On Desirable Properties of the Structural Subsumption-Based Similarity Measure

Suwan Tongphu and Boontawee Suntisrivaraporn(✉)

School of Information, Computer and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University,
Phra Nakhon Si Ayutthaya, Pathumthani, Thailand
stongphu@gmail.com, sun@siit.tu.ac.th

Abstract. Checking for subsumption relation is the main reasoning service readily available in classical DL reasoners. With their binary response stating whether two given concepts are in the subsumption relation, it is adequate for many applications relied on the service. However, in several specific applications, there often exists the case that requires an investigation for concepts that are not directly in a subclass-superclass relation but shared some commonality. In this case, providing merely a crisp response is apparently insufficient. To achieve this, the similarity measure for DL $\mathcal{ELH}$, which is inspired by the homomorphism-based structural subsumption characterization, has been introduced. To ensure that the proposed method reaches the performance, in this work, desirable properties for concept similarity measure are checked and compared with those previously reported in other classical works.

Keywords: Concept similarity · Non-standard reasoner · Description logic · Structural subsumption

1 Introduction

Knowledge representation is one such major research area that has a long range of development and mainly focuses on an investigation for well-founded ways to model, share, and interpret the knowledge. One modeling formalism is an exploitation of the family of Description Logics (DLs) which allows various types of reasoning services. Among those readily available in classical DL reasoners, concept subsumption (i.e. identification of subclass-superclass relationships) is one of the most prominent services. Despite its usefulness, classical subsumption reasoners merely response with a binary result (i.e. whether two given concepts are in a subclass-superclass relation). This capability seems adequate for many applications. However, in some situations, there may be the case that the two concepts do not align that way but share some commonality. This special case, on the other hand, turns into account in many specific applications. For example, in hospitals, once a doctor has diagnosed medical conditions of a patient and identified what the illness is, he/she may need to investigate further for other possible illnesses of similar but not exactly the same conditions.

T. Supnithi et al.(Eds.): JIST 2014, LNCS 8943, pp. 19–32, 2015.
DOI: 10.1007/978-3-319-15615-6_2

Table 1. Syntax and semantics of the Description Logic $\mathcal{ELH}$

Name	Syntax	Semantics
top	$\top$	$\Delta^{\mathcal{I}}$
concept name	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
concept inclusion	$A \sqsubseteq D$	$A^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
concept equivalence	$A \equiv C$	$A^{\mathcal{I}} = C^{\mathcal{I}}$
role inclusion	$r \sqsubseteq s$	$r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$

Our first introduction to $\mathcal{EL}$ concept similarity measure [13] and its sample application [8] have shown its usability specifically in but not limited to one of the most popular medical-domain ontologies, SNOMED CT [12]. In this work, we extend the algorithm to a more expressive DL $\mathcal{ELH}$. Therefore, role names of the same hierarchy are taken into account. Moreover, to ensure that the proposed method reaches the performance and holds satisfactory features, in this work, desirable properties for concept similarity measure are proofed and compared with those previously reported in other works.

In the next section, notions of the DL and necessary backgrounds are introduced. Section 3 and Section 4 provide details on the proposed method and its properties, respectively. Concluding remarks are given in the last section.

2 Preliminaries

Let CN and RN be a set of concept names and a set of role names. In Description Logics (DLs), complex $\mathcal{ELH}$ concept descriptions can be built using a set of constructors shown in the upper part of Table 1. The background knowledge about the domain called *terminology box* or *TBox* can then be devised using a set of ontological axioms shown in the second part of Table 1. A TBox is *unfoldable* if, for each concept name, there is only one concept definition and there is neither direct nor indirect concept definition that refers to the concept itself. Figure 1 shows an example of the unfoldable Tbox $\mathcal{O}_{\mathsf{med}}$.

An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ comprises of *interpretation domain* $\Delta^{\mathcal{I}}$ and *interpretation function* $\cdot^{\mathcal{I}}$. The interpretation function maps every concept name $A \in \mathsf{CN}$ to a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, every role name $r \in \mathsf{RN}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and every individual $x \in \mathsf{Ind}$ to an element $x^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. The last column of Table 1 depicts the semantics for $\mathcal{ELH}$ constructors and terminological axioms. An interpretation $\mathcal{I}$ is a model of a TBox $\mathcal{O}$ if it satisfies every axiom defined in $\mathcal{O}$. Let C, D and E be concept descriptions, C is subsumed by D

(written $C \sqsubseteq D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ in every model $\mathcal{I}$ and if $C \sqsubseteq D \sqsubseteq E$, then $C \sqsubseteq E$. Moreover, C, D are equivalent (written $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$, i.e. $C^{\mathcal{I}} = D^{\mathcal{I}}$ for all interpretations $\mathcal{I}$.

By introducing a set of fresh concept names [13], a concept inclusion can be transformed to an equivalent form. Without losing of generality, we assume that all concept names can be expanded (i.e. they can be replaced by the definition) and has the following form:

$$\bigsqcap_{i \leq m} P_i \sqcap \bigsqcap_{j \leq n} \exists .r_j C_j$$

where $1 \leq i \leq m$ and $1 \leq j \leq n$. This is true both for defined concepts (i.e. concepts that appear on the left-hand side of a definition) and primitive concepts (i.e. concepts that appear only on the right-hand side of a definition). For any primitive concept P, $P \sqsubseteq \top$ therefore $P \equiv X \sqcap \top \equiv X$ where X is a fresh concept name. For convenience, we denote by $\mathcal{P}_C = \{P_1, \ldots, P_m\}$ and $\mathcal{E}_C = \{\exists r_1 C_1, \ldots, \exists r_n C_n\}$ the set of top-level primitive concepts and the set of top-level existential restrictions. Also, we denote by $\mathcal{R}_r = \{s | r \sqsubseteq^* s\}$ the set of all super roles where r and s are role names, $*$ is the transitive closure, and $r \sqsubseteq^* s$ if $r = s$ or $r_i \sqsubseteq r_{i+1} \in \mathcal{O}$ where $r_1 = r$ and $r_n = s$. The following demonstrates the expanded form of the concept $\mathsf{AspirationOfMucus}$ defined in $\mathcal{O}_{\mathsf{med}}$ (see Figure 1).

$$X \sqcap \mathsf{RespiratoryDisorder} \sqcap \exists \mathsf{agent.Mucus}$$

where X is a fresh concept name.

