that use ontology hierarchy structure (e.g., is-a, part-of) to compute the semantic similarity between terms [7], [14], [17], [25] and 2) information content (IC) based measures are those measures that use IC of concepts derived from corpus statistics to measure the semantic similarity between terms.

The primitive approach to find the semantic similarity between two terms in ontology is to find the shortest path length between them in the ontology (shortest path length) giving that the links are of type is-a/part-of. Rada *et al.* [17] proposed path length as a potential measure of semantic distance/similarity in the biomedical domain within the Medical Subject Headings (MeSH) terminology [5], [11], [23], [24], [26]. After Rada *et al.*'s approach, a number of approaches have been developed using ontology/terminology as primary information source, and were mostly applied in the general English domain using, for example, WordNet [13]. However, most of these similarity techniques cannot measure semantic similarity between terms in multiple ontologies to solve the problem of missing terms in a single ontology, or to allow for source integration.

An ontology represents an effective means of knowledge sharing within controlled and structured vocabulary [21]. Many ontologies have been developed in the past few decades for various purposes and domains [3], [12], [15], [21]. For example, WordNet [13] is a lexical database for general English covering most of the general English concepts and supporting various purposes. In the biomedical domain, the Unified Medical Language System (UMLS) framework [5] includes many biomedical ontologies and terminologies (e.g., MeSH, Systemized Nomenclature of Medicine Clinical Term (SNOMED-CT), ICD family [5], [6], [11], [23], [24]). Moreover, within one domain, one concept may exist in one ontology but is missing in another ontology. For example, in Fig. 1, the biomedical concept "stomach cramps" exists in SMOMED-CT but not in MeSH although its parent concept "abdominal pain" exists in both ontologies. In this case, such terms that are missing from the underlying ontology will not be measured for similarity and will be skipped;
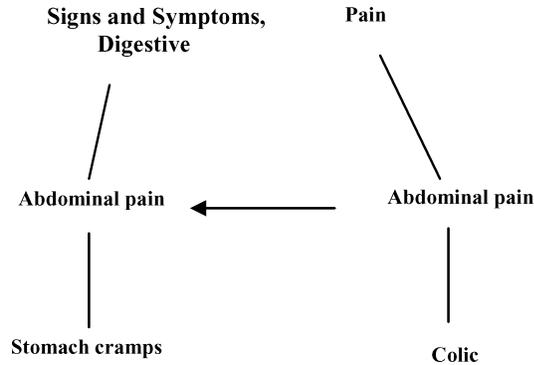
Fig. 1. Two fragments of (left) SNOMED-CT and (right) MeSH.

see, for example, [14] and [16]. Therefore, we would like to explore a way of measuring semantic similarity between terms (such as "stomach cramp" and "colic" in Fig. 1) within multiple ontologies. The existing ontology-based semantic similarity measures cannot handle this case.

In this paper, first, we investigate the basic ontology-based semantic similarity features, and then, we propose new semantic features [1] as well as new semantic similarity techniques. The proposed techniques can measure the similarity between concepts in single ontology and across multiple ontologies in the biomedical domain. We also present a strategy for combining and selecting ontologies in measuring the similarity of terms dispersed across multiple ontologies. The proposed technique relies on one type of relations, i.e., is-a, and all links between nodes have the same weight of 1. We evaluated the proposed measures along with other existing similarity measures that we adapted from general English into the biomedicine field. We used datasets of biomedical term pairs scored for similarity by human experts. The experimental evaluations were conducted relative to correlation with human scores using the MeSH and SNOMED-CT within UMLS framework [6], [11], [23]. The experimental results with the multiple datasets and with multiple ontologies/terminologies confirmed the efficiency of the proposed measure.

In the biomedical and bioinformatics domains, a term is distinguished from a concept in that the "term" is a word form or a lexical form that exists in the (biomedical) texts (like research articles) whereas the "concept" is a node in an ontology or in a controlled/structured vocabulary like MeSH [21]. In the context of measuring semantic similarity between terms, we assume that each term is mapped into one or more concepts (i.e., nodes) in the given set of ontologies.

The biomedical field boasts the availability of huge volumes of data and information resources. Therefore, identifying and measuring similarity between biomedical concepts and terms in these resources will help greatly in utilizing and integrating information sources, for example, in intelligent knowledge discovery and information retrieval applications.

The UMLS project started at the National Library of Medicine (NLM) in 1986 with one of the objectives being to help in-

[1]The proposed semantic similarity measure is available upon request from the first author.

terpret and understand medical meanings across systems [5], [6], [11], [23]. It consists of three main knowledge sources: Metathesaurus, Semantic Network, and SPECIALIST Lexicon and Lexical Tools [5]. Metathesaurus consists of more than one million concepts from over 130 sources, and supports 17 languages [5]. MeSH [5], [11], [23], [24], [26] is one of the core source vocabularies used in UMLS with the primary purpose of supporting indexing, cataloguing, and retrieval of more than 16 million medical literature articles stored in NLM Medline database [10]. SNOMED-CT [5], [6], [23], [24] was included in UMLS in May 2004 (2004AA) [5]. SNOMED-CT is a comprehensive clinical ontology with coverage of diseases, clinical findings, and procedures, and includes concepts, terms, and relationships to represent clinical information. Each node in MeSH is a main heading, which is a concept that belongs to a descriptor in MeSH database [26]. A descriptor is often broader than a single concept, and so, it may consist of a class of concepts.

II. BACKGROUND

Several semantic similarity measures have been proposed in the past few decades. Most of these measures were developed for, and applied into, WordNet [13]. WordNet was developed at Princeton University and includes English nouns, verbs, adjectives, and adverbs organized into synonym sets. In WordNet 2.0, there are nine noun taxonomies with an average depth of 13, and 554 verb taxonomies with an average depth of 2. For more details on WordNet, see [13]. The following semantic similarity measures are based on WordNet and can be roughly classified into two groups as follows.

A. Structure-Based Measures

Most of the measures that are based on the hierarchy structure of ontology are actually based on: 1) path length (i.e., shortest path length/distance between the two concept nodes) and 2) depth of concept nodes in the ontology. Some of the measures that are based on WordNet include Path length [17], Wu and Palmer [25], and Leacock and Chodorow [7].

B. Information-Content-Based Measures

These measures use IC of concepts, sometimes with ontology structure, as their information source. Some of the WordNet-based measures include Resnik [18], Jiang and Conrath [4], and Lin [9].

