Identifying an Agent's Preferences Toward Similarity Measures in Description Logics

Teeradaj Racharak[1,2], Boontawee Suntisrivaraporn[1], and Satoshi Tojo[2]

[1] School of Information, Computer and Communication Technology,
Sirinthorn International Institute of Technology,
Thammasart University, Thailand
r.teeradaj@gmail.com, sun@siit.tu.ac.th
[2] Graduate School of Information Science,
Japan Advanced Institute of Science and Technology, Japan
racharak@jaist.ac.jp, tojo@jaist.ac.jp

Abstract. In Description Logics (DLs), concept similarity measures (CSMs) aim at identifying a degree of commonality between two given concepts and are often regarded as a generalization of the classical reasoning problem of equivalence. That is, any two concepts are equivalent if their similarity degree is one, and vice versa. When two concepts are not quite equivalent but similar, nevertheless, a problem may arise as to which aspects of commonality should play more important role than others. This work presents the so-called *preference profile*, which is design guidelines for an agent's preferences and points out to our preliminary developing stage of sim^{π} [1], in which an agent's preferences can influence the calculation of CSM in DL $\mathcal{ELH}$.

Key words: Preference Profile, Concept Similarity Measures, Non-standard Reasoning Services, Description Logics

1 Introduction and Motivation

Preferences are used in a variety of related, but not identical, ways in human beings' daily life: to express what they like and dislike, to express their desired goals when choosing routes for travelling [2], etc.. In psychology, preferences may be conceived of as an individual's attitude towards a set of objects when making decisions [3]. Alternatively, it can be interpreted as a judgment in a sense of liking or disliking an object [4].

In Description Logics (DLs), concept similarity measures (CSMs) aim at identifying a degree of commonality between two given concept names and are often regarded as a generalization of the classical reasoning problem of equivalence. That is, any two concepts are equivalent if their similarity degree is one, and vice versa. To date, many elegant CSMs have been developed (cf. Subsection 2.2). These developments can induce efficient similarity-oriented DL reasoning services, i.e., to measure if two concepts are similar, to check if an individual is a relaxed instance of a concept, and to retrieve those individuals similar to a given individual.

Unfortunately, those similarity measures may be counter-intuitive when it comes to human perception. A counterexample could be the similarity between animals in accordance with human beings' preferences. A person may perceive that Frog is similar to $\mathsf{TreeLizard}$ because they belong to the same class and their skin colors are similar. On the other hand, another person may rather perceive that Frog is similar to $\mathsf{TreeLizard}$ because they belong to the same class and their natural habitation and current living environment are near. These scenarios reveal that human beings always have *bias* or *preferences* when making judgments about concepts in question.

Example 1. Aforementioned concepts could be modeled in DL as follows:

$$\begin{aligned}
\mathsf{Frog} &\sqsubseteq \mathsf{Reptile} \sqcap \exists \mathsf{hasColor.Green} \sqcap \exists \mathsf{hasHabitat.Forest} \\
\mathsf{TreeLizard} &\sqsubseteq \mathsf{Reptile} \sqcap \exists \mathsf{hasColor.Yellow} \sqcap \exists \mathsf{liveIn.Forest} \\
\mathsf{hasHabitat} &\sqsubseteq \mathsf{hasLocation} \\
\mathsf{liveIn} &\sqsubseteq \mathsf{hasLocation}
\end{aligned}$$

Reasonable perception when considering on the DLs is that Frog and $\mathsf{TreeLizard}$ are not much similar. However, in reality, the similarity value between each of them could be varied, by an agent's preferences as stated above. □

In this paper, we study and formulate essential aspects of preferences which can be expressed by an agent when measuring the similarity between two concept descriptions. A numerical degree yielded from a computation of concept similarity measures under an agent's preferences will thereby comply more with the agent's intuition than that from the base measures do.

The structure of this paper is organized as follows. Section 2 reviews Description Logics, particularly $\mathcal{ELH}$, and the definition of concept similarity measures (CSMs). Section 3 presents potential aspects of preference expressions formalized as *guidelines for developing concept similarity measures under preferences* in DLs and discusses that, to the best of our knowledge, none of existing measures for DLs satisfy all elements of the so-called *preference profile*. Note here that all the proposed preference expressions are each represented as functions and are considered collectively as the preference profile. Finally, section 4 presents the conclusion and our future work.