Let $\mathcal{T} = (V, E, rt, \ell, \rho)$ be the $\mathcal{ELH}$ description tree [13] w.r.t. an unfoldable TBox, where V is a set of nodes, $E \subseteq V \times V$ is a set of edges, rt is the root, $\ell : V \to 2^{\mathsf{CN}^{\mathsf{pri}}}$ is a node labeling function, and $\rho : E \to 2^{\mathsf{RN}}$ is an edge labeling function. Definition 1 defines a homomorphism mapping. Let $\mathcal{T}_C$ and $\mathcal{T}_D$ be $\mathcal{ELH}$ description trees w.r.t. the concept C and D, Theorem 1 depicts a characterization of $C \sqsubseteq D$ based on a homomorphism that maps the root of $\mathcal{T}_D$ to the root of $\mathcal{T}_C$.

Definition 1 (Homomorphism). *Let $\mathcal{T}$ and $\mathcal{T}'$ be two $\mathcal{ELH}$ description trees as previously defined. There exists a homomorphism h from $\mathcal{T}$ to $\mathcal{T}'$ written $h : \mathcal{T} \to \mathcal{T}'$ iff the following conditions are satisfied:*

(i) $\ell(v) \subseteq \ell'(h(v))$
(ii) *For each successor w of v in $\mathcal{T}$, $h(w)$ is a successor of $h(v)$ with $\rho(v, w) \subseteq \rho'(h(v), h(w))$*

Theorem 1 ([9]). *Let C, D be $\mathcal{ELH}$ concept descriptions and $\mathcal{T}_C, \mathcal{T}_D$ be $\mathcal{ELH}$ concept description trees w.r.t. C and D. Then, $C \sqsubseteq D$ iff there exists a homomorphism $h : \mathcal{T}_D \to \mathcal{T}_C$ which maps the root of $\mathcal{T}_D$ to the root of $\mathcal{T}_C$.*

By using Theorem 1 together with properties of homomorphism mapping defined in Definition 1, Corollary 1 and Corollary 2 hold due to an associativity and commutativity of concept conjunction.

Corollary 1. *Let C and D be concept names. Then $C \sqsubseteq D$ iff $\mathcal{P}_D \subseteq \mathcal{P}_C$ and for each $\exists r.D' \in \mathcal{E}_D$ there exists $\exists s.C'$ such that $s \sqsubseteq^* r$ and $C' \sqsubseteq D'$.*

Corollary 2. *Let C and D be concept names, then $\mathcal{E}_D \cong \mathcal{E}_C$ iff for each $\exists r.D' \in \mathcal{E}_D$ there exists $\exists s.C'$ such that $s \sqsubseteq^* r$, $r \sqsubseteq^* s$, $C' \sqsubseteq D'$, and $D' \sqsubseteq C'$. Moreover, $C \equiv D$ iff $\mathcal{P}_D = \mathcal{P}_C$ (i.e. $\mathcal{P}_D \subseteq \mathcal{P}_C$ and $\mathcal{P}_C \subseteq \mathcal{P}_D$) and $\mathcal{E}_D \cong \mathcal{E}_C$.*

$$\begin{aligned}
\mathsf{AspirationOfMucus} &\equiv \mathsf{AspirationSyndromes} \sqcap \exists \mathsf{agent.Mucus} \\
\mathsf{AspirationOfMilk} &\equiv \mathsf{AspirationSyndromes} \sqcap \mathsf{InhalationOfLiquid} \\
&\quad \sqcap \exists \mathsf{agent.Milk} \sqcap \exists \mathsf{assocWith.Milk} \\
\mathsf{Hypoxia} &\equiv \mathsf{RespiratoryDisorder} \sqcap \mathsf{BloodGasDisorder} \\
&\quad \sqcap \exists \mathsf{interprets.OxygenDelivery} \\
\mathsf{Hypoxemia} &\equiv \mathsf{RespiratoryDisorder} \sqcap \mathsf{BloodGasDisorder} \\
&\quad \sqcap \exists \mathsf{interprets.OxygenDelivery} \\
&\quad \sqcap \exists \mathsf{site.ArterialSystem} \\
\mathsf{AspirationSyndromes} &\sqsubseteq \mathsf{RespiratoryDisorder} \\
\mathsf{agent} &\sqsubseteq \mathsf{assocWith}
\end{aligned}$$

Fig. 1. Examples of $\mathcal{ELH}$ concept descriptions defined in $\mathcal{O}_{\mathsf{med}}$

Based on the property of concept subsumption and homomorphism mapping, in the next section, we introduce the notion of homomorphism degree hd and concept similarity sim.

3 Homomorphism Degree

Let $\mathcal{T} = (V, E, rt, \ell, \rho)$ be the $\mathcal{ELH}$ description tree as previously defined. Then, the degree of having a homomorphism from $\mathcal{T}_D$ to $\mathcal{T}_C$ is defined by Definition 2.