C. Semantic Similarity in the Biomedical Domain

Rada *et al.* [17] proposed semantic distance as a potential measure for semantic similarity between two concepts in MeSH. More recently, Caviedes and Cimino [2] implemented the shortest path length measure, called CDist, based on the shortest distance between two concept nodes in the ontology. They evaluated CDist on UMLS Metathesaurus (MeSH, SNOMED, ICD9 [5], [6], [11], [23], [24], [26]), and then compared the CDist similarity scores to human expert scores by correlation coefficients. Another recent work on semantic similarity and relatedness in the biomedicine domain is by Pedersen *et al.* [16]
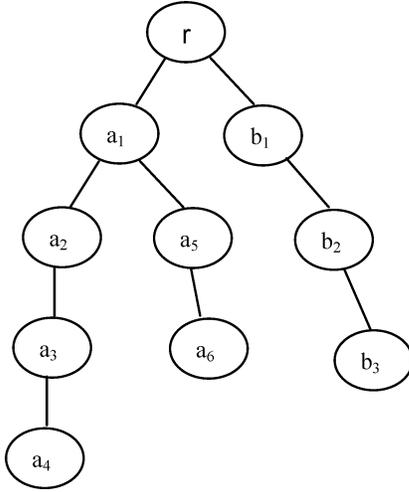
Fig. 2.   Fragment of two clusters in ontology.

in which they applied a corpus-based context vector approach to measure similarity between concepts in SNOMED-CT. Their context vector approach is ontology-free but requires training text, for which, they used text data from Mayo Clinic corpus of medical notes [16], [28].

III. Ontology-Based Semantic Similarity Features

A. Path Length and Common Specificity Features

Path-length-based measures, such as path length [2], [17] and Leacock and Chodorow [7], do not account for the depth of the nodes and give the same similarity value for two pairs of nodes $\{(n1, n2), (n3, n4)\}$ if PathLength($n1$, $n2$) = PathLength($n3$, $n4$) regardless of the depth of the nodes or the depth of their least common subsumer (LCS). The LCS of two nodes $n1$ and $n2$ is the lowest node in the tree that can subsume both $n1$ and $n2$. For example, in Fig. 2, LCS(a_3, a_5) = a_1 and LCS(a_2, b_2) = r. In addition, the measure of Wu and Palmer [25] uses only depth of concept nodes. We, however, want to combine both features, path length and depth, in one measure, i.e., we will take the specificity of concept nodes into account by utilizing the depth feature of the concepts. The LCS of two concept nodes determines the common specificity of two concepts, according to ontology structure only. Furthermore, local density such as link strength/weight also affects the similarity. One way of measuring the local density is using IC of concepts based on corpus statistic [1], [8], [18]. And since there is no standard corpus in biomedical domain, we use only ontology-based features as properties of semantic similarity. The proposed method does not require ontology (essentially) with formal semantic relations between terms; it can be applied onto any terminology structure, structured vocabulary, or, in general, directed acyclic graph (DAG) with links between the nodes.

B. Local Granularity of Clusters and Local Specificity of Concepts

The existing semantic similarity measures do not take into account the local granularity of clusters (i.e., subtrees, category

trees, and taxonomy) containing the concept nodes. We examine the local specificity of a concept node by considering the cluster containing that concept node. The following example explains the effect of cluster on concept specificity. We use the term "cluster" in this paper to refer to a subtree or taxonomy tree in the ontology.

Let us consider the fragment in Fig. 2 that has two (left and right) clusters containing concepts ai and bi with depths of 4 and 3, respectively. The roots of the two clusters are a_1 and b_1, while r is the global root. Define the specificity, spec(c), of concept c in a cluster C as follows:

$$\text{spec}(c) = \frac{\text{depth}(c)}{\text{depth}C} \tag{1}$$

where depth(c) is the depth of concept c and depth$C$ is the depth of cluster C using node counting. We note that spec(c) = 1 when the concept c is a *leaf* node in the cluster (i.e., depth(c) = depth$C$). Following (1), specificity of a_2 and b_2, in Fig. 2, is calculated as follows:

$$\text{spec}(a_2) = \frac{2}{4} = 0.50 \qquad \text{spec}(b_2) = \frac{2}{3} = 0.67.$$

i.e., the local specificity of concept b_2 (0.67) is more than that of concept a_2 (0.50) even though both have the same depth. Thus, b_2 has more specificity (i.e., more specialized) within its cluster than a_2 as it lies further down, toward the bottom, in its cluster. Therefore, the local granularity of clusters affects the similarity scale between concepts in different clusters. Thus, the granularity of clusters should be taken into account and utilized as a contributing feature in measuring semantic similarity.

IV. Cross-Ontology Semantic Similarity

Within the framework of UMLS, the concepts are dispersed in many ontologies and terminologies, and these ontologies overlap in many sets of UMLS concepts. Some concepts (e.g., "colic" in Fig. 1) related to a concept in a given ontology (e.g., SNOMED-CT) may not be seen in that ontology, as shown in Fig. 1; however, these related concepts exist in other ontologies (e.g., MeSH) in the framework. The issue that stands out here is that the ontologies within a framework like UMLS have different granularity degrees, and so, each ontology reflects a different similarity scale. For measuring cross-ontology similarity of concepts, the granularity of ontologies should be taken into account. Rodriguez and Egenhofer proposed a method for determining the most similar (or equivalent) entity (concept) classes in two ontologies [20]. Their method also ranks the similarity of a given concept to a set of concepts based on two ontologies [20]. The method finds similar classes based on matching common features in the two ontologies. In the evaluations, precision and recall were calculated based on number of similar entity classes that their method finds versus number of actual similar classes [20]. In this paper, however, we want to measure the semantic similarity between two terms within two ontologies O1 and O2 having some common (identical) nodes (see Figs. 1 and 3).
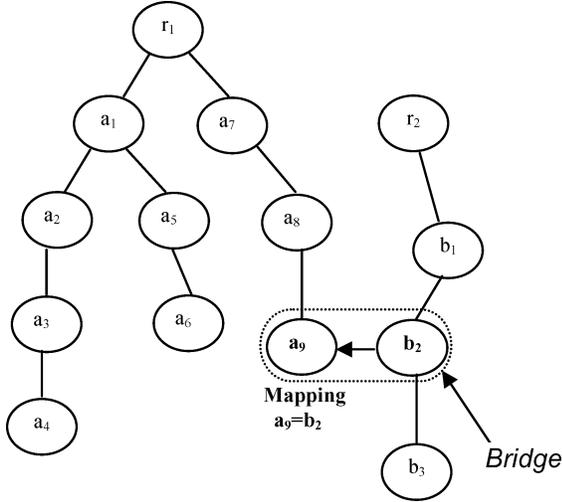
Fig. 3.   Connecting two ontology fragments.