2 Preliminaries

2.1 Description Logics

In Description Logics (DLs), *concept descriptions* are inductively defined by the help of a set of *constructors*, a set of concept names CN, and a set of role names RN. $\mathcal{ELH}$ concept descriptions are formed using the conjunction ($\sqcap$), existential restrictions (e.g., ($\exists r.C$) where $r \in \mathsf{RN}$ and $C \in \mathsf{CN}$), and the top concept ($\top$). The set of concept descriptions, or simply concepts, for a specific DL $\mathcal{L}$ is denoted by $\mathsf{Con}(\mathcal{L})$. For instance, $\mathsf{Con}(\mathcal{ELH})$ is the set of all $\mathcal{ELH}$ concept descriptions.

Conventionally, concept names are denoted by A and B, concept descriptions are denoted by C and D, and role names are denoted by r and s.

A terminology or TBox $\mathcal{T}$ is a finite set of concept definitions (e.g., $A \sqsubseteq D$ or $A \equiv D$ where $A, D \in \mathsf{CN}$) and role hierarchy axioms (e.g., $r \sqsubseteq s$ where $r, s \in \mathsf{RN}$). A TBox is called *unfoldable* if it contains at most one concept definition for each concept name in CN and does not contain cyclic dependencies. Concept names occurring on the left-hand side of a concept definition are called defined concept names (denoted by $\mathsf{CN}^{\mathsf{def}}$), the other concept names are primitive concept names (denoted by $\mathsf{CN}^{\mathsf{pri}}$). Primitive concept definitions are commonly found in realistic terminologies in which necessary conditions of concepts are merely known. Such a primitive definition $A \sqsubseteq D$ can easily be transformed into a semantically equivalent full definitions $A \equiv X \sqcap D$ where X is a fresh concept name. When a TBox $\mathcal{T}$ is unfoldable, concept names can be expanded by exhaustively replacing all defined concept names by their definitions until only primitive concept names remain. Such concept names are called *fully expanded concept names*. In this work, we assume that concepts are fully expanded since TBox can be completely disregarded from decision procedures. Furthermore, a set of statements about the characteristics of roles can be axiomatized by a role hierarchy. Like primitive definitions, a role hierarchy axiom $r \sqsubseteq s$ can be transformed in to a semantically equivalent role definition $r \equiv t \sqcap s$ where t is a fresh role name. Role names occurring on the left-hand side of a role definition are called defined role names (denoted by $\mathsf{RN}^{\mathsf{def}}$).

In order to defined a formal semantics for a specific DL $\mathcal{L}$, we consider an *interpretation* $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, which consists of a nonempty set $\Delta^{\mathcal{I}}$ as the domain of the interpretation and an interpretation function $\cdot^{\mathcal{I}}$ which assigns to every concept name A a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to every role name r a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ (cf. [5, 6] for more details). An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if, for each axiom in $\mathcal{T}$, the conditions corresponding to their semantics are satisfied. One of the main classical reasoning problems is the *subsumption problem*. That is, given two concept descriptions C and D and a TBox $\mathcal{T}$, C is subsumed by D w.r.t. a TBox $\mathcal{T}$ (written as $C \sqsubseteq_{\mathcal{T}} D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ in every model $\mathcal{I}$ of $\mathcal{T}$. Furthermore, C and D are equivalent w.r.t. $\mathcal{T}$ (written as $C \equiv_{\mathcal{T}} D$) if $C \sqsubseteq_{\mathcal{T}} D$ and $D \sqsubseteq_{\mathcal{T}} C$. When a TBox $\mathcal{T}$ is empty or is clear from the context, we omit to denote $\mathcal{T}$, i.e. $C \sqsubseteq D$ and $C \equiv D$.