Definition 2 (Homomorphism degree)
Let $\mathbf{T}^{\mathcal{ELH}}$ be the set of all $\mathcal{ELH}$ description trees. The homomorphism degree function $\mathsf{hd} : \mathbf{T}^{\mathcal{ELH}} \times \mathbf{T}^{\mathcal{ELH}} \rightarrow [0, 1]$ *is inductively defined as follows:*

$$\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C) := \mu \cdot \mathsf{p\text{-}hd}(\mathcal{P}_D, \mathcal{P}_C) + (1 - \mu) \cdot \mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_D, \mathcal{E}_C), \tag{1}$$

where $|\cdot|$ *represents the set cardinality,* $\mu = \frac{|\mathcal{P}_D|}{|\mathcal{P}_D \cup \mathcal{E}_D|}$ *and* $0 \leq \mu \leq 1$*;*

$$\mathsf{p\text{-}hd}(\mathcal{P}_D, \mathcal{P}_C) := \begin{cases} 1 & \text{if } \mathcal{P}_D = \emptyset \\ \frac{|\mathcal{P}_D \cap \mathcal{P}_C|}{|\mathcal{P}_D|} & \text{otherwise,} \end{cases} \tag{2}$$

$$\mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_D, \mathcal{E}_C) := \begin{cases} 1 & \text{if } \mathcal{E}_D = \emptyset \\ \sum\limits_{\epsilon_i \in \mathcal{E}_D} \frac{max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_C\}}{|\mathcal{E}_D|} & \text{otherwise,} \end{cases} \tag{3}$$

where ϵ_i, ϵ_j are existential restrictions; and

$$\mathsf{e\text{-}hd}(\exists r.X, \exists s.Y) := \gamma(\nu + (1-\nu) \cdot \mathsf{hd}(\mathcal{T}_X, \mathcal{T}_Y)) \tag{4}$$

where $\gamma = \frac{|\mathcal{R}_r \cap \mathcal{R}_s|}{|\mathcal{R}_r|}$ *and* $0 \leq \nu < 1$.

The meaning of μ and ν are similar to those defined in our previous work [13] and are set to $\frac{|\mathcal{P}_C|}{|\mathcal{P}_C \cup \mathcal{E}_C|}$ and 0.4, respectively. However, in this work, we introduce the notion of γ which is the proportion of common roles between r and s against all those respect to r. For a special case where $\gamma = 0$, this means that there is no role commonality, therefore, further computation for all successors should be omitted. For the case that $0 < \gamma < 1$, this reveals that there exists some commonality. Moreover, if $\gamma = 1$, both r and s are totally similar and thus considered logically equivalent.

Proposition 1. *Let C, D be $\mathcal{ELH}$ concept descriptions, and $\mathcal{O}$ an $\mathcal{ELH}$ unfoldable TBox. Then, the following are equivalent:*

1. $C \sqsubseteq_{\mathcal{O}} D$
2. $\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C) = 1$,

where X is the equivalent expanded concept description w.r.t. $\mathcal{O}$, and $\mathcal{T}_X$ is its corresponding $\mathcal{ELH}$ description tree, with $X \in \{C, D\}$.

Proof. ($1 \implies 2$) To prove this, we need to show that for each $v \in V_D$, there exists $h(v) \in V_C$ such that $\mathsf{p\text{-}hd}(\cdot,\cdot) = 1$ and $\mathsf{e\text{-}set\text{-}hd}(\cdot,\cdot) = 1$ (only for those non-leaf nodes). Let d be the depth of $\mathcal{T}_D$. Since $C \sqsubseteq_{\mathcal{O}} D$, by Theorem 1 there exists a homomorphism from $\mathcal{T}_D$ to $\mathcal{T}_C$. For the induction base case where $d = 0$ and $C = P_1 \sqcap \ldots \sqcap P_m$, there exists a mapping from rt_D to rt_C such that $\ell_D(v) \subseteq \ell_C(h(v))$ (i.e. $\mathsf{hd} = \mathsf{p\text{-}hd} = 1$). For the induction step where $C = P_1 \sqcap \ldots \sqcap P_m \sqcap \exists r_1 C_1 \sqcap \ldots \sqcap \exists r_n C_n$ there exists a mapping from each v to $h(v)$ such that $\ell_D(v) \subseteq \ell_C(h(v))$ (i.e. $\mathsf{p\text{-}hd}(\cdot,\cdot) = 1$) and $\rho_D(v,w) \subseteq \rho_C(h(v),h(w))$ (i.e. $\mathsf{e\text{-}set\text{-}hd}(\cdot,\cdot) = 1$) where w and $h(w)$ are successors of v and $h(v)$, respectively. For the case where v is a leaf, this is similar to the base case (i.e. $\mathsf{p\text{-}hd}(\cdot,\cdot) = 1$).

($2 \implies 1$) By Definition 2, $\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C) = 1$ means $\mathsf{p\text{-}hd}(\mathcal{P}_D, \mathcal{P}_C) = 1$ and $\mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_D, \mathcal{E}_C) = 1$ (in case that the tree has child nodes), therefore for each $P \in \mathcal{P}_D$ there exists $P \in \mathcal{P}_C$ (i.e. $\mathcal{P}_D \subseteq \mathcal{P}_C$) and for each $\exists r.D' \in \mathcal{E}_D$ there exists $\exists s.C' \in \mathcal{E}_C$ such that $s \sqsubseteq^* r$ and $C' \sqsubseteq D'$. By Corollary 1, this implies that $C \sqsubseteq D$.

The homomorphism degree function provides a numerical value that represents structural similarity of one concept description when compared to another concept description. Since both directions constitute the degree of the two concepts being equivalent, our similarity measure for $\mathcal{ELH}$ concept descriptions is defined by means of these values. Definition 3 defines a similarity between concepts.

Definition 3. *Let C, D be $\mathcal{ELH}$ concept descriptions, and $\mathcal{O}$ an $\mathcal{ELH}$ unfoldable TBox. The* degree of similarity between C and D, *in symbols* $\mathsf{sim}(C, D)$, *is defined as:*

$$\mathsf{sim}(C, D) := \frac{\mathsf{hd}(\mathcal{T}_C, \mathcal{T}_D) + \mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C)}{2}, \tag{5}$$

where X is the equivalent expanded concept description w.r.t. $\mathcal{O}$, and $\mathcal{T}_X$ is its corresponding $\mathcal{ELH}$ description tree, with $X \in \{C, D\}$.