A. Adapted Common Specificity Feature

For cross-ontology semantic similarity, we adapt the concept specificity feature (Section III-B) for cross-ontology. In a previous work, we investigated this specificity feature with cluster granularity within a cluster-based approach [12]. The specificity feature takes into account the depth of the LCS of two concepts and the depth of ontology. The common specificity, CSpec, of two concept nodes C_1 and C_2 can be measured by finding the depth of their LCS node, and then scaling it by the depth D of the ontology as follows:

$$\mathrm{CSpec}(C_1,\ C_2) = D - \mathrm{Depth}(\mathrm{LCS}(C_1,\ C_2)).   \quad (2)$$

Thus, the CSpec feature determines the common specificity of two concepts in the ontology. The less the CSpec value of two concepts, the more they have "shared information," and thus the more they are similar. For example, in Fig. 2, $\mathrm{CSpec}(a_2, a_5) = 4 - 1 = 3$ and $\mathrm{CSpec}(a_2, b_2) = 4 - 0 = 4$.

B. Proposed Cross-Ontology Approach

1) Rules and Assumptions: As each ontology has a different granularity degree affecting its similarity scale, we present the intuitive rules of cross-ontology semantic similarity as follows.
  1) The semantic similarity scale system reflects the degree of similarity of pairs of concepts comparably in single ontology or in cross-ontology. This rule ensures that the mapping of one ontology OB (called secondary ontology) to another ontology OA (called primary ontology) does not deteriorate the similarity scale of the primary ontology.
  2) The semantic similarity must obey local ontology's similarity rules as follows.
      a) The shorter the distance between two concept nodes in the ontology, the more they are similar.
      b) Lower level pairs of concept nodes are semantically closer (more similar) than higher level pairs (i.e., the more the two concept nodes share information, the more similar they are).

  c) The maximum similarity of two concepts is reached when they are in the same node in the ontology.

2) Single Ontology Similarity: Within the single ontology, we do not consider the granularity, and hence, we use the path length and depth features to get semantic distance of two concepts as follows:

$$\mathrm{SemDist}(C_1,C_2) = \log((\mathrm{Path} - 1)^{\alpha} \times (\mathrm{CSpec})^{\beta} + k)   \quad (3)$$

where $\alpha > 0$ and $\beta > 0$ are contribution factors of two features (Path and CSpec), k is a constant, Path is the shortest path length between the two concept nodes, and CSpec feature is calculated as in (2). We use logarithm function (inverse of exponentiation) for semantic distance (3), which is the inverse of semantic similarity. Shepard [29] derived an exponential-decay function as a form of universal law of stimulus generalization for psychological sciences based on which the assumption can be true that logarithm functions are appropriate for semantic distance [8].

3) Cross-Ontology Semantic Similarity: In cross-ontology semantic similarity between two concepts (C_1 and C_2), there are four cases that depend on whether or not the two concepts (C_1 and C_2) occur in one ontology (i.e., if C_1 and C_2 do not exist in one ontology, then we must use two ontologies in the measure), and whether they occur in the *primary* or *secondary* ontology (the proposed cross-ontology measure requires that one of the ontologies be designated as "*primary*" and the rest are *secondary* ontologies; determining primary and secondary ontologies is discussed in Section IV-C). The four cases are as follows.

Case 1 (Similarity within the primary ontology): If the two concepts occur in the *primary* ontology, then the similarity in this case is treated as similarity within single ontology using (3) given before.

Case 2 [Cross-ontology similarity (primary–secondary)]: The common specificity feature: In this case, the two concepts belong to two different ontologies (primary and secondary). The secondary ontology is connected to the primary ontology by joining the common nodes (e.g., a_9 and b_2 in Fig. 3) of two ontologies. However, two ontologies may have many common or equivalent nodes. Two nodes in two ontologies are equivalent if they refer to the same concept. For example, in Fig. 3, suppose that b_2 and a_9 refer to the same concept (i.e., $b_2 = a_9$), then we merge b_2 and a_9 into one node called *bridge*, as in Fig. 3. Thus, since there can be more than one bridge node when mapping two ontologies, there can be more than one LCS node ($\{\mathrm{LCS}_n\}$) for the two concepts. The LCS node of two concept nodes (C_1, C_2) belonging to two ontologies is the LCS of the first node C_1 in primary ontology and the bridge node, i.e.,

$$\mathrm{LCS}_n(C_1, C_2) = \mathrm{LCS}(C_1, \mathrm{bridge}_n)   \quad (4)$$

such that C_1 belongs to the primary ontology while C_2 belongs to the secondary ontology. The path length between two concept nodes in two ontologies passes through two ontologies having different granularity degrees. The part of path length in secondary ontology is then converted (scaled) into primary ontology's scale of path feature as explained next.

The cross-ontology path length feature: The path length between two concept nodes is calculated by adding up two path

lengths from each of them to *bridge* node. For example, the path length between a_3 and b_3 in Fig. 3 is calculated as follows:

$$\text{Path}(a_3, b_3) = d_1 + d_2 - 1 \qquad (5)$$

such that $d_1 = d(a_3, \text{bridge})$ and $d_2 = d(b_3, \text{bridge})$, where $d(a_3, \text{bridge})$ is the shortest path length from a_3 to the bridge, and similarly for $d(b_3, \text{bridge})$ (in this example, $d_1 = 7$ and $d_2 = 2$). In this case, bridge is counted twice because of using node counting approach, so 1 is subtracted in (5). Note that path lengths d_1 and d_2 are in different scales. Let us assume the first ontology (which contains *ai*) to be the primary ontology, and thus, the second ontology (which contains *bi*) is the secondary ontology. We scale the path length and CSpec features in the secondary ontology to the primary ontology's scale. We define PathRate and CSpecRate as follows. The granularity ratio of the primary ontology over the secondary ontology for the common specificity feature is

$$\text{CSpecRate} = \frac{D_1 - 1}{D_2 - 1} \qquad (6)$$

where D_1 and D_2 are the depth of primary ontology and secondary ontology, respectively [hence $(D_1 - 1)$ and $(D_2 - 1)$ are the maximum common specificity values of the primary and secondary ontologies, respectively; see (2)]. The granularity ratio of the primary ontology over the secondary ontology for the path feature is given by

$$\text{PathRate} = \frac{2D_1 - 1}{2D_2 - 1} \qquad (7)$$

where $(2D_1 - 1)$ and $(2D_2 - 1)$ are the maximum path values of two concept nodes in the primary and secondary ontologies, respectively. Following our rules (in Section IV-B1), d_2 [in (5)] in the secondary ontology needs to be scaled to the primary ontology as follows:

$$d_2' = \text{PathRate} \times d_2. \qquad (8)$$

This new path length d_2' reflects the path length of the second concept node to the bridge node relative to the primary ontology granularity scale. Applying (8), the cross path length between the two nodes in primary ontology scale of path feature is given as follows:

$$\text{Path}(C_1, C_2) = d_1 + \text{PathRate} \times d_2 - 1. \qquad (9)$$

And so,

$$\text{Path}(C_1, C_2) = d_1 + \frac{2D_1 - 1}{2D_2 - 1} \times d_2 - 1. \qquad (10)$$

