Indonesian Journal of Electrical Engineering and Computer Science
Vol. 14, No. 3, June 2019, pp. 1356∼1372
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v14i3.pp1356-1372

Toward Semantic Similarity Measure Between Concepts in an Ontology

Suwan Tongphu, Boontawee Suntisrivaraporn, Pakinee Aimmanee
School of Information, Computer, and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University, Thailand

Article Info

Article history:

Received Oct 1, 2018
Revised Dec 10, 2018
Accepted Jan 15, 2019

Keywords:

Description logic
Semantic Analysis
Concept Similarity
Reasoning

ABSTRACT

A concept similarity measure is one classical problem in Description Logic which aims at identifying similarity between concepts in an ontology. Measuring a distance between concepts is an essential process. Most methods used for measuring, they usually do not take semantic for consideration. This work introduces a new method for concept similarity measure. The proposed method semantically analyzes structures of two concepts and then computes the similarity score based on the number of shared structures. The efficiency of the proposed algorithm is measured by means of the satisfaction of desirable properties and intensive experiments on the SNOMED CT ontology.

Corresponding Author:

Suwan Tongphu,
School of Information, Computer, and Communication Technology,
Sirindhorn International Institute of Technology,
Thammasat University, Thailand.
Email: stongphu@gmail.com

1. INTRODUCTION

With a rapid increasing of internet users and a massive online data, retrieving a relevant information from a given query is one of the most challenging topics. Semantic querying [1] is one recent aspect of information retrieval [2], which aims at representing knowledge in a well-found way and incorporating intelligence into the system. With the help of Description Logics (DLs) [3, 4], the use of Web Ontology language (OWL) [5, 6] to model the knowledge has been introduced and is lately recommended as a new standard for knowledge representation by W3C. A family of Description Logics (DLs) is a common tool to formally equip the knowledge base and offers several decidable reasoning services which are sufficient for several scenarios. For example, determining whether or not a concept is a subclass of another one can be done using a concept subsumption. Besides a usefulness of classical reasoning services [7], there are some cases in which the classical reasoners are inapplicable. An example includes a measuring similarity score between concepts. By using a classical DL reasoner, it is evidently insufficient since subsumption reasoning service simply returns a boolean value so they cannot provide a degree of similarity between concepts.

Several methods have been proposed for measuring similarity between concepts. The most well-known techniques are the distance-based [8] and the pattern-based analysis [9, 10]. These methods basically can be used for only learning a new pattern of concept. However, due to the fact that they have a lack of semantic analysis. They can only provide rough concept similarity outputs. To address this problem, modern semantic-based techniques, which aim at quantitatively analyzing values of concepts by means of their definitions, are lately introduced. The techniques are normally equipped to work with a different family of DLs. Distel et al. [11] proposed a new method for concept dissimilarity measure. The difference between two concepts C and D is measured by means of the number of operations required for relaxing a concept D until subsumed by a concept C. If the two concepts are concluded to be totally similar, the method returns 0 as an output. In addition to the method they proposed, the dissimilarity score is computed based on the number of relaxing operations.

Jaccard [12] proposed a simple method for computing similarity between concepts. However, the proposition merely supports the concept conjunction, which is mostly not practical in many real life ontologies. For example, it has been proved that building a large-scaled ontology requires at least a family of DL $\mathcal{ELH}$ (see e.g. SNOMED CT [13] and gene ontology). Recently, a similarity measures for a less-expressive DL $\mathcal{FL}_0$ was proposed by Racharak and Suntisrivaraporn [14]. Lehmann and Turhan [15] extended the work of Jaccard to support more constructors. They proposed a new similarity framework for DL $\mathcal{ELH}$. The operators of the proposed formulas are described by means of desired properties and left for interested users to customize.

In the work proposed by Janowic [16], a more refined semantic measure was proposed to employ high expressive DLs, e.g. $\mathcal{ALN}$. The extension to support DL $\mathcal{SHI}$ is subsequently proposed in the later work [17]. d'Amato et al. [18] introduced a new method for $\mathcal{ALE}$ concept similarity measure. The method satisfies several desirable properties including symmetric, equivalent invariant, structural dependent, and reverse subsumption preserving property. The adoption for DL $\mathcal{ALC}$, which equally satisfies the same properties, is proposed in their later work [19].

In this work, we introduce a new algorithm for computing similarity between concepts based on shared features. Unlike any other approaches which are tailored for a specific domain, this work proposes a new notion for a concept similarity measure for a general domain. The proposed method is designed to work with the knowledge base modeled using at most the lightweight DL $\mathcal{ALEH}$ family. Comparing to more expressive DLs, modeling the knowledge base using the family of $\mathcal{ALEH}$ is more practical since a computing time is polynomially bounded. Moreover, it is more convenient to meet a large-scale expansion. Examples include the modeling of knowledge bases using the DL $\mathcal{EL}$, e.g. the well-known knowledge bases for clinical terms (SNOMED CT), lexical terms (WordNet), and genes (Gene ontology).

To enable semantic measure, we first transform the concept descriptions to their equivalent description trees. The level of similarity from one concept to another is then measured based on how well the two description trees can be mapped. The overall similarity rate is lastly reported as an average of similarity. The effectiveness of the proposed method is measured by means of satisfactory of desirable properties and compared to state-of-the-art methods.

In the next section, we briefly introduce the notion of DLs, describe the expansion process for a concept description, describe the rules which we use to normalize expanded concept description, and also provide steps which we use to construct a so-called concept description tree. Later sections introduce notions of a homomorphism score which measures a similarity from one concept description tree to another. The notion of $\mathcal{ALEH}$ semantic similarity measure is introduced. The example of computation is exemplified by means of a prototypical family ontology. More intensive experiments are performed on the well-known SNOMED CT ontology and reported in the experiment section. The last section gives a conclusion of this work.

2. BACKGROUND

In DL $\mathcal{ALEH}$, concepts are used to describe classes of objects and roles are used to describe their relations. In this work, we use CN to represent a set of concept names and RN to represent a set of role names. Complex concept descriptions can be formulated based on CN, RN, and concept constructors such as a concept conjunction $\sqcap$ (the upper section of Table 1 show all constructors for DL $\mathcal{ALEH}$). Conventionally, we use the symbols r and s to represent role names ($r, s \in \mathsf{RN}$), A and B to represent concept names ($\mathsf{A}, \mathsf{B} \in \mathsf{CN}$), and C and D to represent complex concept descriptions. For example, let $\mathsf{Female}, \mathsf{Male}, \mathsf{Person} \in \mathsf{CN}$ and $\mathsf{child} \in \mathsf{RN}$, we can define a concept of Woman by means of the following concept description:

$$\mathsf{Female} \sqcap \mathsf{Person}.$$

Likewise, we can define a concept of Mother based on the existing concept Woman as follow:

$$\mathsf{Woman} \sqcap \exists \mathsf{child.Person}.$$

Formally, we define the semantics of DL $\mathcal{ALEH}$ by means of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, which is a pair of an interpretation domain $\Delta^{\mathcal{I}}$ (i.e. a finite set of individuals of the domain of interest), and an interpretation function $\cdot^{\mathcal{I}}$ (i.e. a function that maps $A \in \mathsf{CN}$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and $r \in \mathsf{RN}$ to a binary relation $r^{\mathcal{I}}$ on $\Delta^{\mathcal{I}}$). There are two facilities to define a new concept: a) concept equivalence ($\equiv$) and b) concept inclusion ($\sqsubseteq$). See the syntax in the lower part of Table 1. For example, we can define the concept Mother using the concept equivalence as shown below:

$$\mathsf{Mother} \equiv \mathsf{Woman} \sqcap \exists \mathsf{child.Person}.$$

This infers that a mother is a woman who has some child person and vice versa. However, if a concept is defined using the concept inclusion, it will be interpreted merely in a forward direction. For example, if we define a concept Father as follows:

$$\mathsf{Father} \sqsubseteq \mathsf{Man} \sqcap \exists \mathsf{child.Person}$$

this infers that a father is a man who has some child person. However, it is still unknown whether a man who has some child person will be a father. Nevertheless, for each concept inclusion B in which $B \sqsubseteq D$, it can be equally transformed to a concept equivalence $B \equiv F \sqcap D$ where F is a fresh concept name (F is unknown). Therefore, the concept Father can be transformed to the following form:

$$\mathsf{Father} \equiv \mathsf{F} \sqcap \mathsf{Man} \sqcap \exists \mathsf{child.Person}.$$

In addition, assume that each defined concept has only one definition and does not contain any cyclic dependencies, by recursively replacing defined concepts with their definitions, we have a new equivalent concept definition which contains only primitive concept names (concept names that appear only on the right-hand side of concept definitions). Symbolically, we denote by $\mathsf{CN}^{\mathsf{pri}}$ a set of primitive concepts.