Example 1. To be more illustrative, consider concepts defined in $\mathcal{O}_{\mathsf{med}}$ (see Figure 1). From a classical DL reasoner's point of view, it is clear that the concept AspirationOfMilk (AMK) and AspirationOfMucus (AMC) are not in the subsumption relation, i.e. there is no relationship between the two concepts, despite the fact that they are both disorders in a group of AspirationSymdromes. Moreover, it is intuitive to argue that AspirationOfMilk is more similar to AspirationOfMucus than to Hypoxemia or to Hypoxia. Consider the expanded form of AMK and AMC.

$$\begin{aligned}\mathsf{AMK} \equiv X \ &\sqcap \ \mathsf{RespiratoryDisorder} \ \sqcap \ \mathsf{InhalationOfLiquid} \\ &\sqcap \ \exists \mathsf{agent.Milk} \ \sqcap \ \exists \mathsf{assocWith.Milk}\end{aligned}$$

$$\mathsf{AMC} \equiv X \ \sqcap \ \mathsf{RespiratoryDisorder} \ \sqcap \ \exists \mathsf{agent.Mucus}$$

where X is a fresh concept. The following shows sample computation steps for $\mathsf{hd}(\mathcal{T}_{\mathsf{AMK}}, \mathcal{T}_{\mathsf{AMC}})$:

$$\begin{aligned}\mathsf{hd}(\mathcal{T}_{\mathsf{AMK}}, \mathcal{T}_{\mathsf{AMC}}) &:= \tfrac{3}{5}\mathsf{p\text{-}hd}(\mathcal{P}_{\mathsf{AMK}}, \mathcal{P}_{\mathsf{AMC}}) + \tfrac{2}{5}\mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_{\mathsf{AMK}}, \mathcal{E}_{\mathsf{AMC}}) \\ &:= \tfrac{3}{5}(\tfrac{2}{3}) + \tfrac{2}{5} \sum_{\epsilon_i \in \mathcal{E}_{\mathsf{AMK}}} \tfrac{max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_{\mathsf{AMC}}\}}{|\mathcal{E}_{\mathsf{AMK}}|} \\ &:= \tfrac{3}{5}(\tfrac{2}{3}) + \tfrac{2}{5}(\tfrac{1}{2}) \sum_{\epsilon_i \in \mathcal{E}_{\mathsf{AMK}}} max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_{\mathsf{AMC}}\} \\ &:= \tfrac{3}{5}(\tfrac{2}{3}) + \tfrac{2}{5}(\tfrac{1}{2})(\tfrac{2}{5} + \tfrac{2}{5})\end{aligned}$$

$$//\text{Where} \sum_{\epsilon_i \in \mathcal{E}_{\mathsf{AMK}}} max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_{\mathsf{AMC}}\} = \tfrac{2}{5} + \tfrac{2}{5}; \text{ see belows}$$

$$:= 0.56$$

The computation for the sub-description corresponding to $\epsilon_i = \exists\mathsf{agent.Milk}$ and $\epsilon_j = \exists\mathsf{agent.Mucus}$ is as follows:

$$\begin{aligned}\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) &:= \gamma(\nu + (1 - \nu) \cdot \mathsf{hd}(\mathcal{T}_{\mathsf{Milk}}, \mathcal{T}_{\mathsf{Mucus}})) \\ &:= \tfrac{2}{2}(\tfrac{2}{5} + 0) := \tfrac{2}{5}\end{aligned}$$

With the sub-description $\epsilon_i = \exists\mathsf{assocWith.Milk}$ and $\epsilon_j = \exists\mathsf{agent.Mucus}$, we have

$$\begin{aligned}\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) &:= \gamma(\nu + (1 - \nu) \cdot \mathsf{hd}(\mathcal{T}_{\mathsf{Milk}}, \mathcal{T}_{\mathsf{Mucus}})) \\ &:= \tfrac{1}{1}(\tfrac{2}{5} + 0) := \tfrac{2}{5}.\end{aligned}$$

Table 2. Homomorphism degrees among concepts defined in $\mathcal{O}_{\mathsf{med}}$ where HPX, HPM and ASD stand for Hypoxia, Hypoxemia and AspirationSyndromes, respectively

hd(↓, →)	AMC	AMK	HPX	HPM	ASD
AspirationOfMucus	1.0	0.56	0.333	0.25	1.0
AspirationOfMilk	0.8	1.0	0.333	0.25	1.0
Hypoxia	0.333	0.2	1.0	0.75	0.5
Hypoxemia	0.333	0.2	1.0	1.0	0.5
AspirationSyndromes	0.667	0.4	0.333	0.25	1.0

The reverse direction can be computed by:

$$\begin{aligned}
\mathsf{hd}(\mathcal{T}_{\mathsf{AMC}}, \mathcal{T}_{\mathsf{AMK}}) &:= \tfrac{2}{3}\mathsf{p\text{-}hd}(\mathcal{P}_{\mathsf{AMC}}, \mathcal{P}_{\mathsf{AMK}}) + \tfrac{1}{3}\mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_{\mathsf{AMC}}, \mathcal{E}_{\mathsf{AMK}}) \\
&:= \tfrac{2}{3}(\tfrac{2}{2}) + \tfrac{1}{3} \sum_{\epsilon_i \in \mathcal{E}_{\mathsf{AMC}}} \frac{max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_{\mathsf{AMK}}\}}{|\mathcal{E}_{\mathsf{AMC}}|} \\
&:= \tfrac{2}{3}(\tfrac{2}{2}) + \tfrac{1}{3}(\tfrac{1}{1}) \sum_{\epsilon_i \in \mathcal{E}_{\mathsf{AMC}}} max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_{\mathsf{AMK}}\} \\
&:= \tfrac{2}{3}(\tfrac{2}{2}) + \tfrac{1}{3}(\tfrac{1}{1})(\tfrac{2}{5})
\end{aligned}$$