Recall that there can be more than one bridge node; therefore, there can be more than one path length between the two concepts ($\{\text{Path}_n\}$). Finally, the semantic distance, SemDist, between two concept nodes is given as follows:

$$\text{CSpec}_i(C_1, C_2) = D_1 - \text{Depth}(\text{LCS}(C_1, \text{Bridge}_i)) \qquad (11)$$

$$\text{SemDist}_i(C_1, C_2) = \log((\text{Path}_i - 1)^\alpha \times (\text{CSpec}_i)^\beta + k) \qquad (12)$$

$$\text{SemDist}(C_1, C_2) = \text{MIN}_q\{\text{SemDist}_q(C_1, C_2)\} \qquad (13)$$

where D_1 is the depth of the primary ontology and Path_i is the path length of two concepts calculated via Bridge_i. The semantic distance between two concepts is chosen as the minimum among all possible paths.

Case 3 (Similarity within single secondary ontology): This is the case when both concepts are in a single secondary ontology. Then, the semantic distance features in this case must be converted into primary ontology scales of two features as follows:

$$\text{Path}(C_1, C_2) = \text{Path}(C_1, C_2)_{\text{secondary}} \times \text{PathRate} \qquad (14)$$

$$\text{CSpec}(C_1, C_2) = \text{CSpec}(C_1, C_2)_{\text{secondary}} \times \text{CSpecRate} \qquad (15)$$

$$\text{SemDist}(C_1, C_2) = \log((\text{Path} - 1)^\alpha \times (\text{CSpec})^\beta + k) \qquad (16)$$

where $\text{Path}(C_1, C_2)$secondary and $\text{CSpec}(C_1, C_2)$secondary are the Path and CSpec between C_1 and C_2 in the secondary ontology, and PathRate and CSpecRate are computed in (7) and (6), respectively.

Case 4 (Similarity within multiple secondary ontologies): The fourth case is when the two concepts are in two different secondary ontologies (i.e., none of them exists in the primary ontology). In this case, one of the two secondary ontologies acts momentarily (temporarily) as a primary to calculate the semantic features (viz., Path and CSpec) using case 2 before. Then, the semantic similarity is computed using case 3 to scale the features (again) to the scale level of the primary ontology.

C. Choosing the Secondary Ontologies

First, among multiple ontologies, the one with the highest granularity will be chosen as the primary ontology. Now, which ontology is chosen as the secondary ontology? In case 2 (i.e., first concept in the primary ontology while the other concept is in a secondary ontology), the second concept may belong to several ontologies in the unified framework. Assume we have one designated primary ontology and a number of candidate secondary ontologies, then our proposed strategy for choosing the secondary ontology is based on two observations. The first observation is that the more the two ontologies overlap, the better the similarity of two concepts dispersed in these two ontologies. Thus, among the candidate secondary ontologies, we select the one that has the most overlapping with the primary ontology. The second observation is that the secondary ontology should be chosen as the one that has the highest granularity degree. For that, we proposed a metric to measure the goodness of choosing a secondary ontology. The higher the goodness value, the better it is chosen as the secondary ontology. The metric is as follows:

$$\text{goodness}(\text{Op}, \text{Os}) = \frac{|\text{Op} \cap \text{Os}|}{|\text{Op} \cup \text{Os}|} \times \frac{\text{Ds}}{\text{Dp}} \qquad (17)$$

where Op is the primary ontology and Os is the secondary ontology that is examined for goodness as secondary ontology. $\text{Op} \cap \text{Os}$ is the set of common concepts of two ontologies. $\text{Op} \cup \text{Os}$ is the union of two sets of concepts of two ontologies.

Ds and Dp are the depth of primary ontology and secondary ontology, respectively.

In case 4 before, when the two concepts (C_1 and C_2) belong to two different secondary ontologies, there can be multiple secondary ontologies containing the concepts. Therefore, the problem is which secondary ontology is chosen for each of the two concepts? First, for the first concept C_1, among the ontologies containing C_1, the ontology with the highest granularity (say Ox) is chosen for C_1. Then, the goodness metric (17) is used to determine the best ontology for the second concept C_2 (note: when applying (17) to determine the second ontology, we treat the first ontology Ox chosen for the first concept as a primary ontology).

V. EXPERIMENTS AND RESULTS

There are a number of evaluation methodologies to assess the accuracy of similarity values computed by a given similarity measure [1]. One of the methodologies is to employ the similarity measure in an application that requires similarity between words like word sense disambiguation, finding malapropisms, or information retrieval [1], [3], [7], [14], [19]. Another evaluation methodology is to compare the computed similarity values of the measure against the human (expert) similarity scores using, for example, correlation coefficient. The latter methodology requires a dataset of term pairs scored for similarity by human experts. In this paper, we use the second methodology to evaluate the proposed measure as it is more common in this task [1], [3], [7], [8], [14].

A. Datasets

Dataset 1: In the biomedical domain, there are no standard human rating sets of terms for semantic similarity like the Miller and Charles (MC) or Rubenstein and Goodenough (RG) datasets in general English [1]. Thus, to evaluate our methods, we used the set of 30 concept pairs from [16], which was annotated by three physicians and nine medical index experts. Each pair was annotated on a four-point scale: practically synonymous, related, marginally related, and unrelated. The average correlation between physicians is 0.68, and between experts is 0.78. Because the experts are more than the physicians (nine experts versus three physicians) and the correlation (agreement) between experts (0.78) is stronger than the correlation between physicians (0.68), we will assume that the experts rating scores are more reliable than the physician rating scores.

Dataset 2: We used another biomedical dataset containing 36 MeSH term pairs [3]. The human scores in this dataset are the average evaluated scores of reliable doctors.

Dataset 3: Furthermore, we combined dataset 1 and dataset 2 into one dataset; we call it dataset 3 containing 66 biomedical term pairs.

We used the UMLS Knowledge Source (UMLSKS) browser [24] for SNOMED-CT and MeSH browser [11] for MeSH.

RG Dataset: The RG dataset of term pairs contains 65 general English term pairs. RG was collected by Rubenstein and Goodenough [30], and covers 51 subjects containing 65 pairs of words on a scale from "highly synonymous" to "semantically unrelated" [1], [30].