We call a set of concept definitions a knowledge base or a terminology (*TBox*). For instance, we can define the TBox for a family domain as a set of concepts shown in Figure 3. A TBox is *unfoldable* if all concept definitions are expandable. Given, for example, the definition of MotherNoSon:

$$\mathsf{MotherNoSon} \equiv \mathsf{Mother} \sqcap \forall \mathsf{child.Woman}$$

By replacing Mother with Woman $\sqcap$ $\exists$child.Person and Woman with Female $\sqcap$ Person, we then have an equivalent definition of MotherNoSon as follows:

$$\mathsf{MotherNoSon} \equiv \mathsf{Female} \sqcap \mathsf{Person} \sqcap \forall \mathsf{child.(Female} \sqcap \mathsf{Person)} \sqcap \exists \mathsf{child.Person}$$

where Person, Female $\in \mathsf{CN}^{\mathsf{pri}}$. In symbol, for every $\mathcal{ALEH}$ concept which defined in an unfoldable TBox, we assume without lost of generality in the following form:

$$\prod_{i=1}^{l} P_i \quad \sqcap \quad \prod_{j=1}^{m} \exists r_j.C_j \quad \sqcap \quad \prod_{k=1}^{n} \forall s_k.D_k \tag{1}$$

where $P_i \in \mathsf{CN}^{\mathsf{pri}}$, $r_j, s_k \in \mathsf{RN}$, and $C_j, D_k \in \mathsf{CN} \cup \{\top, \bot\}$. For simplicity, we assign $\mathcal{P}_C := \{P_1, \ldots, P_l\}$, $\mathcal{E}_C := \{\exists r_1.C_1, \ldots, \exists r_m.C_m\}$, and $\mathcal{A}_C := \{\forall s_1.D_1, \ldots, \forall s_n.D_n\}$ where l is the size of $\mathcal{P}_C$, m is the size of $\mathcal{E}_C$, and n is the size of $\mathcal{A}_C$. Additionally, given that $\sqsubseteq^*$ be the transitive closure of $\sqsubseteq$ over the role names, we use the symbols $\mathcal{R}_{\exists r} = \{s \in \mathsf{RN} \mid r \sqsubseteq^* s\}$ to represent a set of super-roles of r and $\mathcal{R}_{\forall r} = \{t \in \mathsf{RN} \mid t \sqsubseteq^* r\}$ to represent a set of sub-roles of r.

Table 1. Syntax and semantics of the DL $\mathcal{ALEH}$

Name	Syntax	Semantics
bottom	$\bot$	$\emptyset$
top	$\top$	$\Delta^{\mathcal{I}}$
concept name	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
atomic negation	$\neg A$	$\Delta^{\mathcal{I}} \backslash A$
concept conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
universal restriction	$\forall r.C$	$\{x \mid \forall y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}\}$
concept inclusion	$B \sqsubseteq D$	$A^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
concept equivalent	$B \equiv D$	$A^{\mathcal{I}} = D^{\mathcal{I}}$
role hierarchy	$r \sqsubseteq s$	$r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$

In addition to the expanded form of the $\mathcal{ALEH}$ concept description, there may exists a case which makes the description implicit. This can be eliminated by applying the following rules over the expanded description:

$$
\begin{aligned}
\forall r.C \sqcap \exists r.D &\rightarrow \forall r.C \sqcap \exists r.(C \sqcap D),\\
\forall r.C \sqcap \forall r.D &\rightarrow \forall r.(C \sqcap D),\\
\forall r.\top &\rightarrow \top,\\
\exists r.\bot &\rightarrow \bot,\\
C \sqcap \bot &\rightarrow \bot.\\
A \sqcap \neg A &\rightarrow \bot,\\
C \sqcap \top &\rightarrow C,
\end{aligned}
$$

To be more illustrative, by applying the rules above to the expanded form of MotherNoSon, we have the following normalized definition:

$$\mathsf{Female} \sqcap \mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Female} \sqcap \mathsf{Person}) \sqcap \forall \mathsf{child}.(\mathsf{Female} \sqcap \mathsf{Person})$$

3. RESEARCH METHOD

In the work proposed by Baader and Kusters [20], a characterization using homomorphism for an unfoldable $\mathcal{ALEH}$ TBox has been proposed. The authors proved that if the concept C is subsumed by D, then there must exist a homomorphism from a concept description tree of D to that of C. Our proposed concept similarity measure is directly derived from a concept homomorphism, which is one important characterization of a concept subsumption. The measure is, however, extended for the case where the two concepts are out of a subsumption relation but there still exist some shared structures.

Definition 1. (**$\mathcal{ALEH}$ concept subsumption**) *Let C and D are $\mathcal{ALEH}$ concept descriptions which defined in the terminology $\mathcal{O}$, we say that $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. Moreover, $C \equiv D$ if $C \sqsubseteq D$ and $D \sqsubseteq C$.*

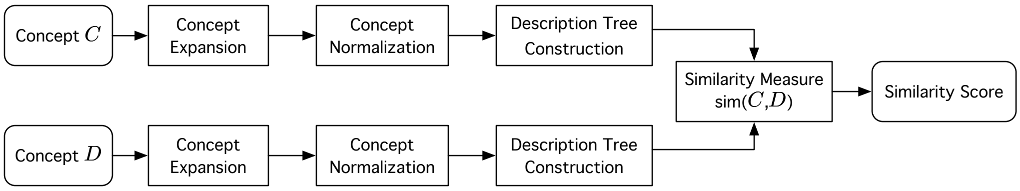

Figure 1. An overview of the similarity measure system

Figure 1 depicts the overview of our similarity measure system. Starting with two input concept descriptions, we expand and transform them into the normal forms. A so-called $\mathcal{ALEH}$ description tree is then constructed. For example, given C an expanded and normalized concept description, we construct a concept description tree $\mathcal{G}_C := (V, E, v_0, \ell, \rho)$ where V is a set of nodes, $E \subseteq V \times V$ is a set of edges, v_0 is the root, $\ell : V \to 2^{\mathsf{CN}^{\mathsf{pri}}}$ is a function representing a set of node labels, and $\rho : E \to 2^{\mathsf{RN}}$ is a function representing a set of edge labels. The following shows the steps for constructing an $\mathcal{ALEH}$ description tree:

i. Create a new node v_0 and assign $\mathcal{P}_C$ to $\ell(v_0)$.

ii. For each $\exists r.D_j \in \mathcal{E}_C$, create a new node w and then introduce a new edge (v_0, w) with w an r-successor of v_0 and assign $\mathcal{R}_{\exists r}$ to $\rho(v_0, w)$. Repeat from step (i) by treating D_j as C and w as v_0.

iii. For each $\forall s.D_k \in \mathcal{A}_C$, create a new node w' and then introduce a new edge (v_0, w') with w' an s-successor of v_0 and assign $\mathcal{R}_{\forall s}$ to $\rho(v_0, w')$. Repeat from step (i) by treating D_k as C and w' as v_0.

Theorem 1 shows that the concept subsumption can be characterized by means of a homomorphism mapping from an opposite direction.

Theorem 1 (Let C and D be $\mathcal{ALEH}$ concept descriptions, and $\mathcal{G}_C$ and $\mathcal{G}_D$ be the corresponding $\mathcal{ALEH}$ concept description trees. We say that $C \sqsubseteq D$ if there is a homomorphism $h : \mathcal{G}_D \to \mathcal{G}_C$ which maps all nodes and edges of $\mathcal{G}_D$ to the corresponding nodes and edges of $\mathcal{G}_C$ [21]**)**. .

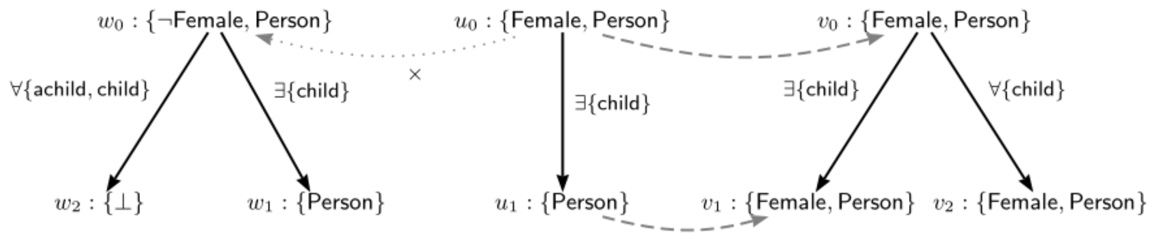

Figure 2. A homomorphism (dashed arrows) mapping $\mathcal{G}_{\mathsf{Mother}}$ to $\mathcal{G}_{\mathsf{MotherNoSon}}$ and a failure of mapping (dotted arrows) $\mathcal{G}_{\mathsf{Mother}}$ to $\mathcal{G}_{\mathsf{NonAdoptiveFather}}$.

To be more visible, consider the normalized description of the concept MotherNoSon and the following normalized description of the concept Mother and NonAdoptiveFather:

$$\begin{aligned} \mathsf{Mother} &\equiv \mathsf{Female} \sqcap \mathsf{Person} \sqcap \exists \mathsf{child}.\mathsf{Person}, \\ \mathsf{NonAdoptiveFather} &\equiv \neg \mathsf{Female} \sqcap \mathsf{Person} \sqcap \exists \mathsf{child}.\mathsf{Person} \sqcap \forall \mathsf{achild}.\bot. \end{aligned} \quad (2)$$

We can construct the $\mathcal{ALEH}$ description trees $\mathcal{G}_{\mathsf{MotherNoSon}}$, $\mathcal{G}_{\mathsf{Mother}}$, and $\mathcal{G}_{\mathsf{NonAdoptiveFather}}$ using the process previously described. Figure 2 shows a successful attempt of the homomorphism mapping from $\mathcal{G}_{\mathsf{Mother}}$

to $\mathcal{G}_{\mathsf{MotherNoSon}}$. It is obvious that all nodes and edges $\mathcal{G}_{\mathsf{Mother}}$ can be mapped to $\mathcal{G}_{\mathsf{MotherNoSon}}$. The figure also shows a failed attempt of a homomorphism mapping from $\mathcal{G}_{\mathsf{Mother}}$ to $\mathcal{G}_{\mathsf{NonAdoptiveFather}}$. By Theorem 1, we can conclude that $\mathsf{MotherNoSon} \sqsubseteq \mathsf{Mother}$ but $\mathsf{NonAdoptiveFather} \not\sqsubseteq \mathsf{Mother}$.