//Where $max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_{\mathsf{AMC}}\} = \frac{2}{5}$; see belows

$$:= 0.8$$

The computation for the sub-description corresponding to $\epsilon_i = \exists\mathsf{agent.Mucus}$ and $\epsilon_j = \exists\mathsf{agent.Milk}$ is as follows:

$$\begin{aligned}
\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) &:= \gamma(\nu + (1-\nu) \cdot \mathsf{hd}(\mathcal{T}_{\mathsf{Mucus}}, \mathcal{T}_{\mathsf{Milk}})) \\
&:= \tfrac{2}{2}(\tfrac{2}{5} + 0) := \tfrac{2}{5}.
\end{aligned}$$

With $\epsilon_i = \exists\mathsf{agent.Mucus}$ and $\epsilon_j = \exists\mathsf{assocWith.Milk}$, the computation for the sub-description is as follows;

$$\begin{aligned}
\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) &:= \gamma(\nu + (1-\nu) \cdot \mathsf{hd}(\mathcal{T}_{\mathsf{Mucus}}, \mathcal{T}_{\mathsf{Milk}})) \\
&:= \tfrac{1}{2}(\tfrac{2}{5} + 0) := \tfrac{1}{5}.
\end{aligned}$$

Table 2 and Table 3 show homomorphism degrees and similarity degrees among all concepts defined in $\mathcal{O}_{\mathsf{med}}$. It is obvious that the results we obtained are as expected.

It is to be mentioned that the similarity measure sim first introduced in [13] is quite similar to *simi* proposed by [10] since they are both recursive-based method. In fact, the meaning of the weighting parameter ν used in sim and ω in *simi* are identical and similarly defined. Likewise, the operators that represent the t-conorm, and fuzzy connector are relatively used but differently defined. However, unlike the work proposed by [10], the use of μ and the way

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

$\mathsf{sim}(\downarrow, \rightarrow)$	AMC	AMK	HPX	HPM	ASD
AspirationOfMucus	1.0	0.68	0.333	0.292	0.833
AspirationOfMilk	-	1.0	0.267	0.225	0.7
Hypoxia	-	-	1.0	0.875	0.417
Hypoxemia	-	-	-	1.0	0.375
AspirationSyndromes	-	-	-	-	1.0

it is weighted, which determines how important the primitive concepts are to be considered, is defined. The other is obviously the distinction of their inspirations. While *simi* is inspired by the Jaccard Index [5], sim is, on the other hand, motivated by the homomorphism-based structural subsumption characterization. In sim, as a pre-process, concept names are to be transformed into an $\mathcal{ELH}$ concept description tree. Taking this as an advantage, a bottom-up approach, which allows rejection of unnecessary recursive calls and reuses of solutions to subproblems, can be alternatively devised.

4 Desirable Properties for Concept Similarity Measure

This section describes desirable properties for concept similarity measure and provides corresponding mathematical proofs. At the end of the section, a comparison of satisfactory properties between sim and those significantly reported in other classical works is made available.

Definition 4 determines important properties for concept similarity measure introduced by [10]. These are believed to be desirable features and thus checked for satisfaction. Theorem 2 states the characteristics of sim.

Definition 4. *Let C, D and E be $\mathcal{ELH}$ concept, the similarity measure is*

i. symmetric *iff* $\mathsf{sim}(C, D) = \mathsf{sim}(D, C)$,
ii. equivalence closed *iff* $\mathsf{sim}(C, D) = 1 \Longleftrightarrow C \equiv D$,
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 *Let C_i and C_j be atoms in C where $C_i \not\sqsubseteq C_j$, the concept $D' := \bigsqcap_{i \leq n} C_i \sqcap D$ and $E' := \bigsqcap_{i \leq n} C_i \sqcap E$ satisfies the condition* $\lim_{n \to \infty} \mathsf{sim}(D', E') = 1$,
vii. *satisfying* triangle inequality *iff* $1 + \mathsf{sim}(D, E) \geq \mathsf{sim}(D, C) + \mathsf{sim}(C, E)$.

Theorem 2. *The similarity-measure* sim *is:*

i. symmetric,
ii. equivalence closed,

iii. equivalence invariant,
iv. subsumption preserving,
v. structurally dependent,
vi. not reverse subsumption preserving, and
vii. not satisfying triangle inequality.

Proof. i. By Definition 3, it is obvious that $\mathsf{sim}(C, D) = \mathsf{sim}(D, C)$.

ii. ($\Longrightarrow$) By Definition 2, $\mathsf{sim}(C, D) = 1$ iff $\mathsf{hd}(\mathcal{T}_C, \mathcal{T}_D) = 1$ and $\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C) = 1$. By Proposition 1, these imply that $C \sqsubseteq D$ and $D \sqsubseteq C$. Therefore, $C \equiv D$. ($\Longleftarrow$) Assume $C \equiv D$, then $C \sqsubseteq D$ and $D \sqsubseteq C$. Using the same proposition, this ensures that $\mathsf{hd}(\mathcal{T}_C, \mathcal{T}_D) = 1$, and $\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C) = 1$, which means $\mathsf{sim}(C, D) = 1$.