TABLE I
ABSOLUTE CORRELATIONS WITH HUMAN SCORES FOR FOUR MEASURES
USING SNOMED-CT ON DATASET 1, DATASET 2, AND DATASET 3

Measure	SNOMED-CT		
	Dataset 1 correl (rank)	Dataset 2 correl (rank)	Dataset 3 correl (rank)
SemDist (proposed)	0.665 (1)	0.735 (1)	0.726 (1)
Leacock and Chodorow	0.431 (2)	0.677 (3)	0.600 (2)
Wu and Palmer	0.296 (3)	0.686 (2)	0.498 (3)
Path length	0.254 (4)	0.586 (4)	0.422 (4)
Average	**0.412**	**0.671**	**0.562**

To evaluate the approach in cross-ontology, a dataset containing term pairs as in cases 2–4 should be used. For example, concept pairs like (C_1, C_2), such that C_1 belongs to ontology O1 whereas C_2 exists in ontology O2 and both ontologies are in the unified framework, should be used for testing. Since there is no such dataset with human ratings, we simulated this case and combined different datasets to make sure that there are pairs [e.g., (C_1, C_2)] for which we must use two ontologies (cross-ontology) to measure their similarity. For that, dataset 1, dataset 2, and RG dataset [1] were used for experiments.

B. Tools and Information Sources

We implemented the proposed measures using the C# programming language. WordNet 2.0 was used as primary ontology while MeSH and SNOMED-CT as secondary ontologies. The Perl module WordNet::Similarity [22] was used in our implementation. MeSH database and MeSH browser (http://www.nlm.nih.gov/mesh/meshhome.html) were used to get information on the terms/concepts in MeSH, and UMLSKS browser (http://umlsks.nlm.nih.gov) was used to get information on the terms in SNOMED-CT. The MeSH browser is a convenient application with a Web-based interface that allows the users to do online search and get all the information in an organized and neat way from the MeSH database. The UMLSKS browser is an Internet application that provides the users with flexibility to get and browse through the databases within the UMLS. For example, it allows the user to request information about particular Metathesaurus concepts, including attributes such as the concept definition, its semantic types, concepts that are related to it, hierarchical context details, and more.

C. Experimental Results

In single ontology, the approach performed very well surpassing other existing measures in biomedical domain. Tables I and II show the results of correlations (Pearson) with human scores for our proposed measure (with default parameters $\alpha = 1$, $\beta = 1$, $k = 1$) using the three datasets, experimented on MeSH and SNOMED-CT, and compared with three other measures: Path length, Wu and Palmer, and Leacock and Chodorow [7], [17], [25].

TABLE II
ABSOLUTE CORRELATIONS WITH HUMAN SCORES FOR FOUR MEASURES
USING MeSH ON DATASET 1, DATASET 2, AND DATASET 3

Measure	MeSH		
	Dataset 1 correl (rank)	Dataset 2 correl (rank)	Dataset 3 correl (rank)
SemDist (proposed)	0.863 (1)	0.825 (1)	0.841 (1)
Leacock and Chodorow	0.857 (2)	0.820 (2)	0.836 (2)
Wu and Palmer	0.794 (3)	0.811 (3)	0.808 (3)
Path Length	0.744 (4)	0.765 (4)	0.764 (4)
Average	**0.815**	**0.805**	**0.812**
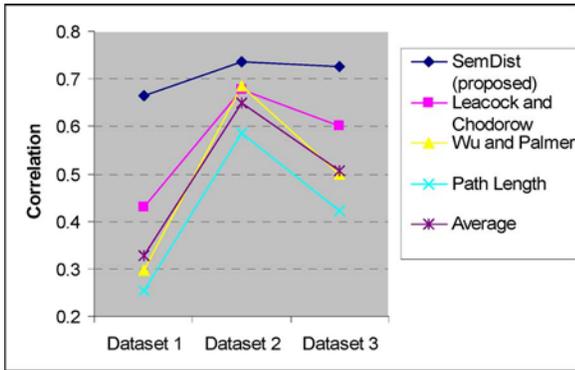
(a)
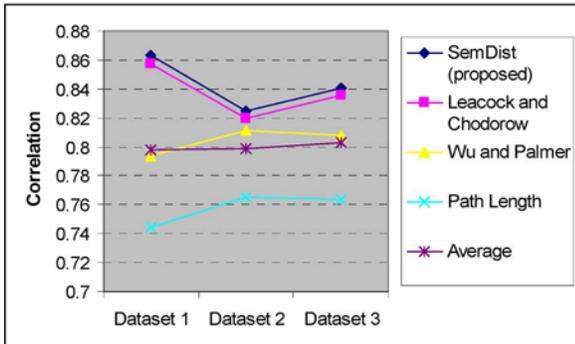
(b)

Fig. 4. Illustration of the correlations with human scores for four measures with dataset 1, dataset 2, and dataset 3 using (a) SNOMED-CT and (b) MeSH.

TABLE III
IMPROVEMENTS THAT SEMDIST ACHIEVED OVER THE AVERAGE OF THE THREE
OTHER SIMILAR TECHNIQUES USING SNOMED-CT WITH THREE DATASETS

	Correlations using SNOMED-CT		
	Dataset 1	Dataset 2	Dataset 3
Average of the 3 similar measures	0.327	0.650	0.507
SemDist	0.665	0.735	0.726
Improvement	103%	13%	43%

TABLE IV
IMPROVEMENTS THAT SEMDIST ACHIEVED OVER THE AVERAGE OF THE THREE
OTHER SIMILAR TECHNIQUES USING MeSH WITH THREE DATASETS

	Correlations using MeSH		
	Dataset 1	Dataset 2	Dataset 3
Average of the 3 similar measures	0.798	0.799	0.803
SemDist	0.863	0.825	0.841
Improvement	8.1%	3.3%	4.8%

TABLE V
ABSOLUTE CORRELATION OF THE PROPOSED APPROACH ON THE RG DATASET
AND WORDNET 2.0

Measure	Parameters	Correlation
SemDist (Proposed)	$\alpha=1, \beta=1, k=1$	0.815

TABLE VI
ABSOLUTE CORRELATIONS OF THE PROPOSED APPROACH IN
CROSS-ONTOLOGY (WORDNET AND MeSH)

No.	Dataset	Correlation
1	WordNet (RG, 65 pairs) + MeSH (Dataset 1, 25 pairs) (90 pairs)	0.808
2	WordNet (RG, 65 pairs) + MeSH (Dataset 2, 36 pairs) (101 pairs)	0.804
3	WordNet (RG, 65 pairs) + MeSH (Dataset 1, 25 pairs) + MeSH (Dataset 2, 36 pairs) (126 pairs)	0.814
Average number of tested pairs: Average correlation:	**105.7 0.809**	