2.2 Concept Similarity Measure

Concept similarity measure (CSM) is one of non-standard DL reasoning services. It determines how similar two concepts are. Formally, given $C, D \in \mathsf{Con}(\mathcal{L})$ be two concept descriptions for a specific DL $\mathcal{L}$. Then, a *concept similarity measure* w.r.t. a TBox $\mathcal{T}$ is a function $\sim_{\mathcal{T}} : \mathsf{Con}(\mathcal{L}) \times \mathsf{Con}(\mathcal{L}) \to [0, 1]$ such that $C \sim_{\mathcal{T}} D = 1$ iff $C \equiv_{\mathcal{T}} D$ (total similarity) and $C \sim_{\mathcal{T}} D = 0$ indicates total dissimilarity between C and D. When a TBox $\mathcal{T}$ is empty or is clear from the context, we simply write $C \sim D$.

There exist many state-of-the-art measures and those can be seen as actual instances of CSMs. For instance, two elegant measures, viz. $\sim^s$ and $\sim^c$, based on an automata-theoretic characterization of subsumption in $\mathcal{FL}_0$ are defined in [7] to calculate the similarity degree between two $\mathcal{FL}_0$ concept descriptions w.r.t. different levels of strongness. The measure sim from [5] for $\mathcal{EL}$ concept descriptions is defined based on a characterization of subsumption by tree homomorphism. The work from [5] is continued to define the similarity-based instance checking [8] and to define for measuring the similarity between $\mathcal{ELH}$ concept descriptions [9, 6] together with two concrete algorithms for implementing the proposed measure. Another measure for DL $\mathcal{ELH}$ is the parameterizable measure called *simi* [10] which allows calibrating via various parameters of the measure to fit the expectation. A set of well-defined properties for CSMs is also collected and introduced in [10]. Those formally defined properties are believed to desirable properties for CSMs, i.e., actual instances of CSMs complying with those properties can produce predictable outcomes for CSM users. Fortunately, sim [9, 6] and *simi* [10] are theoretically proven to fulfill most of those properties. To illustrate an application of CSMs, applying sim [9, 6] on concepts given in Example 1 yields $\mathsf{sim}(\mathsf{Frog}, \mathsf{TreeLizard}) = 0.475$ (See more details in [1]).

3 Preference Profile

A numerical degree value obtained by CSMs indicates the similarity degree value between two concept descriptions. For instance, $\mathsf{sim}(\mathsf{Frog}, \mathsf{TreeLizard}) = 0.475$ indicates that the similarity between Frog and TreeLizard is 47.5%. Unfortunately, the finding reported by CSMs might not be intuitive and reasonable concerning different perceiving agents. Consider two aforementioned agents from Section 1:

Agent 1: Frog is similar to TreeLizard due to classes and skin color;
Agent 2: Frog is similar to TreeLizard due to classes and surrounding.

Most modern CSMs including sim reveal Frog and TreeLizard are not quite similar. Hence, they per se are not appropriate to be used in our scenario. In this section, we explore various aspects of preference expressions which can be seamlessly captured in CSMs and then formalize them as guidelines for *CSMs under preferences*. These aspects of preferences are compiled together as the *preference profile* π. Any CSMs which expose a syntax and satisfy semantics of these aspects are appropriate to be used under an agent's preferences.

In the following, we present five aspects of preference expressions, which can be adopted in and thereby influencing the calculation of CSMs. The syntax and semantics for each aspect of a preference profile are given in term of partial functions since the exact domain of each aspect is varied from agents to agents.

1. Primitive concept importance;
2. Role importance;
3. Primitive concepts similarity;
4. Primitive roles similarity; and
5. Role discount factor.

Definition 1 (Primitive Concept Importance). *Let* $\mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ *be a set of primitive concept names occurring in* $\mathcal{T}$. *Then, a* primitive concept importance *is a partial function* $\mathfrak{i}^{\mathfrak{c}} : \mathsf{CN} \to \mathbb{R}_{\geq 0}$, *where* $\mathsf{CN} \subseteq \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$.