iii. $C \equiv D$ iff $C \sqsubseteq D$ and $D \sqsubseteq C$. By using Corollary 2, we have $\mathcal{P}_C = \mathcal{P}_D$ and $\mathcal{E}_C \cong \mathcal{E}_D$. Therefore, $\mathcal{T}_C = \mathcal{T}_D$ and this implies $\mathsf{hd}(\mathcal{T}_C, \mathcal{T}_E) = \mathsf{hd}(\mathcal{T}_D, \mathcal{T}_E)$ and $\mathsf{hd}(\mathcal{T}_E, \mathcal{T}_C) = \mathsf{hd}(\mathcal{T}_E, \mathcal{T}_D)$. Such that $\mathsf{sim}(C, E) = \mathsf{sim}(D, E)$.

iv. We need to show that

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

Since $C \sqsubseteq D$ and $D \sqsubseteq E$, then $C \sqsubseteq E$. By Proposition 1, $\mathsf{hd}(\mathcal{T}_E, \mathcal{T}_C) = 1$ and $\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_C) = 1$. Therefore, it suffices to show that

$$\mathsf{hd}(\mathcal{T}_C, \mathcal{T}_D) \geq \mathsf{hd}(\mathcal{T}_C, \mathcal{T}_E)$$

If expanded, on both sizes of the upper equation, we have $\mu = \frac{|\mathcal{P}_C|}{|\mathcal{P}_C \cup \mathcal{E}_C|}$. Hence, it is adequate to show that $\mathsf{p\text{-}hd}(\mathcal{P}_C, \mathcal{P}_D) \geq \mathsf{p\text{-}hd}(\mathcal{P}_C, \mathcal{P}_E)$ and $\mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_C, \mathcal{E}_D) \geq \mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_C, \mathcal{E}_E)$. For the first part, we show that

$$\frac{|\mathcal{P}_C \cap \mathcal{P}_D|}{|\mathcal{P}_C|} \geq \frac{|\mathcal{P}_C \cap \mathcal{P}_E|}{|\mathcal{P}_C|} \tag{6}$$

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

By Corollary 1, $C \sqsubseteq D \sqsubseteq E$ ensures that $\mathcal{P}_E \subseteq \mathcal{P}_D \subseteq \mathcal{P}_C$. Therefore

$$|\ \mathcal{P}_D\ | \geq |\ \mathcal{P}_E\ |$$

and Equation 6 is true. For the second part, we show that

$$\sum_{\epsilon_i \in \mathcal{E}_C} \frac{max\{\mathsf{e\text{-}hd}(\epsilon_i,\epsilon_j):\epsilon_j \in \mathcal{E}_D\}}{|\mathcal{E}_C|} \geq \sum_{\epsilon_i \in \mathcal{E}_C} \frac{max\{\mathsf{e\text{-}hd}(\epsilon_i,\epsilon_j):\epsilon_j \in \mathcal{E}_E\}}{|\mathcal{E}_C|} \tag{7}$$

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

Let $\hat{\epsilon}_i \in \mathcal{E}_E$ such that $\mathsf{e\text{-}hd}(\epsilon_i, \hat{\epsilon}_i) = max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_E\}$, but since $\hat{\epsilon}_i \in \mathcal{E}_E \subseteq \mathcal{E}_D$, then $max\{\mathsf{e\text{-}hd}(\epsilon_i, \epsilon_j) : \epsilon_j \in \mathcal{E}_D\} \geq \mathsf{e\text{-}hd}(\epsilon_i, \hat{\epsilon}_i)$. Therefore, Equation 7 is true.

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

(a) if $n_{\mathcal{P}} \to \infty$ and $n_{\mathcal{E}}$ is finite, it suffices to show that $\lim_{n_{\mathcal{P}} \to \infty} \mu = 1$ and $\lim_{n_{\mathcal{P}} \to \infty} \mathsf{p\text{-}hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) = 1$. Therefore, $\mathsf{hd}(D', E') = \mathsf{hd}(E', D') = 1$ and these imply that $\mathsf{sim}(D', E') = 1$. From Equation 2, we have

$$\begin{aligned} \mu &= \frac{|\mathcal{P}_{D'}|}{|\mathcal{P}_{D'} \cup \mathcal{E}_{D'}|} \\ &= \frac{|\mathcal{P}_{D'}|}{|\mathcal{P}_{D'}| + |\mathcal{E}_{D'}|} \\ &= \frac{|\mathcal{P}_C| + |\mathcal{P}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + |\mathcal{E}_{D'}|} \\ &= \frac{n_{\mathcal{P}} + |\mathcal{P}_D|}{n_{\mathcal{P}} + |\mathcal{P}_D| + |\mathcal{E}_{D'}|} \end{aligned} \tag{8}$$

Since $\mid \mathcal{P}_D \mid$ and $\mid \mathcal{E}_{D'} \mid$ are constant, $\lim_{n_{\mathcal{P}} \to \infty} \mu = \lim_{n_{\mathcal{P}} \to \infty} \frac{n_{\mathcal{P}} + |\mathcal{P}_D|}{n_{\mathcal{P}} + |\mathcal{P}_D| + |\mathcal{E}_{D'}|} = 1$. For the second part, we have

$$\begin{aligned} \mathsf{p\text{-}hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) &= \frac{|\mathcal{P}_{D'} \cap \mathcal{P}_{E'}|}{|\mathcal{P}_{D'}|} \\ &= \frac{|\mathcal{P}_C| + |\mathcal{P}_D \cap \mathcal{P}_E|}{|\mathcal{P}_C| + |\mathcal{P}_D|} \\ &= \frac{n_{\mathcal{P}} + |\mathcal{P}_D \cap \mathcal{P}_E|}{n_{\mathcal{P}} + |\mathcal{P}_D|} \end{aligned}$$

where $\mid \mathcal{P}_D \cap \mathcal{P}_E \mid$ and $\mid \mathcal{P}_D \mid$ are constant. Thus,

$$\lim_{n_{\mathcal{P}} \to \infty} \mathsf{p\text{-}hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) = \lim_{n_{\mathcal{P}} \to \infty} \frac{n_{\mathcal{P}} + \mid \mathcal{P}_D \cap \mathcal{P}_E \mid}{n_{\mathcal{P}} + \mid \mathcal{P}_D \mid} = 1. \tag{9}$$