By employing a classical subsumption reasoning service, it is obvious that MotherNoSon is Mother and NonAdoptiveFather is not Mother. However, by analyzing the structure of $\mathcal{G}_{\mathsf{NonAdoptiveFather}}$ and $\mathcal{G}_{\mathsf{Mother}}$, there are some shared structures (e.g. both are person and have some child). Thus, there must exist some similarity between these two concepts though out of subsumption relation. Our interest is to measure their degree of similarity.

3.1. Homomorphism score

From Theorem 1, it is obvious that a subsumption relation can be characterized by means of a homomorphism mapping in a reverse direction. In this section, we consider a case where the homomorphism condition is not fully satisfied but there is some shared structure between two description trees.

Symbolically, let C and D be $\mathcal{ALEH}$ concept descriptions, $\mathcal{P}_C$ and $\mathcal{P}_D$ be sets of primitive concepts, $\mathcal{E}_C$ and $\mathcal{E}_D$ be sets of existential restrictions, $\mathcal{A}_C$ and $\mathcal{A}_D$ be as sets of universal restrictions, and $\mathcal{G}_C$ and $\mathcal{G}_D$ be $\mathcal{ALEH}$ concept description trees. We measure the similarity from C to D by means of the homomorphism score $\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C)$. The *homomorphism score function* $\mathsf{hd} : \mathbf{G}^{\mathcal{ALEH}} \times \mathbf{G}^{\mathcal{ALEH}} \to [0, 1]$ is mathematically defined as follows:

$$\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C) := \quad (1 - \mu^{\mathsf{e}} - \mu^{\mathsf{a}}) \cdot \mathsf{p\_hd}(\mathcal{P}_D, \mathcal{P}_C) + \ \mu^{\mathsf{e}} \cdot \mathsf{e\_set\_hd}(\mathcal{E}_D, \mathcal{E}_C) + \ \mu^{\mathsf{a}} \cdot \mathsf{a\_set\_hd}(\mathcal{A}_D, \mathcal{A}_C)$$

Where each component constituting this function is defined in the following manners. The parameter $\mu^{\mathsf{e}} = \frac{|\mathcal{E}_D|}{|\mathcal{P}_D \cup \mathcal{E}_D \cup \mathcal{A}_D|}$ and $\mu^{\mathsf{a}} = \frac{|\mathcal{A}_D|}{|\mathcal{P}_D \cup \mathcal{E}_D \cup \mathcal{A}_D|}$ assign the weights indicating how important the existentially and universally quantified subconcepts are to be considered. Intuitively, if the number of top-level primitive concepts $\mathcal{P}_D$ is greater than the number top-level existential restrictions $\mathcal{E}_D$ and the number of top-level existential restriction $\mathcal{A}_D$, we consider that the similarity between nodes is more important than the similarity between edges, which results in an increasing of μ. Otherwise, the similarity between edges is more important than that of between nodes, which results a decreasing of μ. Additionally, the homomorphism score hd is a measure from $\mathcal{G}_D$ to $\mathcal{G}_C$. It is defined as a weighted summation of the similarity between nodes (p_hd), existential restrictions (e_set_hd), and universal restrictions (a_set_hd). The function p_hd determines the similarity score between nodes and is defined as follows:

$$\mathsf{p\_hd}(\mathcal{P}_D, \mathcal{P}_C) := \begin{cases} 1 & \text{if } \mathcal{P}_D = \emptyset \text{ or } \mathcal{P}_C = \{\bot\} \\ \frac{|\mathcal{P}_D \cap \mathcal{P}_C|}{|\mathcal{P}_D|} & \text{otherwise,} \end{cases} \tag{3}$$

where $|\cdot|$ represents the set cardinality. To identify the similarity among edges, we consider the similarity from $\mathcal{E}_D$ to $\mathcal{E}_C$, and also from $\mathcal{A}_D$ to $\mathcal{A}_C$ using the function $\mathsf{e\_set\_hd}(\mathcal{E}_D, \mathcal{E}_C)$ and $\mathsf{a\_set\_hd}(\mathcal{A}_D, \mathcal{A}_C)$, respectively. The function $\mathsf{e\_set\_hd}(\mathcal{E}_D, \mathcal{E}_C)$ is defined as follow:

$$\mathsf{e\_set\_hd}(\mathcal{E}_D, \mathcal{E}_C) := \begin{cases} 1 & \text{if } \mathcal{E}_D = \emptyset \\ 0 & \text{if } \mathcal{E}_D \neq \emptyset, \mathcal{E}_C = \emptyset \\ \sum\limits_{\epsilon_i \in \mathcal{E}_D} \frac{max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_C\}}{|\mathcal{E}_D|} & \text{otherwise,} \end{cases} \tag{4}$$

where ϵ_i, ϵ_j are existential restrictions; Note that all $\exists r.\bot$ will be transformed to $\bot$ during the normalization process. Therefore, we need not to treat this case in Equation 4. For each existential restriction ϵ_i, we compute the similarity to each ϵ_j using the function e_hd.

$$\mathsf{e\_hd}(\exists r.X, \exists s.Y) := \gamma^{\mathsf{e}}(\nu^{\mathsf{e}}(r) \ + (1 - \nu^{\mathsf{e}}(r)) \cdot \mathsf{hd}(\mathcal{G}_X, \mathcal{G}_Y)) \tag{5}$$

where $\nu^{\mathsf{e}} : \mathsf{RN} \to [0, 1)$ is a role weight function. It assigns different weight to each role name. Moreover, we use $\gamma^{\mathsf{e}} = \frac{|\mathcal{R}_{\exists r} \cap \mathcal{R}_{\exists s}|}{|\mathcal{R}_{\exists r}|}$ to indicate an inclusion score between labels of two edges. For the case $\gamma^{\mathsf{e}} = 0$, this infers that the two edges have no feature in common. Therefore, the homomorphism score of a successor can be omitted. The function e_hd returns the similarity score if there is an existential edge label in common; However, the successors' structures must be recursively checked using the function $\mathsf{hd}(\mathcal{G}_X, \mathcal{G}_Y)$. Similar to the existential restrictions, we apply the same operations for the universal restrictions as shown below:

$$\mathsf{a\_set\_hd}(\mathcal{A}_D, \mathcal{A}_C) := \begin{cases} 1 & \text{if } \mathcal{A}_D = \emptyset, \\ 0 & \text{if } \mathcal{A}_D \neq \emptyset, \mathcal{A}_C = \emptyset, \\ \sum\limits_{\alpha_i \in \mathcal{A}_D} \frac{max\{\mathsf{a\_hd}(\alpha_i, \alpha_j) : \alpha_j \in \mathcal{A}_C\}}{|\mathcal{A}_D|} & \text{otherwise} \end{cases} \quad (6)$$

where α_i, α_j are universal restrictions; and

$$\mathsf{a\_hd}(\forall r.X, \forall s.Y) := \begin{cases} \gamma^{\mathsf{a}} & \text{if } \mathcal{P}_Y = \{\bot\}, \\ \gamma^{\mathsf{a}}(\nu^{\mathsf{a}}(r) + (1 - \nu^{\mathsf{a}}(r)) \cdot \mathsf{hd}(\mathcal{G}_X, \mathcal{G}_Y)) & \text{otherwise} \end{cases} \quad (7)$$

where $\gamma^{\mathsf{a}} = \frac{|\mathcal{R}_{\forall r} \cap \mathcal{R}_{\forall s}|}{|\mathcal{R}_{\forall r}|}$ and $\nu^{\mathsf{a}} : \mathsf{RN} \to [0, 1)$.

ω_1	Woman ≡ Female ⊓ Person
ω_2	Man ≡ ¬Female ⊓ Person
ω_3	Parent ≡ Person ⊓ ∃child.Person
ω_4	Mother ≡ Woman ⊓ Parent
ω_4	Father ≡ Man ⊓ Parent
ω_5	MotherNoSon ≡ Mother ⊓ ∀child.Woman
ω_5	MotherNoDaughter ≡ Mother ⊓ ∀child.Man
ω_5	AdoptiveFather ≡ Man ⊓ ∃achild.Person
ω_5	NonAdoptiveFather ≡ Father ⊓ ∀achild.⊥
ω_5	achild ⊑ child

Figure 3. An example $\mathcal{ALEH}$ terminology $\mathcal{O}_{\mathsf{family}}$; here child, achild are shorthands for hasChild and hasAdoptedChild, respectively.