For any $A \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$, $\mathfrak{i}^{\mathfrak{c}}(A) = 1$ captures an expression of normal importance for A, $\mathfrak{i}^{\mathfrak{c}}(A) > 1$ (and $\mathfrak{i}^{\mathfrak{c}}(A) < 1$) indicates that A has higher (and lower, respectively) importance, and $\mathfrak{i}^{\mathfrak{c}}(A) = 0$ indicates that A is entirely ignored by an agent. For example, suppose both agents consider heavily whether the two concepts are in the same class, i.e., $\mathsf{Reptile}$. Therefore, they might express as $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Reptile}) = 2$ for their own preference profiles.

Practically, many primitive concept names might be not assigned the important values by agents, i.e. those concept names are not mapped to the corresponding values. Hence, we assign the default importance value of 1 for $A \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ in case $\mathfrak{i}^{\mathfrak{c}}(A)$ is not defined. Furthermore, the total function $\mathfrak{i}_0^{\mathfrak{c}}(A) = 1$ for all $A \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ is called the *default primitive concept importance.*

Definition 2 (Role Importance). *Let* $\mathsf{RN}(\mathcal{T})$ *be a set of role names occurring in* $\mathcal{T}$*. Then, a* role importance *is a partial function* $\mathfrak{i}^{\mathfrak{r}} : \mathsf{RN} \to \mathbb{R}_{\geq 0}$*, where* $\mathsf{RN} \subseteq \mathsf{RN}(\mathcal{T})$*.*

For any $r \in \mathsf{RN}(\mathcal{T})$, $\mathfrak{i}^{\mathfrak{r}}(r) = 1$ captures an expression of normal importance for r, $\mathfrak{i}^{\mathfrak{r}}(r) > 1$ (and $\mathfrak{i}^{\mathfrak{r}}(r) < 1$) indicates that r has higher (and lower, respectively) importance, and $\mathfrak{i}^{\mathfrak{r}}(r) = 0$ indicates that r is entirely ignored by an agent. For example, **Agent 1** may consider heavily on their skin colors, i.e., $\mathfrak{i}^{\mathfrak{r}}(\mathsf{hasColor}) = 2$. In addition, **Agent 2** may consider heavily on their surrounding, i.e., $\mathfrak{i}^{\mathfrak{r}}(\mathsf{hasHabitat}) = 2$ and $\mathfrak{i}^{\mathfrak{r}}(\mathsf{liveIn}) = 2$.

Practically, many role names might be not assigned the important values by agents, i.e. those role names are not mapped to the corresponding values. Hence, we use the default importance value of 1 for $r \in \mathsf{RN}(\mathcal{T})$ in case $\mathfrak{i}^{\mathfrak{r}}(r)$ is not defined. Furthermore, the total function $\mathfrak{i}_0^{\mathfrak{r}}(r) = 1$ for all $r \in \mathsf{RN}(\mathcal{T})$ is called the *default role importance.*

Definition 3 (Primitive Concepts Similarity). *Let* $\mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ *be a set of primitive concept names occurring in* $\mathcal{T}$*. For* $A, B \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$*, a* primitive concepts similarity *is a partial function* $\mathfrak{s}^{\mathfrak{c}} : \mathsf{CN} \times \mathsf{CN} \to [0, 1]$*, where* $\mathsf{CN} \subseteq \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$*, such that* $\mathfrak{s}^{\mathfrak{c}}(A, B) = \mathfrak{s}^{\mathfrak{c}}(B, A)$ *and* $\mathfrak{s}^{\mathfrak{c}}(A, A) = 1$*.*

For $A, B \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$, $\mathfrak{s}^{\mathfrak{c}}(A, B) = 1$ captures an expression of total similarity between A and B and $\mathfrak{s}^{\mathfrak{c}}(A, B) = 0$ captures an expression of total dissimilarity between A and B. For example, **Agent 1** may feel that there is similarity between Green and Yellow. Hence, $\mathfrak{s}^{\mathfrak{c}}(\mathsf{Green}, \mathsf{Yellow}) = 0.5$.