(b) if $n_{\mathcal{E}} \to \infty$ and $n_{\mathcal{P}}$ is finite, it suffices to show that $\lim_{n_{\mathcal{E}} \to \infty} \mu = 0$ and $\lim_{n_{\mathcal{E}} \to \infty} \mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) = 1$ which implies $\mathsf{hd}(D', E') = \mathsf{hd}(E', D') = 1$, and $\mathsf{sim}(D', E') = 1$. From Equation 8, the value of μ is as follows:

$$\begin{aligned} \mu &= \frac{|\mathcal{P}_C| + |\mathcal{P}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + |\mathcal{E}_{D'}|} \\ &= \frac{|\mathcal{P}_C| + |\mathcal{P}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + n_{\mathcal{E}} + |\mathcal{E}_D|} \end{aligned}$$

Since $\mid \mathcal{P}_C \mid$, $\mid \mathcal{P}_D \mid$ and $\mid \mathcal{E}_D \mid$ are constant, by taking limit, we have

$$\lim_{n_{\mathcal{E}} \to \infty} \mu = \lim_{n_{\mathcal{E}} \to \infty} \frac{|\mathcal{P}_C| + |\mathcal{P}_D|}{|\mathcal{P}_C| + |\mathcal{P}_D| + n_{\mathcal{E}} + |\mathcal{E}_D|} = 0.$$

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

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

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,

$$\begin{aligned}
\mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) &= \frac{n_{\mathcal{E}} + p}{|\mathcal{E}_C| + |\mathcal{E}_D|} \\
&= \frac{n_{\mathcal{E}} + p}{n_{\mathcal{E}} + |\mathcal{E}_D|}
\end{aligned}$$

where $p = \sum_{e_i \in \mathcal{E}_D} \max\{\mathsf{e\text{-}hd}(e_i, e_j) : e_j \in \mathcal{E}_{E'}\}$, and $p \leq \mid \mathcal{E}_D \mid$. Therefore, the following is true.

$$\lim_{n_{\mathcal{E}} \to \infty} \mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) = \lim_{n_{\mathcal{E}} \to \infty} \frac{n_{\mathcal{E}} + p}{n_{\mathcal{E}} + \mid \mathcal{E}_D \mid} = 1. \qquad (10)$$

(c) if $n_{\mathcal{P}} \to \infty$ and $n_{\mathcal{E}} \to \infty$, it suffices to show that $\lim_{n_{\mathcal{P}} \to \infty} \mathsf{p\text{-}hd}(\mathcal{P}_{D'}, \mathcal{P}_{E'}) = 1$ and $\lim_{n_{\mathcal{E}} \to \infty} \mathsf{e\text{-}set\text{-}hd}(\mathcal{E}_{D'}, \mathcal{E}_{E'}) = 1$. But these follow from Equation 9 and Equation 10.

vi. Consider a counter example defined in Figure 2. It is obvious that $C \sqsubseteq D \sqsubseteq E$. By definition,

$$\begin{aligned}
\mathsf{sim}(C, E) &:= \frac{\mathsf{hd}(\mathcal{T}_C, \mathcal{T}_E) + \mathsf{hd}(\mathcal{T}_E, \mathcal{T}_C)}{2} \\
&:= \frac{0.4250 + 1}{2} \\
&:= 0.7125
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{sim}(D, E) &:= \frac{\mathsf{hd}(\mathcal{T}_D, \mathcal{T}_E) + \mathsf{hd}(\mathcal{T}_E, \mathcal{T}_D)}{2} \\
&:= \frac{0.3333 + 1}{2} \\
&:= 0.6667.
\end{aligned}$$

Apparently, there exists the case $\mathsf{sim}(C, E) \not\leq \mathsf{sim}(D, E)$.

$$\begin{array}{lcl} \mathsf{E} & \equiv & \exists \mathsf{r}.(\mathsf{F} \sqcap \mathsf{G}) \\ \mathsf{D} & \equiv & \exists \mathsf{r}.(\mathsf{F} \sqcap \mathsf{G}) \sqcap \exists s.\mathsf{F} \sqcap \exists s.\mathsf{G} \\ \mathsf{C} & \equiv & \exists \mathsf{r}.(\mathsf{F} \sqcap \mathsf{G}) \sqcap \exists s.\mathsf{F} \sqcap \exists s.\mathsf{G} \sqcap \exists \mathsf{r}.(\mathsf{F} \sqcap \mathsf{H}) \end{array}$$

Fig. 2. Examples of $\mathcal{ELH}$ concept descriptions

vii. Providing the concept description C, D, and E defined in Figure 2, the following demonstrates the case $1 + \mathsf{sim}(D,E) \not\geq \mathsf{sim}(D,C) + \mathsf{sim}(C,E)$. Here, we have

$$\begin{aligned} \mathsf{sim}(D,E) &:= \tfrac{\mathsf{hd}(\mathcal{T}_D,\mathcal{T}_E)+\mathsf{hd}(\mathcal{T}_E,\mathcal{T}_D)}{2} := \tfrac{0.3333+1}{2} \\ &:= 0.6667 \end{aligned}$$

and

$$\begin{aligned} \mathsf{sim}(D,C) &:= \tfrac{\mathsf{hd}(\mathcal{T}_D,\mathcal{T}_C)+\mathsf{hd}(\mathcal{T}_C,\mathcal{T}_D)}{2} := \tfrac{1+0.9250}{2} \\ &:= 0.9625 \end{aligned}$$

and

$$\begin{aligned} \mathsf{sim}(C,E) &:= \tfrac{\mathsf{hd}(\mathcal{T}_C,\mathcal{T}_E)+\mathsf{hd}(\mathcal{T}_E,\mathcal{T}_C)}{2} := \tfrac{0.4250+1}{2} \\ &:= 0.7125. \end{aligned}$$

By applying a summation, it is obvious that $1.6667 \not\geq 1.675$.