These results are also illustrated in Fig. 4. Moreover, the improvements of the proposed measure, SemDist, achieved over the average of the three other measures, with both SNOMED-CT and MeSH, are shown in Tables III and IV, respectively. Leacock and Chodorow measure achieves the second best correlations (after SemDist) in five of the six experiments (Tables I and II), while Wu and Palmer measure gives the third best correlations in five of six experiments and the second best correlation in one experiment. Path length achieves the lowest correlations in all six experiments. These results seem realistic since Leacock and Chodorow measure uses both path length and depth features, and thus, outperforms both the path length measure and Wu and Palmer measure that uses only depth feature. To be more specific, the measure of Leacock and Chodorow computes the similarity by using the path length scaled by the maximum depth of the ontology, whereas Wu and Palmer measure uses depth of LCS of the two concepts scaled by the summation of the depths of the two concepts [7], [25]. We note that SemDist outperforms the other measures with SNOMED-CT more significantly than with MeSH because of the higher specificity of SNOMED-CT (with depth around 18) compared to MeSH (with a depth of around 12). From the results in Tables III and IV, we observe that SemDist achieved an average improvement of 53% using SNOMED-CT, while using MeSH, the average improvement is 5.3%. This suggests that SemDist is a good choice for ontologies with high specificities where the new CSpec feature will have

TABLE VII
BIOMEDICAL DATASET 2 (36 PAIRS) AND RG DATASET (65 PAIRS, IN ITALICS) WITH HUMAN SIMILARITY SCORES (HUMAN) AND SEMDIST'S SCORES USING WORDNET AND MESH

Concept 1	Concept 2	Score		Concept 1	Concept 2	Score	
		Human	SemDist			Human	SemDist
Anemia	Appendicitis	0.031	4.69	Diabetic Nephropathy	Diabetes Mellitus	0.500	3.25
Meningitis	Tricuspid Atresia	0.031	4.69	Pulmonary Valve Stenosis	Aortic Valve Stenosis	0.531	3.12
Sinusitis	Mental Retardation	0.031	4.69	Hepatitis B	Hepatitis C	0.562	2.97
Dementia	Atopic Dermatitis	0.062	4.83	Vaccines	Immunity	0.593	4.79
Acquired Immunodeficiency Syndrome	Congenital Heart Defects	0.062	4.54	Psychology	Cognitive Science	0.593	2.47
Bacterial Pneumonia	Malaria	0.156	4.69	Failure to Thrive	Malnutrition	0.625	4.69
Osteoporosis	Patent Ductus Arteriosus	0.156	4.83	Urinary Tract Infection	Pyelonephritis	0.656	3.92
Amino Acid Sequence	Anti Bacterial Agents	0.156	5.24	Migraine	Headache	0.718	4.72
Otitis Media	Infantile Colic	0.156	4.94	Myocardial Ischemia	Myocardial Infarction	0.750	2.33
Hyperlipidemia	Hyperkalemia	0.156	3.92	Carcinoma	Neoplasm	0.750	3.75
Neonatal Jaundice	Sepsis	0.156	4.69	Breast Feeding	Lactation	0.843	0.00
Asthma	Pneumonia	0.187	3.64	Seizures	Convulsions	0.843	0.00
Hypothyroidism	Hyperthyroidism	0.357	3.25	Pain	Ache	0.875	0.00
Sarcoidosis	Tuberculosis	0.406	5.05	Malnutrition	Nutritional Deficiency	0.875	0.00
Sickle Cell Anemia	Iron Deficiency Anemia	0.406	4.01	Down Syndrome	Trisomy 21	0.875	0.00
Adenovirus	Rotavirus	0.437	4.14	Measles	Rubeola	0.906	0.00
Lactose Intolerance	Irritable Bowel Syndrome	0.468	4.01	Antibiotics	Antibacterial Agents	0.937	0.00
Hypertension	Kidney Failure	0.500	4.83	Chicken Pox	Varicella	0.968	0.00
Cord	*smile*	*0.005*	*5.26*	*car*	*journey*	*0.388*	*5.40*
Rooster	*voyage*	*0.010*	*5.78*	*cemetery*	*mound*	*0.423*	*5.08*
Noon	*string*	*0.010*	*5.24*	*glass*	*jewel*	*0.445*	*4.51*
Fruit	*furnace*	*0.013*	*4.44*	*magician*	*oracle*	*0.455*	*4.44*
Autograph	*shore*	*0.015*	*5.32*	*crane*	*implement*	*0.593*	*3.97*
Automobile	*wizard*	*0.028*	*5.20*	*brother*	*lad*	*0.603*	*4.04*
Mound	*stove*	*0.035*	*4.44*	*sage*	*wizard*	*0.615*	*4.26*
Grin	*implement*	*0.045*	*5.40*	*oracle*	*sage*	*0.653*	*4.19*
Asylum	*fruit*	*0.048*	*4.44*	*bird*	*crane*	*0.658*	*3.33*
Asylum	*monk*	*0.098*	*5.02*	*bird*	*cock*	*0.658*	*2.30*
Graveyard	*madhouse*	*0.105*	*5.42*	*food*	*fruit*	*0.673*	*4.73*
Glass	*magician*	*0.110*	*4.66*	*brother*	*monk*	*0.685*	*2.48*
Boy	*rooster*	*0.110*	*4.97*	*asylum*	*madhouse*	*0.760*	*2.30*
Cushion	*jewel*	*0.113*	*4.44*	*furnace*	*stove*	*0.778*	*4.60*
Monk	*slave*	*0.143*	*4.04*	*magician*	*wizard*	*0.803*	*0.00*
Asylum	*cemetery*	*0.198*	*5.18*	*hill*	*mound*	*0.823*	*0.00*
Coast	*forest*	*0.213*	*4.51*	*cord*	*string*	*0.853*	*2.56*
Grin	*lad*	*0.220*	*5.40*	*glass*	*tumbler*	*0.863*	*2.48*
Shore	*woodland*	*0.225*	*4.33*	*grin*	*smile*	*0.865*	*0.00*
Monk	*oracle*	*0.228*	*4.60*	*serf*	*slave*	*0.865*	*3.69*
Boy	*sage*	*0.240*	*4.26*	*journey*	*voyage*	*0.895*	*2.48*
Automobile	*cushion*	*0.243*	*4.65*	*autograph*	*signature*	*0.898*	*2.48*
Mound	*shore*	*0.243*	*3.97*	*coast*	*shore*	*0.900*	*2.56*
Lad	*wizard*	*0.248*	*4.04*	*forest*	*woodland*	*0.913*	*0.00*
Forest	*graveyard*	*0.250*	*4.98*	*implement*	*tool*	*0.915*	*2.56*
Food	*rooster*	*0.273*	*5.34*	*cock*	*rooster*	*0.920*	*0.00*
Cemetery	*woodland*	*0.295*	*4.98*	*boy*	*lad*	*0.955*	*2.56*
Shore	*voyage*	*0.305*	*5.32*	*cushion*	*pillow*	*0.960*	*2.56*
Bird	*woodland*	*0.310*	*4.80*	*cemetery*	*graveyard*	*0.970*	*0.00*
Coast	*hill*	*0.315*	*3.97*	*automobile*	*car*	*0.980*	*0.00*
Furnace	*implement*	*0.343*	*4.26*	*midday*	*noon*	*0.985*	*0.00*
Crane	*rooster*	*0.353*	*4.16*	*gem*	*jewel*	*0.985*	*0.00*
Hill	*woodland*	*0.370*	*4.33*	--	--	--	--

more positive impact on the correlation results. Even with MeSH, where the average improvement is 5.4%, this improvement can be considered significant given the existing limited resources of human scored datasets in this domain. Furthermore, Tables I and II show that all four measures performed better in MeSH than in SNOMED-CT.