To demonstrate how the algorithm works, we consider the similarity measure between the concepts Mother and NonAdoptiveFather depicted in Figure 2. By using μ^{e}, μ^{a}, γ^{e}, and γ^{a} as previously defined and fixing $\nu^{\mathsf{e}}(r)$ and $\nu^{\mathsf{a}}(r)$ to 0.4 for each $r \in \mathsf{RN}$, the following show the computing steps. Note that, for simplicity, the abbreviations of concept names M and NAF for Mother and NonAdoptiveFather are used, respectively.

$$\begin{aligned} \mathsf{hd}(\mathcal{G}_{\mathsf{M}}, \mathcal{G}_{\mathsf{NAF}}) &:= \tfrac{2}{3}\mathsf{p\_hd}(\mathcal{P}_{\mathsf{M}}, \mathcal{P}_{\mathsf{NAF}}) + \tfrac{1}{3}\mathsf{e\_set\_hd}(\mathcal{E}_{\mathsf{M}}, \mathcal{E}_{\mathsf{NAF}}) + (0)\mathsf{a\_set\_hd}(\mathcal{A}_{\mathsf{M}}, \mathcal{A}_{\mathsf{NAF}}) \\ &:= \tfrac{2}{3}[\tfrac{1}{2}] + \tfrac{1}{3}\mathsf{e\_hd}(\epsilon_i, \epsilon_j) \end{aligned}$$

// with $\mu^{\mathsf{e}} = \frac{1}{3}$, $\mu^{\mathsf{a}} = 0$, $\epsilon_i = \exists\mathsf{child.Person}$ and $\epsilon_j = \exists\mathsf{child.Person}$

$$:= \tfrac{2}{3}[\tfrac{1}{2}] + \tfrac{1}{3}[\tfrac{1}{1}][\tfrac{2}{5} + \tfrac{3}{5}\mathsf{hd}(\mathcal{G}_{\mathsf{Person}}, \mathcal{G}_{\mathsf{Person}})] := \tfrac{2}{3}[\tfrac{1}{2}] + \tfrac{1}{3}[\tfrac{2}{5} + \tfrac{3}{5}[\tfrac{1}{1}]] := 0.67$$

The homomorphism score of the opposite direction is computed as follows:

$$\begin{aligned} \mathsf{hd}(\mathcal{G}_{\mathsf{NAF}}, \mathcal{G}_{\mathsf{M}}) &:= \tfrac{2}{4}\mathsf{p\_hd}(\mathcal{P}_{\mathsf{NAF}}, \mathcal{P}_{\mathsf{M}}) + \tfrac{1}{4}\mathsf{e\_set\_hd}(\mathcal{E}_{\mathsf{NAF}}, \mathcal{E}_{\mathsf{M}}) + \tfrac{1}{4}\mathsf{a\_set\_hd}(\mathcal{A}_{\mathsf{NAF}}, \mathcal{A}_{\mathsf{M}}) \\ &:= \tfrac{2}{4}[\tfrac{1}{2}] + \tfrac{1}{4}\mathsf{e\_hd}(\epsilon_i, \epsilon_j) + \tfrac{1}{4}\mathsf{a\_hd}(\alpha_i, \alpha_j) \end{aligned}$$

// with $\mu^{\mathsf{e}} = \frac{1}{4}$, $\mu^{\mathsf{a}} = \frac{1}{4}$, $\epsilon_i = \exists\mathsf{child.Person}$ and $\epsilon_j = \exists\mathsf{child.Person}$ $\alpha_i = \forall\mathsf{achild}.\bot$ and $\alpha_j = \emptyset$

$$:= \tfrac{2}{4}[\tfrac{1}{2}] + \tfrac{1}{4}[\tfrac{1}{1}][\tfrac{2}{5} + \tfrac{3}{5}\mathsf{hd}(\mathcal{G}_{\mathsf{Person}}, \mathcal{G}_{\mathsf{Person}})] + \tfrac{1}{4}[0] := 0.50$$

By applying the above computation steps, the homomorphism score from $\mathcal{G}_{\mathsf{Mother}}$ to $\mathcal{G}_{\mathsf{NonAdoptiveFather}}$ is 0.67, and that from the $\mathcal{G}_{\mathsf{NonAdoptiveFather}}$ to $\mathcal{G}_{\mathsf{Mother}}$ is 0.50. For the other pairs of concepts defined in $\mathcal{O}_{\mathsf{family}}$, we apply the same steps. Table 2 shows the homomorphism scores among concepts in $\mathcal{O}_{\mathsf{family}}$.

Table 2. Homomorphism scores among defined concepts in $\mathcal{O}_{\mathsf{family}}$.

hd(↓, →)	Woman	Man	Parent	Mother	Father	MNS	MND	AF	NAF
Woman	1.00	0.50	0.50	0.67	0.33	0.50	0.50	0.33	0.25
Man	0.50	1.00	0.50	0.33	0.67	0.25	0.25	0.67	0.50
Parent	0.50	0.50	1.00	0.67	0.67	0.43	0.43	0.50	0.50
Mother	1.00	0.5	1.00	1.00	0.67	0.68	0.68	0.50	0.50
Father	0.50	1.00	1.00	0.67	1.00	0.43	0.43	0.83	0.75
MotherNoSon (MNS)	1.00	0.50	1.00	1.00	0.67	1.00	0.85	0.50	0.55
MotherNoDaughter (MND)	1.00	0.50	1.00	1.00	0.67	0.85	1.00	0.50	0.55
AdoptiveFather (AF)	0.50	1.00	1.00	0.67	1.00	0.43	0.43	1.00	0.75
NonAdoptiveFather (NAF)	0.50	1.00	1.00	0.67	1.00	0.68	0.68	0.83	1.00

By observing the values in Table 2 and by using Proposition 2, it is obvious that that the closer the $\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C)$ is equal to 1, the more likely the subsumption may hold in a reverse direction. Moreover, if $C \sqsubseteq D$, this means that $\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C) = 1$ and vice versa. From Theorem 1 [20, 22], it implies Proposition 2 stated as follows;

Proposition 2. *Let C and D be $\mathcal{ALEH}$ concept descriptions, and $\mathcal{G}_C$ and $\mathcal{G}_D$ be concept description trees, the following are similar:*

1. $\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C) = 1$.
2. $C \sqsubseteq D$,

3.2. $\mathcal{ALEH}$ Semantic Similarity

The homomorphism score function returns a value that represents the similarity of a concept comparing to another concept. The value, however, measures the similarity only in one direction. For example, $\mathsf{hd}(\mathcal{G}_{\mathsf{M}}, \mathcal{G}_{\mathsf{NAF}}) = 0.67$, whereas $\mathsf{hd}(\mathcal{G}_{\mathsf{NAF}}, \mathcal{G}_{\mathsf{M}}) = 0.50$. Since the homomorphism scores of both the forward and the backward direction indicates the similarity score of the two concepts, we therefore define the similarity for $\mathcal{ALEH}$ concept descriptions using the average value. The following Defintion 2 provides the definition of the $\mathcal{ALEH}$ similarity measure. The proposed measure is the average of the homomorphism score in both directions, which ensures that $\mathsf{sim}(C, D) = \mathsf{sim}(D, C)$. Table 3 shows the similarity score among concepts in $\mathcal{O}_{\mathsf{family}}$.

Definition 2. *Let C and D be $\mathcal{ALEH}$ concepts. A similarity score between C and D is calculated as follows:*

$$\mathsf{sim}(C, D) := \frac{\mathsf{hd}(\mathcal{G}_C, \mathcal{G}_D) + \mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C)}{2}, \tag{8}$$

Table 3. Similarity score among defined concepts in $\mathcal{O}_{\mathsf{family}}$.

sim(↓, →)	Woman	Man	Parent	Mother	Father	MNS	MND	AF	NAF
Woman	1.00	0.50	0.50	0.83	0.42	0.75	0.75	0.42	0.38
Man		1.00	0.50	0.42	0.83	0.38	0.38	0.83	0.75
Parent			1.00	0.83	0.83	0.71	0.71	0.75	0.75
Mother				1.00	0.67	0.84	0.84	0.75	0.71
Father					1.00	0.55	0.55	0.92	0.88
MotherNoSon (MNS)						1.00	0.85	0.46	0.61
MotherNoDaughter (MND)							1.00	0.46	0.59
AdoptiveFather (AF)								1.00	0.79
NonAdoptiveFather (NAF)									1.00

Corollary 3. *Let C and D be concept descriptions, $\mathcal{P}_C$ and $\mathcal{P}_D$ be sets of primitive concepts, $\mathcal{E}_C$ and $\mathcal{E}_D$ be sets of existential restrictions, and $\mathcal{A}_C$ and $\mathcal{A}_D$ be sets of universal restrictions. We say that $C \sqsubseteq D$ if*

(a) $\mathcal{P}_D \subseteq \mathcal{P}_C$,

(b) for each $\exists r.D' \in \mathcal{E}_D$ *there exists* $\exists s.C'$ *such that* $s \sqsubseteq^* r$ *and* $C' \sqsubseteq D'$, *and*

(c) for each $\forall r.D' \in \mathcal{A}_D$ *there exists* $\forall s.C'$ *such that* $s \sqsubseteq^* r$ *and* $C' \sqsubseteq D'$.

Corollary 4. *Let* C, C', D, *and* D' *be concept descriptions, we say that* $\mathcal{E}_D \cong \mathcal{E}_C$ *iff for each* $\exists r.D' \in \mathcal{E}_D$ *there exists* $\exists s.C' \in \mathcal{E}_C$ *such that* $s \sqsubseteq^* r$, $r \sqsubseteq^* s$, $C' \sqsubseteq D'$, *and* $D' \sqsubseteq C'$.

Corollary 5. *Let* C, C', D, *and* D' *be concept descriptions, we say that* $\mathcal{A}_D \cong \mathcal{A}_C$ *iff for each* $\forall r.D' \in \mathcal{A}_D$ *there exists* $\forall s.C' \in \mathcal{A}_C$ *such that* $s \sqsubseteq^* r$, $r \sqsubseteq^* s$, $C' \sqsubseteq D'$, *and* $D' \sqsubseteq C'$.

Corollary 6. *Let* C *and* D *be concept descriptions,* $C \equiv D$ *iff* $\mathcal{P}_D = \mathcal{P}_C$, $\mathcal{E}_D \cong \mathcal{E}_C$, *and* $\mathcal{A}_D \cong \mathcal{A}_C$.

3.3. Desirable Properties for Concept Similarity

To identify whether the proposed similarity measure has a good performance, it is important to check the satisfactory of desirable properties. This section describes all important similarity properties and gives mathematical proofs.