Practically, many pairs of primitive concept names might be not assigned the similarity values by agents, i.e. those pairs of concept names are not mapped to the corresponding values. Hence, we assign the default similarity value of 0 for $(A, B) \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}) \times \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ in case $\mathfrak{s}^{\mathfrak{c}}(A, B)$ is not defined. Furthermore, the total function $\mathfrak{s}_0^{\mathfrak{c}}(A, B) = 0$ for all $(A, B) \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}) \times \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ is called the *default primitive concept similarity.*

Definition 4 (Primitive Roles Similarity). *Let* $\mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$ *be a set of primitive role names occurring in* $\mathcal{T}$*. For* $r, s \in \mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$*, a* primitive roles similarity *is a partial function* $\mathfrak{s}^{\mathfrak{r}} : \mathsf{RN} \times \mathsf{RN} \to [0, 1]$*, where* $\mathsf{RN} \subseteq \mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$*, such that* $\mathfrak{s}^{\mathfrak{r}}(r, s) = \mathfrak{s}^{\mathfrak{r}}(s, r)$ *and* $\mathfrak{s}^{\mathfrak{r}}(r, r) = 1$*.*

For $r, s \in \mathsf{RN}(\mathcal{T})$, $\mathfrak{s}^{\mathfrak{r}}(r,s) = 1$ captures an expression of total similarity between r and s and $\mathfrak{s}^{\mathfrak{r}}(r,s) = 0$ captures an expression of total dissimilarity between r and s. For example, **Agent 2** may feel that there is similarity between knowing the natural habitat of the animals, i.e., hasHabitat, and knowing the living environment of the animals, i.e., liveIn. Hence, expressing the similarity between their corresponding new primitive role names can be exploited, i.e., $\mathfrak{s}^{\mathfrak{r}}(t,u) = 0.1$.

Practically, many pairs of primitive role names might be not assigned the similarity values by agents, i.e. those pairs of role names are not mapped to the corresponding values. Hence, we assign the default similarity value of 0 for $(r,s) \in \mathsf{RN}^{\mathsf{pri}}(\mathcal{T}) \times \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ in case $\mathfrak{s}^{\mathfrak{r}}(r,s)$ is not defined. Furthermore, the total function $\mathfrak{s}^{\mathfrak{r}}_0(r,s) = 0$ for all $(r,s) \in \mathsf{RN}^{\mathsf{pri}}(\mathcal{T}) \times \mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$ is called the *default primitive role similarity.*

Definition 5 (Role Discount Factor). *Let* $\mathsf{RN}(\mathcal{T})$ *be a set of role names occurring in* $\mathcal{T}$*. Then, a* role discount factor *is a partial function* $\mathfrak{d} : \mathsf{RN} \to [0,1]$*, where* $\mathsf{RN} \subseteq \mathsf{RN}(\mathcal{T})$*.*

For any $r \in \mathsf{RN}(\mathcal{T})$, $\mathfrak{d}(r) = 1$ captures an expression of total importance on a role (over a corresponding nested concept) and $\mathfrak{d}(r) = 0$ captures an expression of total importance on a nested concept (over a corresponding role). This notion is inspired by [6] in which sim is used with different values of the discount factors in the similarity application on Snomed ct (cf. Section 5 of [1]). For example, **Agent 2** may believe that knowing actual surrounding information is more important. Thus, $\mathfrak{d}(\mathsf{hasHabitat}) = 0.3$ and $\mathfrak{d}(\mathsf{liveIn}) = 0.3$ might be expressed this situation.

Like others, many role names might be not assigned the discount values by agents, i.e. those role names are not mapped to the corresponding values. Here, we use the default discount value of 0.4 for $r \in \mathsf{RN}(\mathcal{T})$ in case $\mathfrak{d}(r)$ is not defined. This amount of fixed value is influenced by sim [5] where the value of 0.4 is used for the discount factor when the similarity between two existential restrictions is considered. The total function $\mathfrak{d}_0(r) = 0.4$ for all $r \in \mathsf{RN}(\mathcal{T})$ is called the *default role discount factor.*

Hence, we now conclude that a preference profile π is a quintuple of preference functions, viz. $\mathfrak{i}^{\mathfrak{c}}, \mathfrak{i}^{\mathfrak{r}}, \mathfrak{s}^{\mathfrak{c}}, \mathfrak{s}^{\mathfrak{r}}$, and $\mathfrak{d}$. When a preference profile π is given, π can thereby influence the calculation of CSMs (cf. Section 5 of [1]).