In another single ontology experiment, the RG dataset and WordNet 2.0 were used, and the results are shown in Table V using the default parameters ($\alpha = 1$, $\beta = 1$, $k = 1$). The purpose of this experiment (Table V) is to show that the proposed measure achieves sound results using a standard dataset (RG) and a large and reliable terminology (WordNet).

TABLE VIII
ABSOLUTE CORRELATIONS OF THE PROPOSED APPROACH IN
CROSS-ONTOLOGY (WORDNET AND SNOMED-CT)

No.	Dataset	Correlation
1	WordNet (RG, 65 pairs) + SNOMED-CT (Dataset 1, 29 pairs) (94 pairs)	0.778
2	WordNet (RG, 65 pairs) + SNOMED-CT (Dataset 2, 34 pairs) (99 pairs)	0.700
3	WordNet (RG, 65 pairs) + SNOMED-CT (Dataset 1, 29 pairs) + SNOMED-CT (Dataset 2, 34 pairs) (128 pairs)	0.757
Average number of tested pairs:	107	
Average correlation:	0.745	

To evaluate the approach in cross-ontology, the RG dataset (65 pairs) was combined with the two biomedical datasets (dataset 1 and dataset 2) in three combinations as follows.

1) RG (65 pairs) + Dataset 1 (30 pairs): Total 95 pairs.
2) RG (65 pairs) + Dataset 2 (36 pairs): Total 101 pairs.
3) RG + Dataset 1 + Dataset 2: Total 131 pairs.

WordNet was used for RG terms, and MeSH or SNOMED-CT was used for terms of dataset 1 and dataset 2. Moreover, WordNet was considered the primary while MeSH/SNOMED-CT was the secondary ontology. Then, on these three dataset combinations 1)–3), two evaluations (I and II given later) were conducted: one using WordNet–MeSH and the other using WordNet–SNOMED-CT as primary–secondary ontologies.

Evaluation I: Three experiments [using combinations 1)–3)] were conducted in this evaluation. In the first experiment, using combination 1), only 25 pairs (out of the 30 pairs in dataset 1) were found in MeSH. Thus, the similarity of 65 pairs in WordNet obtained as in single ontology model, and 25 term pairs as cross-ontology technique (case 3). In the second experiment, dataset combination 2) was tested using WordNet and MeSH. In the third experiment, combination 3), the three datasets were combined with a total of 126 pairs distributed between WordNet and MeSH. The results are shown in Table VI. In these experiments, the proposed method achieved an average of ~81% correlation with human scores using, on average, ~106 term pairs and two ontologies. The complete results of correlation for each pair using combination 2) (the second experiment in Table VI) are shown in Table VII. The human rating scores in RG dataset are converted into [0–1] scale to be compatible with human ratings in dataset 2.

Evaluation II: In these experiments, 29 out of 30 pairs in dataset 1 and 34 out of 36 pairs in dataset 2 were found in SNOMED-CT. We calculate 65 pairs in WordNet as in single terminology, and the biomedical term pairs in dataset 1 and dataset 2 using the cross-ontology technique. The results are shown in Table VIII. From these results, we note that using MeSH for biomedical terms (dataset 1 and dataset 2) produces better correlation with human judgments than with SNOMEDCT; this also holds true for single ontology results in Tables I and II. The results in Tables VI and VIII prove that the cross-ontology approach is very promising and efficient with very good correlations with human ratings in combined datasets.

VI. DISCUSSION AND CONCLUSION

Measuring semantic similarity between terms and concepts has many great uses and applications in the biomedical domain and yet faced by many difficulties. One of the problems in measuring semantic similarity between concepts, using ontology, is that certain concepts maybe missing from the underlying ontology. This problem stands out prominently in certain domains (e.g., bioinformatics). For example, in biomedical SIR, there is a great need for measuring the semantic similarity between biomedical concepts and document terms [27], and there are several potential ontologies. It can very well be that not all the concepts are found in single terminology (i.e., the concepts are dispersed on more than one terminology) [16]. We presented a new way to measure the semantic similarity between terms using multiple ontologies. The new measure nonlinearly combines ontology-based features with new features that address the similarity across multiple ontologies. We further presented a strategy for combining and selecting ontologies in measuring the similarity between terms dispersed in multiple ontologies. The proposed measures were evaluated using biomedical datasets and multiple different ontologies within the UMLS, and in single-ontology as well as in cross-ontologies. In the experimental results, the proposed measures achieved the best correlations with human scores in all experiments, with three datasets and with different ontologies. In some experiments, our measure achieved more than 100% improvement over the average of the other similar techniques (Table III), and using MeSH, our proposed measure gave high correlations (slightly exceeding 0.86) in one of the experiments (Table II). And overall, the proposed measure achieved improvements between 0.5% (Table II) and 54% (Table I) over the measure of Leacock and Chodorow, which was the best performer among the three evaluated ontology-based measures (Tables I–IV). To the best of our knowledge, these are the best reported results on this problem in the biomedical domain to date. The experimental results further demonstrated that MeSH terminology produces better semantic similarity correlations with human ratings than SNOMED-CT in all of the tested measures. The reason for this is the fact that MeSH is closer to human perspective of relations between the biomedical terms than SNOMED-CT. This is also the case not only with the proposed measure, but also with the three other measures that we tested (Leacock–Chodorow, Wu–Palmer, and Path length, as in Tables I and II). We exerted research effort to examine this point and found that the structure of these medical terms in MeSH terminology is closer to human perspective than the arrangement of these terms in SNOMED-CT. The experimental results with cross-ontology measure were also quite impressive. On average, the measure achieved a correlation of about 0.81 with human scores of three cross-ontology experiments (Table VI). These results are fairly significant giving the difficulty of the task of measuring the similarity of concepts within several ontologies that overlap in many nodes. This is an interesting work that provides a good contribution in this task of biomedical term similarity across multiple ontologies in a unified framework, and will provoke more progress in this direction in ontology integration and cross-ontology research in

the biomedical domain. In the future work of this research, we would like to investigate a few more aspects related to semantic similarity; for example, we like to comprehensively examine these semantic similarity measures compared with what UMLS provides as clues of semantic similarity. We would also like to involve more terminologies and vocabularies from the biomedical domain. For example, we can expand this research to involve terms from gene and protein structured databases, vocabularies, and ontologies (e.g., gene ontology, Swiss-Prot, TrEMBL, and UniProt). Furthermore, we would like to study the semantic similarity with different types of relations than is-a and part-of, and with different weights of edges between terms.