Let C, D and E be $\mathcal{ALEH}$ concept descriptions, we say that the similarity measure is:

i. *symmetrical* if $\mathsf{sim}(C, D) = \mathsf{sim}(D, C)$,

ii. *equivalence closed* if $\mathsf{sim}(C, D) = 1$ iff $C \equiv D$,

iii. *equivalence invariant* if $C \equiv D$ then $\mathsf{sim}(C, E) = \mathsf{sim}(D, E)$,

iv. *subsumption preserving* if $C \sqsubseteq D \sqsubseteq E$ then $\mathsf{sim}(C, D) \geq \mathsf{sim}(C, E)$,

v. *reverse subsumption preserving* if $C \sqsubseteq D \sqsubseteq E$ then $\mathsf{sim}(C, E) \leq \mathsf{sim}(D, E)$,

vi. *structurally dependent* if $\lim\limits_{n\to\infty} \mathsf{sim}(D', E') = 1$ where $D' := \prod\limits_{i \leq n} C_i \sqcap D$, $E' := \prod\limits_{i \leq n} C_i \sqcap E$, C_i and C_j are atom comcepts in C where $C_i \not\sqsubseteq C_j$.

vii. *triangle inequality* if $1 + \mathsf{sim}(D, E) \geq \mathsf{sim}(D, C) + \mathsf{sim}(C, E)$.

The proposed similarity measure $\mathsf{sim}(\cdot, \cdot)$ are symmetric, equivalence closed, equivalence invariant, subsumption preserving, structurally dependent, not reverse subsumption preserving, and not satisfying triangle inequality. The following are the proofs

i. From Definition 2, it is obvious that $\mathsf{sim}(C, D) = \mathsf{sim}(D, C)$.

ii. ($\Longrightarrow$) By Equation 8, $\mathsf{sim}(C, D) = 1$ implies that $\mathsf{hd}(\mathcal{G}_C, \mathcal{G}_D) = 1$ and $\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C) = 1$. From Proposition 2, we have $C \sqsubseteq D$ and $D \sqsubseteq C$. Therefore, $C \equiv D$. ($\Longleftarrow$) Given that $C \equiv D$, using the same proposition, we have $C \sqsubseteq D$ and $D \sqsubseteq C$. This implies that $\mathsf{hd}(\mathcal{G}_C, \mathcal{G}_D) = 1$, and $\mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C) = 1$, therefore $\mathsf{sim}(C, D) = 1$.

iii. Given that $C \equiv D$, from Corollary 6, we have $\mathcal{P}_C = \mathcal{P}_D$, $\mathcal{E}_C \cong \mathcal{E}_D$, and $\mathcal{A}_D \cong \mathcal{A}_C$. Therefore, $\mathcal{G}_C = \mathcal{G}_D$ and this implies $\mathsf{hd}(\mathcal{G}_C, \mathcal{G}_E) = \mathsf{hd}(\mathcal{G}_D, \mathcal{G}_E)$ and $\mathsf{hd}(\mathcal{G}_E, \mathcal{G}_C) = \mathsf{hd}(\mathcal{G}_E, \mathcal{G}_D)$. Such that $\mathsf{sim}(C, E) = \mathsf{sim}(D, E)$.

iv. From Definition 2, it is sufficient to prove that

$$\frac{\mathsf{hd}(\mathcal{G}_C,\mathcal{G}_D)+\mathsf{hd}(\mathcal{G}_D,\mathcal{G}_C)}{2} \geq \frac{\mathsf{hd}(\mathcal{G}_C,\mathcal{G}_E)+\mathsf{hd}(\mathcal{G}_E,\mathcal{G}_C)}{2}$$

Given that $C \sqsubseteq D \sqsubseteq E$, this implies that $C \sqsubseteq E$. From Proposition 2, we have $\mathsf{hd}(\mathcal{G}_E, \mathcal{G}_C) = \mathsf{hd}(\mathcal{G}_D, \mathcal{G}_C) = 1$. Therefore, it is sufficient to show that $\mathsf{hd}(\mathcal{G}_C, \mathcal{G}_D) \geq \mathsf{hd}(\mathcal{G}_C, \mathcal{G}_E)$.

On both sides of the inequality, if expanded, we have the same μ^{e} and μ^{a}, where

$$\mu^{\mathsf{e}} = \frac{|\mathcal{E}_C|}{|\mathcal{P}_C \cup \mathcal{E}_C \cup \mathcal{A}_C|}, \text{ and } \mu^{\mathsf{a}} = \frac{|\mathcal{A}_C|}{|\mathcal{P}_C \cup \mathcal{E}_C \cup \mathcal{A}_C|};$$

Therefore, it is enough to prove that

a) $\mathsf{p\_hd}(\mathcal{P}_C, \mathcal{P}_D) \geq \mathsf{p\_hd}(\mathcal{P}_C, \mathcal{P}_E)$
b) $\mathsf{e\_set\_hd}(\mathcal{E}_C, \mathcal{E}_D) \geq \mathsf{e\_set\_hd}(\mathcal{E}_C, \mathcal{E}_E)$
c) and $\mathsf{a\_set\_hd}(\mathcal{A}_C, \mathcal{A}_D) \geq \mathsf{a\_set\_hd}(\mathcal{A}_C, \mathcal{A}_E)$

In a), we need to show that $\frac{|\mathcal{P}_C \cap \mathcal{P}_D|}{|\mathcal{P}_C|} \geq \frac{|\mathcal{P}_C \cap \mathcal{P}_E|}{|\mathcal{P}_C|}$. In short, we need to show that

$$|\ \mathcal{P}_C \cap \mathcal{P}_D\ | \geq |\ \mathcal{P}_C \cap \mathcal{P}_E\ | \tag{9}$$

By Corollary 3, $C \sqsubseteq D \sqsubseteq E$ ensures that $\mathcal{P}_E \subseteq \mathcal{P}_D \subseteq \mathcal{P}_C$. Therefore $|\ \mathcal{P}_D\ | \geq |\ \mathcal{P}_E\ |$ and Equation 9 is true. To prove that b) is true, we show that

$$\sum_{\epsilon_i \in \mathcal{E}_C} \frac{max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_D\}}{|\mathcal{E}_C|} \geq \sum_{\epsilon_i \in \mathcal{E}_C} \frac{max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_E\}}{|\mathcal{E}_C|} \tag{10}$$

$$\sum_{\epsilon_i \in \mathcal{E}_C} max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_D\} \geq \sum_{\epsilon_i \in \mathcal{E}_C} max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_E\}.$$

Let $\hat{\epsilon}_i \in \mathcal{E}_E$ such that $\mathsf{e\_hd}(\epsilon_i, \hat{\epsilon}_i) = max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_E\}$, but since $\hat{\epsilon}_i \in \mathcal{E}_E \subseteq \mathcal{E}_D$, then $max\{\mathsf{e\_hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_D\} \geq \mathsf{e\_hd}(\epsilon_i, \hat{\epsilon}_i)$.

Therefore, Equation 10 is true. By applying the same steps, it implies that c) is also true.

v. Let $D' := \bigsqcap_{i \leq n} C_i \sqcap D$, $E' := \bigsqcap_{i \leq n} C_i \sqcap E$, and $n = n_{\mathcal{P}} + n_{\mathcal{E}} + n_{\mathcal{A}}$ be the number of all atomic sequences in C where $n_{\mathcal{P}}, n_{\mathcal{E}}, n_{\mathcal{A}}$ be the number of primitive concepts, the number of existential restrictions, and the number of universal restrictions, respectively. To prove this, we consider the following case distinctions.

(a) If $n_{\mathcal{P}} \to \infty$, and both $n_{\mathcal{E}}$ and $n_{\mathcal{A}}$ are finite, it suffices to show i) $\lim_{n_{\mathcal{P}} \to \infty} \mu^{\mathsf{e}} = 0$, ii) $\lim_{n_{\mathcal{P}} \to \infty} \mu^{\mathsf{a}} = 0$ and iii) $\lim_{n_{\mathcal{P}} \to \infty} \mathsf{p\_hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) = 1$. Therefore, $\mathsf{hd}(\mathcal{G}_{D'}, \mathcal{G}_{E'}) = \mathsf{hd}(\mathcal{G}_{E'}, \mathcal{G}_{D'}) = 1$ which implies $\mathsf{sim}(D', E') = 1$. Starting from

$$\mu^{\mathsf{e}} = \frac{|\mathcal{E}_{D'}|}{|\mathcal{P}_{D'} \cup \mathcal{E}_{D'} \cup \mathcal{A}_{D'}|} = \frac{|\mathcal{E}_{D'}|}{|\mathcal{P}_C| + |\mathcal{P}_D| + |\mathcal{E}_{D'}| + |\mathcal{A}_{D'}|} = \frac{|\mathcal{E}_{D'}|}{n_{\mathcal{P}} + |\mathcal{P}_D| + |\mathcal{E}_{D'}| + |\mathcal{A}_{D'}|} \tag{11}$$

since $|\ \mathcal{P}_D\ |$, $|\ \mathcal{E}_{D'}\ |$ and $|\ \mathcal{A}_{D'}\ |$ are constants, $\lim_{n_{\mathcal{P}} \to \infty} \mu^{\mathsf{e}} = \lim_{n_{\mathcal{P}} \to \infty} \frac{|\mathcal{E}_{D'}|}{n_{\mathcal{P}} + |\mathcal{P}_D| + |\mathcal{E}_{D'}| + |\mathcal{A}_{D'}|} = 0$. To show ii) $\lim_{n_{\mathcal{P}} \to \infty} \mu^{\mathsf{a}} = 0$, consider the formula defining

$$\mu^{\mathsf{a}} = \frac{|\mathcal{A}_{D'}|}{|\mathcal{P}_{D'} \cup \mathcal{E}_{D'} \cup \mathcal{A}_{D'}|} = \frac{|\mathcal{A}_{D'}|}{n_{\mathcal{P}} + |\mathcal{P}_D| + |\mathcal{E}_{D'}| + |\mathcal{A}_{D'}|}$$