Definition 6 (Preference Profile). *A* preference profile*, in symbol* π*, is a quintuple* $\langle \mathfrak{i}^{\mathfrak{c}}, \mathfrak{i}^{\mathfrak{r}}, \mathfrak{s}^{\mathfrak{c}}, \mathfrak{s}^{\mathfrak{r}}, \mathfrak{d} \rangle$ *where* $\mathfrak{i}^{\mathfrak{c}}, \mathfrak{i}^{\mathfrak{r}}, \mathfrak{s}^{\mathfrak{c}}, \mathfrak{s}^{\mathfrak{r}}$*, and* $\mathfrak{d}$ *are as defined above and the* default preference profile*, in symbol* π_0*, is the quintuple* $\langle \mathfrak{i}^{\mathfrak{c}}_0, \mathfrak{i}^{\mathfrak{r}}_0, \mathfrak{s}^{\mathfrak{c}}_0, \mathfrak{s}^{\mathfrak{r}}_0, \mathfrak{d}_0 \rangle$ *where* $\mathfrak{i}^{\mathfrak{c}}_0, \mathfrak{i}^{\mathfrak{r}}_0, \mathfrak{s}^{\mathfrak{c}}_0, \mathfrak{s}^{\mathfrak{r}}_0$*, and* $\mathfrak{d}_0$ *are as defined above.*

For example, let denote a preference profile of the **Agent 1** and the **Agent 2** by π_1 and π_2, respectively. We conclude that $\pi_1 = \langle \mathfrak{i}^{\mathfrak{c}}, \mathfrak{i}^{\mathfrak{r}}, \mathfrak{s}^{\mathfrak{c}}, \mathfrak{s}^{\mathfrak{r}}, \mathfrak{d} \rangle$, where $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Reptile}) = 2$, $\mathfrak{i}^{\mathfrak{r}}(\mathsf{hasColor}) = 2$, and $\mathfrak{s}^{\mathfrak{c}}(\mathsf{Green}, \mathsf{Yellow}) = 0.5$ indicating
1. *Reptile* is important; 2. *Having skin color* is important; and
3. *Green* and *Yellow* are similar, respectively.

Table 1: State-of-the-art CSMs intrinsically use a preference profile

CSM	$\mathfrak{i}^{\mathfrak{c}}$	$\mathfrak{i}^{\mathfrak{r}}$	$\mathfrak{s}^{\mathfrak{c}}$	$\mathfrak{s}^{\mathfrak{r}}$	$\mathfrak{d}$
sim^{π}	✔	✔	✔	✔	✔
sim [6]					✔
simi [10]	✓		✔		

In addition, we conclude that $\pi_2 = \langle \mathfrak{i}^{\mathfrak{c}}, \mathfrak{i}^{\mathfrak{r}}, \mathfrak{s}^{\mathfrak{c}}, \mathfrak{s}^{\mathfrak{r}}, \mathfrak{d} \rangle$, where $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Reptile}) = 2$, $\mathfrak{i}^{\mathfrak{r}}(\mathsf{hasHabitat}) = 2$, $\mathfrak{i}^{\mathfrak{r}}(\mathsf{liveIn}) = 2$, $\mathfrak{s}^{\mathfrak{r}}(\mathsf{t}, \mathsf{u}) = 0.1$, $\mathfrak{d}(\mathsf{hasHabitat}) = 0.3$, and $\mathfrak{d}(\mathsf{liveIn}) = 0.3$ indicating
1. *Reptile* is important;
2. *Having habitat* is important; and
3. *Having living environment* is important, respectively.