ACKNOWLEDGMENT

REFERENCES

[1] A. Budanitsky and G. Hirst, "Evaluating WordNet-based measures of semantic distance," *Comput. Linguistics*, vol. 32, no. 1, pp. 13–47, 2006.

[2] J. E. Caviedes and J. J. Cimino, "Towards the development of a conceptual distance metric for the UMLS," *J. Biomed. Inf.*, vol. 37, no. 2, pp. 77–85, 2004.

[3] A. Hliaoutakis, "Semantic similarity measures in MeSH ontology and their application to information retrieval on Medline," Master's thesis, Tech. Univ. Crete, Chaniá, Crete, 2005.

[4] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," in *Proc. ROCLING X*, Taiwan, 1997.

[5] R. Kleinsorge, C. Tilley, and J. Willis. (2000). Unified Medical Language System (UMLS) Basics [Online]. Available: http://www.nlm.nih.gov/research/umls/pdf/UMLS_Basics.pdf

[6] H. Kuntz M. V. Berkum. SNOMED CT A Standard Terminology for Healthcare. (2005). [Online]. Available: http://www.sst.dk/upload/informatik_og_sundhedsdata/sundhedsinformatik/terminologi/kuntz_vanberkum_snomedct_30mar05.pdf

[7] C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," in *WordNet: An Electronic Lexical Database*, C. Fellbaum, Ed. Cambridge, MA: MIT Press, 1998, pp. 265–283.

[8] Y. Li, Z. A. Bandar, and D. McLean, "An approach for measuring semantic similarity between words using multiple information sources," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 4, pp. 871–882, Jul./Aug. 2003.

[9] D. Lin, "An information-theoretic definition of similarity," in *Proc. Int. Conf. Mach. Learning*, Jul. 1998, pp. 296–304.

[10] Medline. (2007). [Online]. Available: http://medline.cos.com/

[11] Medical Subject Heading MeSH. (2007). [Online]. Available: http://www.nlm.nih.gov/mesh

[12] H. Al-Mubaid and H. A. Nguyen, "Similarity computation using multiple UMLS ontologies in a unified framework," in *Proc. 22nd ACM Symp. Appl. Comput. (SAC)*, South Korea, 2007.

[13] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, pp. 39–41, 1995.

[14] H. A. Nguyen and H. Al-Mubaid, "New ontology-based semantic similarity measure for the biomedical domain," in *Proc. IEEE GrC 2006*, pp. 623–628.

[15] K. Ishida and T. Ohta, "An approach for organizing knowledge according to terminology and representing it visually," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 4, pp. 366–373, Nov. 2002.

[16] T. Pedersen, S. Pakhomov, and S. Patwardhan, "Measures of semantic similarity and relatedness in the medical domain," *J. Biomed. Inf.*, vol. 40, no. 3, 2007.

[17] R. Rada, H. Mili, E. Bichnell, and M. Blettner, "Development and application of a metric on semantic nets," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 17–30, Jan./Feb. 1989.

[18] P. Resnik, "Using information content to evaluate semantic similarity," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, Montreal, QC, Canada, 1995, pp. 448–453.

[19] K. M. Sim and P. T. Wong, "Toward agency and ontology for web-based information retrieval," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 257–269, Aug. 2004.

[20] M. A. Rodriguez and M. J. Egenhofer, "Determining semantic similarity among entity classes from different ontologies," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 2, pp. 442–456, Mar./Apr. 2003.

[21] I. Spasic, S. Ananiadou, J. McNaught, and A. Kumar, "Text mining and ontologies in biomedicine: Making sense of raw text," *Briefing Bioinf.*, vol. 6, no. 3, pp. 239–251, 2005.

[22] T. Pedersen, S. Patwardhan, and J. Michelizzi, "WordNet::Similarity-measuring the relatedness of concepts," presented at the 19th Nat. Conf. Artif. Intell. (AAAI-04), San Jose, CA, 2004.

[23] UMLS: Unified Medical Language System. (2007). [Online]. Available: http://www.nlm.nih.gov/research/umls/

[24] UMLS Browser-UMLS Knowledge Source Server. (2007). [Online]. Available: http: //umlsks.nlm.nih.gov

[25] Z. Wu and M. Palmer, "Verb semantics and lexical selection," in *Proc. 32nd Annu. Meeting Assoc. Comput. Linguistics*, 1994, pp. 133–138.

[26] XML MeSH. (2007). [Online]. Available: http://www.nlm.nih.gov/mesh/xmlmesh.html

[27] W. Mao and W. W. Chu, "Free-text medical document retrieval via phrase-based vector space model," in *Proc. AMIA Symp.*, 2002, pp. 489–493.

[28] S. Patwardhan, "Incorporating dictionary and corpus information into a context vector measure of semantic relatedness," Master's thesis, Univ. Minnesota, Minneapolis, 2003.

[29] R. N. Shepard, "Towards a universal law of generalisation for psychological science," *Science*, vol. 237, pp. 1317–1323, 1987.

[30] H. Rubenstein and J. B. Goodenough, "Contextual correlates of synonymy," *Commun. ACM*, vol. 8, pp. 627–633, 1965.

Hisham Al-Mubaid (M'02) received the M.S. and Ph.D. degrees in computer science from the University of Texas at Dallas, Dallas, in 1997 and 2000, respectively.

He was an Assistant Professor at the State University of New York (SUNY), Geneso, for one year. In 2001 he joined the University of Houston-Clear Lake (UHCL), Houston, TX, where he is currently an Associate Professor of computer science and the Chair of computer information systems. His current research interests include natural language processing, text mining, machine learning, bioinformatics, semantics, and ontology. He was a reviewer for numerous journals.

Prof. Al-Mubaid was the Program/Technical Chair for numerous conferences.

Hoa A. Nguyen (M'05) received the B.Sc. degree in computer science from Ho Chi Minh National University, Ho Chi Minh, Vietnam, in 2003, and the M.S. degree in computer science from the University of Houston-Clear Lake, Houston, TX, in 2006.

He is currently with the University of Houston-Clear Lake, where he was with the Department of Computer Science as a Research Assistant from 2005 to 2006. His current research interests include natural language processing, text mining, text categorization, feature selection, machine learning, bioinformatics, biomedical ontology and terminologies, and data mining.