Therefore,

$$\lim_{n_{\mathcal{P}} \to \infty} \mu^{\mathsf{a}} = \lim_{n_{\mathcal{P}} \to \infty} \frac{|\mathcal{A}_{D'}|}{n_{\mathcal{P}} + |\mathcal{P}_D| + |\mathcal{E}_{D'}| + |\mathcal{A}_{D'}|} = 0.$$

For iii) $\lim_{n_{\mathcal{P}} \to \infty} \mathsf{p\_hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) = 1$, we consider the definition of $\mathsf{p\_hd}$.

$$\mathsf{p\_hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) \;=\; \frac{|\mathcal{P}_{D'} \cap \mathcal{P}_{E'}|}{|\mathcal{P}_{D'}|} = \frac{|\mathcal{P}_C| + |\mathcal{P}_D \cap \mathcal{P}_E|}{|\mathcal{P}_C| + |\mathcal{P}_D|} = \frac{n_\mathcal{P} + |\mathcal{P}_D \cap \mathcal{P}_E|}{n_\mathcal{P} + |\mathcal{P}_D|}$$

where $|\, \mathcal{P}_D \,\cap\, \mathcal{P}_E \,|$ and $|\, \mathcal{P}_D \,|$ are constants. Thus,

$$\lim_{n_\mathcal{P}\to\infty} \mathsf{p\_hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) = \lim_{n_\mathcal{P}\to\infty} \frac{n_\mathcal{P} + \mid \mathcal{P}_D \,\cap\, \mathcal{P}_E \mid}{n_\mathcal{P} + \mid \mathcal{P}_D \mid} = 1. \tag{12}$$

(b) If $n_\mathcal{E} \to \infty$, $n_\mathcal{A}$ and $n_\mathcal{P}$ are finite, it suffices to show that $\lim_{n_\mathcal{E}\to\infty} \mu^\mathsf{e} = 1$ and $\lim_{n_\mathcal{E}\to\infty} \mathsf{e\_set\_hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) = 1$ which implies $\mathsf{hd}(\mathcal{G}_{D'}, \mathcal{G}_{E'}) = \mathsf{hd}(\mathcal{G}_{E'}, \mathcal{G}_{D'}) = 1$, and $\mathsf{sim}(D', E') = 1$. From Equation 3.1., we have μ^e of the following form:

$$\begin{aligned}\mu^\mathsf{e} \;&=\; \frac{|\mathcal{E}_{D'}|}{|\mathcal{P}_{D'} \cup \mathcal{E}_{D'} \cup \mathcal{A}_{D'}|} = \frac{|\mathcal{E}_C| + |\mathcal{E}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + |\mathcal{E}_C| + |\mathcal{E}_D| + |\mathcal{A}_C| + |\mathcal{A}_D|} \\ &=\; \frac{n_\mathcal{E} + |\mathcal{E}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + n_\mathcal{E} + |\mathcal{E}_D| + |\mathcal{A}_C| + |\mathcal{A}_D|}\end{aligned} \tag{13}$$

Since $|\, \mathcal{P}_C \,|, |\, \mathcal{P}_D \,|, |\, \mathcal{E}_D \,|, |\, \mathcal{A}_C \,|$, and $|\, \mathcal{A}_D \,|$ are constants, by taking limit, we have

$$\lim_{n_\mathcal{E}\to\infty} \mu^\mathsf{e} = \lim_{n_\mathcal{E}\to\infty} \frac{n_\mathcal{E} + |\mathcal{E}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + n_\mathcal{E} + |\mathcal{E}_D| + |\mathcal{A}_C| + |\mathcal{A}_D|} = 1$$

To show that $\lim_{n_\mathcal{E}\to\infty} \mathsf{e\_set\_hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) = 1$, we have

$$\begin{aligned}\mathsf{e\_set\_hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) \;&=\; \sum_{e_i \in \mathcal{E}_{D'}} \frac{\max\{\mathsf{e\_hd}(e_i, e_j) : e_j \in \mathcal{E}_{E'}\}}{|\mathcal{E}_{D'}|} = \frac{\sum_{e_i \in \mathcal{E}_{D'}} \max\{\mathsf{e\_hd}(e_i, e_j) : e_j \in \mathcal{E}_{E'}\}}{|\mathcal{E}_{D'}|} \\ &=\; \frac{\sum_{e_i \in \mathcal{E}_C} \max\{\mathsf{e\_hd}(e_i, e_j) : e_j \in \mathcal{E}_{E'}\}}{|\mathcal{E}_C \cup \mathcal{E}_D|} + \frac{\sum_{e_i \in \mathcal{E}_D} \max\{\mathsf{e\_hd}(e_i, e_j) : e_j \in \mathcal{E}_{E'}\}}{|\mathcal{E}_C \cup \mathcal{E}_D|}\end{aligned}$$

Since $\mathcal{E}_C \subseteq \mathcal{E}_{E'}$, for each $\epsilon_i \in \mathcal{E}_C$ there exists $\epsilon_j \in \mathcal{E}_{E'}$ such that $\epsilon_i = \epsilon_j$. Thus,

$$\mathsf{e\_set\_hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) \;=\; \frac{n_\mathcal{E} + p}{|\mathcal{E}_C| + |\mathcal{E}_D|} = \frac{n_\mathcal{E} + p}{n_\mathcal{E} + |\mathcal{E}_D|}$$

where $p = \sum_{e_i \in \mathcal{E}_D} \max\{\mathsf{e\_hd}(e_i, e_j) : e_j \in \mathcal{E}_{E'}\}$, and $p \;\leq |\, \mathcal{E}_D \,|$. Therefore, the following relationship is true.

$$\lim_{n_\mathcal{E}\to\infty} \mathsf{e\_set\_hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) = \lim_{n_\mathcal{E}\to\infty} \frac{n_\mathcal{E} + p}{n_\mathcal{E} + \mid \mathcal{E}_D \mid} = 1.$$

(c) If $n_\mathcal{A} \to \infty$, and $n_\mathcal{E}$ and $n_\mathcal{P}$ are finite, by applying the same steps as described in (b), it is obvious that $\lim_{n_\mathcal{A}\to\infty} \mu^\mathsf{a} = 1$ and $\lim_{n_\mathcal{A}\to\infty} \mathsf{a\_set\_hd}(\mathcal{A}_{D'}, \mathcal{A}_{E'}) = 1$ which implies $\mathsf{hd}(\mathcal{G}_{D'}, \mathcal{G}_{E'}) = \mathsf{hd}(\mathcal{G}_{E'}, \mathcal{G}_{D'}) = 1$, and $\mathsf{sim}(D', E') = 1$.

(d) For the other cases (e.g. $n_\mathcal{P} \to \infty$, $n_\mathcal{E} \to \infty$, and $n_\mathcal{A} \to \infty$), these follow the previous cases and can be concluded that $\mathsf{sim}(D', E') = 1$.

vi. To prove this, we show a counter example. Given concepts C, D, and E as defined in Figure 4. From the definitions, it is obvious that $C \sqsubseteq D \sqsubseteq E$. By computing, we have $\mathsf{sim}(C, E)$:= 0.7125 and $\mathsf{sim}(D, E)$:= 0.6667. Hence, there is the case in which $\mathsf{sim}(C, E) \not\leq \mathsf{sim}(D, E)$.

$$\begin{aligned}\mathsf{C} \;&\equiv\; \exists \mathsf{r}.(\mathsf{G} \sqcap \mathsf{H}) \sqcap \exists s.\mathsf{G} \sqcap \exists s.\mathsf{H} \sqcap \exists \mathsf{r}.(\mathsf{G} \sqcap \mathsf{I}) \\ \mathsf{D} \;&\equiv\; \exists \mathsf{r}.(\mathsf{G} \sqcap \mathsf{H}) \sqcap \exists s.\mathsf{G} \sqcap \exists s.\mathsf{H} \\ \mathsf{E} \;&\equiv\; \exists \mathsf{r}.(\mathsf{G} \sqcap \mathsf{H})\end{aligned}$$

Figure 4. Examples of concept descriptions.

vii. Given the concept descriptions C, D, and E shown in Figure 4, we have $1 + \mathsf{sim}(D, E) \ngeq \mathsf{sim}(D, C) + \mathsf{sim}(C, E)$. We have $\mathsf{sim}(D, E)$:= 0.6667, $\mathsf{sim}(D, C)$:= 0.9625, and $\mathsf{sim}(C, E)$:= 0.7125. It is obvious that 1 + 0.6667 $\ngeq$ 1.675.

To ensure that the proposed method have hold sufficient properties comparing to those reported in the-state-of-the-art methods, Table 4 lists all properties of sim and compares to those reports in other methods. It is obvious that the proposed method together with that of Lehmann and Turhan [15] have hold several significant features.

Table 4. A comparison of desirable properties of different concept similarity measures.

Similarity measure methods	DL	symmetric	equivalent closed	equivalent invariant	sub. preserving	structural dependent	reverse subsumption preserving	triangle inequality
sim	$\mathcal{ALEH}$	✓	✓	✓	✓	✓		
Lehmann and Turhan [15]	$\mathcal{ELH}$	✓	✓	✓	✓	✓		
Janowicz and Wilkes [17]	$\mathcal{SHI}$	✓						✓
d'Amato et al. [19]	$\mathcal{ALC}$	✓		✓		✓	✓	
d'Amato et al. [18]	$\mathcal{ALE}$	✓		✓		✓	✓	
Janowicz [16]	$\mathcal{ALCHQ}$	✓						✓
d'Amato et al. [23]	$\mathcal{ALC}$							
Fanizzi and d'Amato [24]	$\mathcal{ALN}$	✓		✓		✓	✓	

Table 5. The number of concepts in each category of SNOMED CT.