Concept Similarity Measures under Preferences

Preference profile π intends to be a generic guideline for a development of concept similarity measures used under an agent's preferences. It suggests concept similarity measures for any DLs (e.g., $\mathcal{FL}_0$, $\mathcal{ELH}$, $\mathcal{ALC}$, and so on) which permit an agent's preferences to influence the calculation should expose all elements of π. Given an arbitrary CSM $\sim$, a *concept similarity measure under preference profile* π is a function $\overset{\pi}{\sim} : \mathsf{Con}(\mathcal{L}) \times \mathsf{Con}(\mathcal{L}) \to [0, 1]$. A CSM $\sim$ is called *preference invariant w.r.t. equivalence* if $C \overset{\pi}{\sim} D = 1$ iff $C \equiv D$ for any π (cf. Section 4 of [1] for its formal definition and properties).

Developing such functions can be done from scratch or by generalizing existing CSMs. For example, we have generalized sim as a function called sim^{π} and published our theoretical development in [1]. Our sim^{π} is also *preference invariant w.r.t. equivalence*, meaning that similarity between two equivalent $\mathcal{ELH}$ concepts is always one regardless of agents' preferences. When π_0 (cf. Definition 6) has been used as the value of a preference profile, $\mathsf{sim}^{\pi_0}(C, D) = \mathsf{sim}(C, D)$ for $C, D \in \mathsf{Con}(\mathcal{ELH})$. Table 1 shows our investigation on existing CSMs and found that none of them, to the best of our knowledge, comply with our preference profile.

4 Conclusion and Future Work

We present the preference profile π as design guidelines for a development of concept similarity measures under preferences in DLs. CSMs which exposes all elements of preference profile will be appropriate to use when human perceptions are involved (See more details in [1]). This work is still in preliminary stage. We have intended to explore the possibility of implementations on realistic ontologies formulated in DLs, especially DL $\mathcal{ELH}$. We are also interested to investigate deeply desirable properties concept similarity measures under preference profiles must have. It would also interesting to investigate preference profile π when used beyond other kinds of similarity-based reasoning services. i.e., relaxed instance checking and relaxed instance retrieval.

Acknowledgments

This research is partially supported by Thammasart University Research Fund under the TU Research Scholar, Contract No. TOR POR 1/13/2558; the Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation (CILS), Thammasat University; and the JAIST-NECTEC-SIIT dual doctoral degree program.

References

1. Racharak, T., Suntisrivaraporn, B., Tojo, S.: Identifying an agent's preferences toward similarity measures in description logics. Technical-Report BS2015-01, School of Information, Computer and Communication Technology, Sirinthorn International Institute of Technology, Thammasart University, Sirinthorn International Institute of Technology, Thailand (2015) See `http://ict.siit.tu.ac.th/~sun/dw/doku.php?id=publist`.
2. Son, T.C., Cao, T., Pontelli, E.: Planning with preferences using logic programming. In: In Proc. LPNMR'04, Springer (2003) 247–260
3. Lichtenstein, S., Slovic, P., eds.: The construction of preference. Cambridge University Press, New York (2006)
4. Scherer, K.: What are emotions? and how can they be measured? Social Science Information (2005)
5. Suntisrivaraporn, B.: A similarity measure for the description logic el with unfoldable terminologies. In INCoS (2013) 408–413
6. Tongphu, S., Suntisrivaraporn, B.: Algorithms for measuring similarity between elh concept descriptions: A case study on snomed ct. Journal of Computing and Informatics (accepted on May 7; to appear) (2015)
7. Racharak, T., Suntisrivaraporn, B.: Similarity Measures for FL0 Concept Descriptions from an Automata-Theoretic Point of View. In Proceedings of The 6th annual International Conference on Information and Communication Technology for Embedded Systems (ICICTES 2015) (2015)
8. Tongphu, S., Suntisrivaraporn, B.: A Non-Standard Instance Checking for the Description Logic ELH. In Proceedings of the International Conference on Knowledge Engineering and Ontology Development (KEOD 2014) (2014)
9. Tongphu, S., Suntisrivaraporn, B.: On desirable properties of the structural subsumption-based similarity measure. In Lecture Notes in Computer Science (LNCS) **8943** (2015) 19–32
10. Lehmann, K., Turhan, A.Y.: A framework for semantic-based similarity measures for elh-concepts. In del Cerro, L.F., Herzig, A., Mengin, J., eds.: JELIA. Volume 7519 of Lecture Notes in Computer Science., Springer (2012) 307–319