Table 4. A comparison on concept-similarity properties

Similarity Measure	DL	symmetric	equi. closed	equi invariant	sub. preserving	struc. dependent	rev. sub. preserving	triangle inequality
sim	$\mathcal{ELH}$	✓	✓	✓	✓	✓		
Lehmann and Turhan [10]	$\mathcal{ELH}$	✓	✓	✓	✓	✓		
Jaccard [5]	$\mathcal{L}_0$	✓	✓	✓	✓	✓	✓	✓
Janowicz and Wilkes [7]	$\mathcal{SHI}$	✓						✓
Janowicz [6]	$\mathcal{ALCHQ}$	✓						✓
d'Amato et al. [2]	$\mathcal{ALC}$							
Fanizzi and d'Amato [4]	$\mathcal{ALN}$	✓		✓		✓	✓	
d'Amato et al. [1]	$\mathcal{ALC}$	✓		✓		✓	✓	
d'Amato et al. [3]	$\mathcal{ALE}$	✓		✓		✓	✓	

To ensure that our proposed method reaches the performance, Table 4 compares desirable properties of sim and those previously reported in other classical works. Except than the work proposed by [5], which allows only concept conjunction, our approach and that proposed by [10] apparently hold significant features.

5 Conclusion

To this end, we have expanded a concept similarity measure for $\mathcal{EL}$ to take into account also role hierarchy. Comparing to other related works, the measure has been proved that it is outperforming and indeed identical to *simi* in terms of satisfaction of desirable properties.

Particularly, the proposed algorithm is inspired by the homomorphism-based structural subsumption characterization. With the top-down approach, a similarity degree is recursively computed, and as a nature of recursion, there is a chance that the number of unnecessary recursive calls will be greatly increase. Fortunately, as being computed based on description trees, an optimized version of the algorithm that allows rejection of needless computation can be alternatively devised in a reversed direction and this is regarded as one target in our future works. The other directions of possible future works are an extension of the algorithm to a general TBox (i.e. a handling to concepts with cyclic definition) and to a more expressive DL. Lastly, we also aim at setting up experiments on comprehensive terminologies (e.g. SNOMED CT [12] and Gene Ontology [11]) and making a comparison among results obtained from different methods.

Acknowledgments. This work is partially supported by the National Research University (NRU) project of Thailand Office for Higher Education Commission; and by Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation (CILS), Thammasat University.

References

1. d'Amato, C., Fanizzi, N., Esposito, F.: A semantic similarity measure for expressive description logics. In: Proceedings of Convegno Italiano di Logica Computazionale, CILC 2005 (2005)
2. d'Amato, C., Fanizzi, N., Esposito, F.: A dissimilarity measure for $\mathcal{ALC}$ concept descriptions. In: Proceedings of the 2006 ACM Symposium on Applied Computing, SAC 2006, pp. 1695–1699. ACM, New York (2006). http://doi.acm.org/10.1145/1141277.1141677
3. d'Amato, C., Staab, S., Fanizzi, N.: On the influence of description logics ontologies on conceptual similarity. In: Gangemi, A., Euzenat, J. (eds.) EKAW 2008. LNCS (LNAI), vol. 5268, pp. 48–63. Springer, Heidelberg (2008). http://dx.doi.org/10.1007/978-3-540-87696-0_7
4. Fanizzi, N., d'Amato, C.: A similarity measure for the aln description logic. Proceedings of Convegno Italiano di Logica Computazionale, CILC 2006, pp. 26–27 (2006)

5. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles **37**, 547–579 (1901)
6. Janowicz, K.: Sim-DL: towards a semantic similarity measurement theory for the description logic $\mathcal{ALCNR}$ in geographic information retrieval. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4278, pp. 1681–1692. Springer, Heidelberg (2006). http://dx.doi.org/10.1007/11915072_74
7. Janowicz, K., Wilkes, M.: SIM-DL$_A$: a novel semantic similarity measure for description logics reducing inter-concept to inter-instance similarity. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) ESWC 2009. LNCS, vol. 5554, pp. 353–367. Springer, Heidelberg (2009). http://dx.doi.org/10.1007/978-3-642-02121-3_28
8. Jirathitikul, P., Nithisansawadikul, S., Tongphu, S., Suntisrivaraporn, B.: A similarity measuring service for snomed ct: structural analysis of concepts in ontology. In: 2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1–6, May 2014
9. Lehmann, J., Haase, C.: Ideal downward refinement in the $\mathcal{EL}$ description logic. Technical report, University of Leipzig (2009). http://jens-lehmann.org/files/2009_ideal_operator_el_tr.pdf
10. Lehmann, K., Turhan, A.-Y.: A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts. In: del Cerro, L.F., Herzig, A., Mengin, J. (eds.) JELIA 2012. LNCS, vol. 7519, pp. 307–319. Springer, Heidelberg (2012)
11. Michael Ashburner, C.A.: Creating the Gene Ontology Resource: Design and Implementation. Genome Research **11**(8), 1425–1433 (2001). http://dx.doi.org/10.1101/gr.180801
12. Schulz, S., Suntisrivaraporn, B., Baader, F., Boeker, M.: SNOMED reaching its adolescence: Ontologists' and logicians' health check. International Journal of Medical Informatics **78**(Supplement 1), S86–S94 (2009)
13. Suntisrivaraporn, B.: A similarity measure for the description logic $\mathcal{EL}$ with unfoldable terminologies. In: International Conference on Intelligent Networking and Collaborative Systems (INCoS-13). pp. 408–413 (2013)