Snomed-concept categories	Number of concepts
Social context	10,482
Procedure	54,624
Physical force	798
Substance	61,083
Body structure	99,262
Specimen	36
Situation	1,529
Attribute	1,121
Staging and scales	1,108
Physical object	4,351
Event	1,641
Environment	1,665
Qualifier value	12,144
Observable entity	15,228
Special concept	63,660
Pharmaceutical product	4,329
Clinical finding	20,798
Organism	25,810

4. EXPERIMENTS ON SNOMED CT

SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) is a large-scale clinical knowledge base, which stores definitions of clinical terms used by physicians. The terminology is divided into 18 categories containing 379,691 clinical terms. Table 5 depicts the majority of concepts in each category.

4.1. Computing similarity score between SNOMED CT concepts

In this experiment, we randomly select 30 concepts from each concept category and compute the similarity score for each possible pair. For each category, we compute similarity scores for 16,200 pairs of concepts. Table 6 reports the average similarity scores between concept categories. From the results, it is obvious that the average scores of concepts from the same category are higher than those from different categories. This is because the shared structures in the same category are higher than those across categories.

As shown in Table 6, the average similarity scores of concepts from different categories become zero (or nearly) since they are mostly out of subsumption relations (i.e., there is no shared feature). The accumulated time taken for each concept category is measured and reported in Figure 5. From the graph, it reveals that the time taken by the *Procedure* and *Body structure* category are a bit higher than those of other categories. This should infer that these two categories are likely to have complicated concept definitions, which would require some time for both the description tree construction and the similarity measure.

To ensure that the time used by the system is reasonable, we randomly select 10000 pairs of concepts C and D from SNOMED CT (C and D are different concepts). Table 7 shows the number of concept pairs. One may have noticed that the number of concept pairs between each category is directly proportional to the majority of concepts reported in Table 5. In addition, for each pair of C and D, we measure the time taken by the system. From the results reported in Figure 6, it is obvious that the system requires only a few seconds. Though, the worst case has shown that the time taken by the system can be up to 18 seconds, this is a rare case which occurs only once during the entire experiment. Figure 7 shows that the accumulated time required for all 10000 of concept pairs is merely 343 seconds, which is about 0.034 second in average.

5. SYSTEM USABILITY EVALUATION

By performing experiments on the large scale and so-called SNOMED CT ontology, in this section we describe the correctness of the proposed similarity measure through a manual assessment by physicians. For each concept category K in SNOMED CT, we randomly select concepts $D_1, D_2, \ldots, D_{10}$. Moreover, for each concept D_i, we randomly select concepts $F_1, F_2, ...F_{10}$ from the same category such that $D_i \not\sqsubseteq F_i$ and $F_i \not\sqsubseteq D_i$. This is in order to focus on overall concept similarity and not merely similarity due to concept subsumption. Figure 8 shows the example of questions. We then let the physicians manually pick the most similar concept H out of 10 randomized concepts $F_1, F_2, \ldots, F_{10}$ where $F_1, F_2, \ldots, F_{10}$ are ordered descendingly according to the similarity score to the given concept D_i. The following score is computed.

$$score = \begin{cases} 1.0 & \text{if } H = F_1 \\ 0.5 & \text{if } H = F_2 \\ 0.33 & \text{if } H = F_3 \\ 0.0 & otherwise \end{cases}$$

As shown in Table 8, it is obvious that the average score obtained from each physician is slightly variant and essentially depended on the background knowledge of the physicians. However, in average, they are all above 0.5. This indicates that the choices selected by the 4 physicians are mostly within the first three most similar concepts F.

Table 6. The average similarity scores between each concept category.

Snomed-concept categories	Snomed-concept categories																	
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
Social context (C1)	0.2173	0.0087	0	0	0	0	0.0522	0	0	0	0	0	0	0	0	0	0	0
Procedure (C2)	0.0087	0.3294	0.0190	0.0026	0.0161	0	0	0	0	0	0	0	0.0032	0.0025	0	0.0010	0.0081	0
Physical force (C3)	0	0.0190	0.3875	0.0313	0.2735	0	0	0	0	0	0	0	0.0434	0.0323	0	0.0152	0.0917	0
Substance (C4)	0	0.0026	0.0313	0.1723	0.0160	0	0	0	0	0	0	0	0.0079	0.0057	0	0.1176	0.0169	0
Body structure (C5)	0	0.0161	0.2735	0.0160	0.3254	0	0	0	0	0	0	0	0.038	0.0369	0	0.0068	0.0892	0
Specimen (C6)	0	0	0	0	0	0.4938	0	0	0	0	0	0	0	0	0	0	0	0
Situation (C7)	0.0522	0	0	0	0	0	0.3481	0	0	0	0	0	0	0	0	0	0	0
Attribute (C8)	0	0	0	0	0	0	0	0.5352	0	0	0	0	0	0	0	0	0	0
Staging and scales (C9)	0	0	0	0	0	0	0	0	0.4917	0	0	0	0	0	0	0	0	0
Physical object (C10)	0	0	0	0	0	0	0	0	0	0.3934	0	0	0	0	0	0	0	0
Event (C11)	0	0	0	0	0	0	0	0	0	0	0.4108	0	0	0	0	0	0	0
Environment (C12)	0	0	0	0	0	0	0	0	0	0	0	0.3219	0	0	0	0	0	0
Qualifier value (C13)	0	0.0032	0.0434	0.0079	0.0380	0	0	0	0	0	0	0	0.1578	0.0360	0	0.0035	0.0537	0
Observable entity (C14)	0	0.0025	0.0323	0.0057	0.0369	0	0	0	0	0	0	0	0.036	0.1670	0	0.0024	0.0794	0
Special concept (C15)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5988	0	0	0
Pharmaceutical product (C16)	0	0.0010	0.0152	0.1176	0.0068	0	0	0	0	0	0	0	0.0035	0.0024	0	0.4294	0.0072	0
Clinical finding (C17)	0	0.0081	0.0917	0.0169	0.0892	0	0	0	0	0	0	0	0.0537	0.0794	0	0.0072	0.2266	0
Organism (C18)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2361

Table 7. The number of concept pairs randomly selected from SNOMED CT.

Snomed-concept categories	Snomed-concept categories																	
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
Social context (C1)	8	41	0	50	75	0	0	2	0	4	0	1	10	11	54	6	9	12
Procedure (C2)	48	226	3	217	363	0	5	5	1	15	6	7	63	63	225	18	79	93
Physical force (C3)	0	0	0	2	7	0	1	0	0	0	0	0	0	1	4	0	3	1
Substance (C4)	47	239	2	264	408	0	4	4	3	18	1	6	58	65	249	15	81	116
Body structure (C5)	75	360	14	447	623	0	8	8	11	29	13	11	73	116	488	34	128	162
Specimen (C6)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Situation (C7)	1	4	0	10	7	0	0	0	0	0	0	0	1	0	14	0	3	2
Attribute (C8)	1	1	0	1	8	0	0	0	0	0	0	0	0	3	1	1	2	1
Staging and scales (C9)	1	11	0	4	6	0	0	0	0	0	0	0	3	0	7	0	0	2
Physical object (C10)	5	14	0	24	33	0	0	0	0	0	1	1	3	3	26	1	10	6
Event (C11)	1	8	0	7	14	0	0	0	0	0	0	0	0	0	7	1	2	4
Environment (C12)	1	8	0	5	6	0	0	0	0	0	0	0	0	2	5	0	1	7
Qualifier value (C13)	8	59	0	65	71	0	0	0	0	4	2	4	9	13	58	1	23	24
Observable entity (C14)	10	55	0	59	94	0	1	2	1	4	0	1	10	15	71	4	23	28
Special concept (C15)	37	230	4	259	444	0	8	6	4	16	9	7	41	66	280	13	94	114
Pharmaceutical product (C16)	3	15	1	14	24	0	0	1	0	3	0	2	4	5	18	3	5	7
Clinical finding (C17)	15	78	0	96	150	0	2	3	2	2	0	1	19	23	117	5	39	38
Organism (C18)	21	107	0	109	200	0	5	3	4	6	3	1	28	32	119	11	43	41

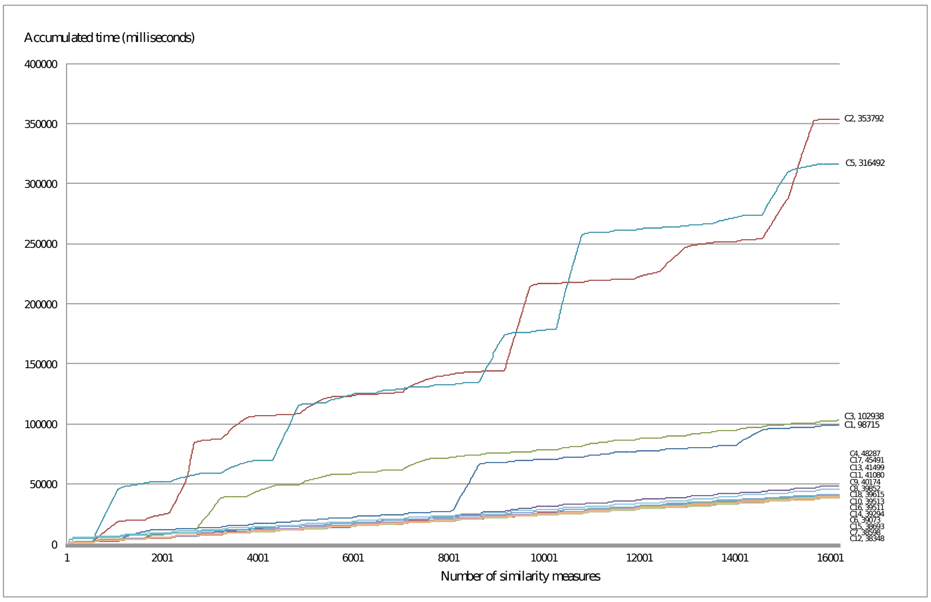

Figure 5. The accumulated time for each concept category.

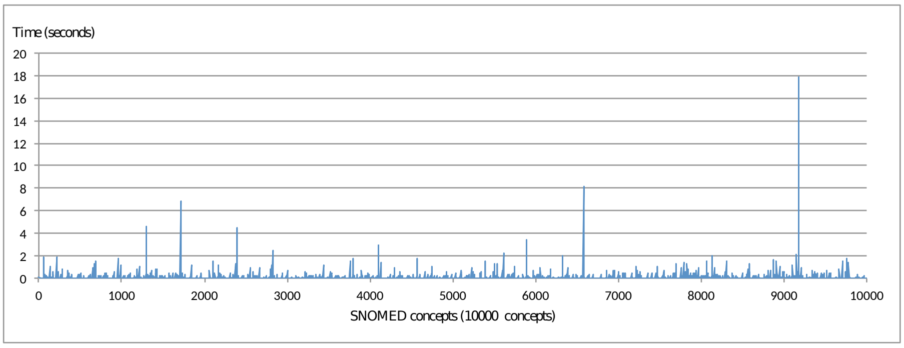

Figure 6. The time used for similarity measure for each pair of concept.

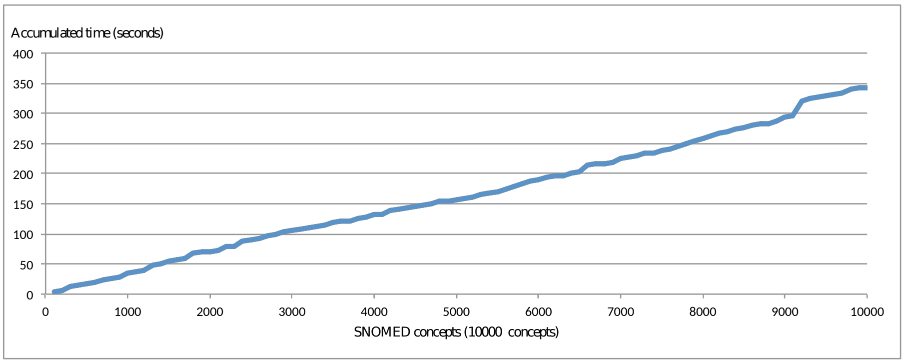

Figure 7. The accumulated time used for computing similarity for 10000 concept pairs.

Table 8. Average similarity scores of medical terms evaluated by 4 physicians.

	Physician ♯1	Physician ♯2	Physician ♯3	Physician ♯4
Average similarity score	0.599	0.783	0.716	0.860

```
Instruction: From the given keyword, select the most similar concept
from the given list.

     Keyword: "Exposure of tooth pulp"
       1) Abrasion or friction burn of gum with infection
       2) Cutaneous actinobacillosis of sheep AND/OR cattle
       3) On ground bystander victim in aircraft accident
       4) Edema bullosum vesicae
       5) Open fracture of ilium
       6) Hemolytic jaundice
       7) Onychotillomania
       8) Displacement of cardiac device
       9) Edema of nasopharynx
       10) Alveolitis of jaw
```

Figure 8. Example of questions for physicians to evaluate.

6. CONCLUSION

This paper presents a new method for measuring similarity between concepts in an ontology. The score obtained from the proposed method is in a range between 0 and 1 which will facilitate users to adopt a good strategy for concept categorization. To demonstrate the usability of the proposed algorithm, the well-known clinical domain SNOMED CT is employed in our experiments. The effectiveness of the proposed algorithm has been revealed by means of desirable properties, the time consumption, and the corresponding results obtained from the assessment by the domain experts. Apart from the scenario expressed in the experiments, we believe that the proposed similarity measure can be useful in different cases. Examples include a checking for similarity between genes, an identification for a disease with similar symptoms, and a checking for similarity between texts known as plagiarism checking.

For future works, there are some possible steps which we can focus on. One is an extension to more expressive features, such as a concept disjunction and a concept negation. The other possible direction is to enhance the algorithm to work with cyclic TBox, such as the one containing a set of general concept inclusions and cyclic definitions. Moreover, it is also possible to extend the capability of the proposed method to measure similarity between individuals. An example application is a query for the most similar image given a sample image.

REFERENCES

[1] J. Bailey, F. Bry, T. Furche, and S. Schaffert, "Web and semantic web query languages: A survey," in *Proceedings of the First International Conference on Reasoning Web*. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 35–133.

[2] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[3] D. Nardi and R. Brachman, "Chapter 1: An introduction to Description Logics," in *The Description Logic Handbook [25]*. Cambridge University Press, 2007, pp. 1–43.

[4] F. Baader and W. Nutt, "Chapter 2: Basic Description Logics," in *The Description Logic Handbook [25]*. Cambridge University Press, 2007, pp. 47–104.

[5] W. Recommendation, "OWL 2 web ontology language: Document overview," 2009.

[6] B. Motik, B. C. Grau, I. Horrocks, Z. Wu, A. Fokoue, and C. Lutz, "OWL 2 web ontology language profiles," 2009.

[7] C. Lutz, F. Baader, E. Franconi, D. Lembo, R. Möller, R. Rosati, U. Sattler, B. Suntisrivaraporn, and S. Tessaris, "Reasoning support for ontology design," in *Proceedings of the second international workshop OWL: Experiences and Directions*, B. C. Grau, P. Hitzler, C. Shankey, and E. Wallace, Eds., 2006.

[8] F. Nurifan, R. Sarno, and C. S. Wahyuni, "Developing corpora using word2vec and wikipedia for word sense disambiguation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 12, pp. 1239–1246, 2018.

[9] M. J. Basha and K. Kaliyamurthie, "An improved similarity matching based clustering framework for short and sentence level text," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, pp. 551–558, 2017.

[10] Y. Al Amrani, M. Lazaar, and E. Kamal El Kadiri, "A novel hybrid classification approach for sentiment analysis of text document," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, pp. 4554–4567, 2018.

[11] F. Distel, J. Atif, and I. Bloch, "Concept dissimilarity with triangle inequality," in *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'14)*. Vienna, Austria: AAAI Press, 2014, short Paper. To appear.

[12] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.

[13] T. S. De Silva, D. MacDonald, G. Paterson, K. C. Sikdar, and B. Cochrane, "Systematized nomenclature of medicine clinical terms (snomed ct) to represent computed tomography procedures," *Comput. Methods Prog. Biomed.*, vol. 101, no. 3, pp. 324–329, Mar. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2011.01.002

[14] T. Racharak and B. Suntisrivaraporn, "Similarity measures for fl 0 concept descriptions from an automata-theoretic point of view," in *Information and Communication Technology for Embedded Systems (IC-ICTES), 2015 6th International Conference of.* IEEE, 2015, pp. 1–6.

[15] K. Lehmann and A.-Y. Turhan, "A framework for semantic-based similarity measures for $\mathcal{ELH}$ -concepts." in *JELIA*, ser. Lecture Notes in Computer Science, L. F. del Cerro, A. Herzig, and J. Mengin, Eds., vol. 7519. Springer, 2012, pp. 307–319.

[16] K. Janowicz, *Sim-DL: Towards a Semantic Similarity Measurement Theory for the Description Logic $\mathcal{ALCNR}$ in Geographic Information Retrieval*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1681–1692.

[17] K. Janowicz and M. Wilkes, *SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 353–367.

[18] C. d'Amato, S. Staab, and N. Fanizzi, *On the Influence of Description Logics Ontologies on Conceptual Similarity*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 48–63.

[19] C. d'Amato, N. Fanizzi, and F. Esposito, "A semantic similarity measure for expressive description logics," *CoRR*, vol. abs/0911.5043, 2009.

[20] F. Baader and R. Küsters, "Matching in description logics with existential restrictions," in *Proceedings of the Seventh International Conference on Knowledge Representation and Reasoning (KR2000)*, A.G. Cohn, F. Giunchiglia, and B. Selman, Eds. San Francisco, CA: Morgan Kaufmann Publishers, 2000, pp. 261–272.

[21] F. Baader and R. Küsters, "Nonstandard inferences in description logics: The story so far," in *Mathematical Problems from Applied Logic I*, ser. International Mathematical Series, D. Gabbay, S. Goncharov, and M. Zakharyaschev, Eds. Springer-Verlag, 2006, vol. 4, pp. 1–75.

[22] F. Baader, "Terminological cycles in a description logic with existential restrictions," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, G. Gottlob and T. Walsh, Eds. Morgan Kaufmann, 2003, pp. 325–330.

[23] C. d'Amato, N. Fanizzi, and F. Esposito, "A dissimilarity measure for alc concept descriptions," in *Proceedings of the 2006 ACM Symposium on Applied Computing*, ser. SAC '06. New York, NY, USA: ACM, 2006, pp. 1695–1699.

[24] N. Fanizzi and C. DAmato, "A similarity measure for the $\mathcal{ALN}$ description logic," *Proceedings of CILC 2006 - Italian Conference on Computational Logic*, pp. 26–27, 2006.

[25] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, Eds., *The Description Logic Handbook: Theory, Implementation and Applications*, 2nd ed. Cambridge University Press, 2007.