Tuning Agent's Profile for Similarity Measure in Description Logic $\mathcal{ELH}$

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

[1]*School of Information, Computer, and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, Thailand*

[2]*School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan*

Keywords: Concept Similarity Measure, Semantic Web Ontology, Preference Profile, Description Logics.

Abstract: In Description Logics (DLs), concept similarity measure aims at identifying a degree of commonality of two given concepts and is often regarded as a generalization of the classical reasoning problem of equivalence. That is, any two concepts are equivalent if and only if their similarity degree is one. When two concepts are not equivalent, the level of similarity varies depending not only on the objective factors (e.g. the structure of concept descriptions) but also on the subjective factors (i.e. the agent's preferences). Realistic ontologies are generally complex. Methodologies for tuning a measure to conform with the agent's preferences should be practical, i.e. it is doable in practice. In this work, we investigate and formalize the task of tuning the preference functions based on the information defined in a TBox and an ABox. We also show how the proposed approaches can be reconciled with the measure sim^{π}, i.e. a concept similarity measure under preference profile for DL $\mathcal{ELH}$. Finally, the paper relates the approach to others and discusses future direction.

1 INTRODUCTION

Most Description Logics (DLs) are decidable fragments of first-order logic (FOL) (Baader et al., 2010) with clearly defined computational properties. DLs are the logical underpinnings of the DL flavor of OWL and OWL 2. The advantage of this close connection is that the extensive DLs literature and implementation experiences can be directly exploited by OWL tools. More specifically, DLs provide unambiguous semantics to the modeling constructs available in the DL flavor of OWL and OWL 2. These semantics make it possible to formalize and design algorithms for a number of reasoning services, which enable the development of ontology applications to become prominent. For instance, ontology classification (or ontology alignment) organizes concepts in an ontology into a subsumption hierarchy and assists in detecting potential errors of a modeling ontology. Though this subsumption hierarchy inevitably benefits ontology modeling, it merely gives two-valued responses, i.e. inferring a concept is subsumed by another concept or not. However, certain pairs of concepts may share commonality even though they are not subsumed. As a consequence, a considerable amount of research effort has been devoted on measuring similarity of two given concepts, i.e. a problem of *concept similarity measure* in DLs.

Basically, a concept similarity measure (CSM) is a function mapping from a concept pair to a unit interval (i.e. $0 \leq x \leq 1$ for any real number x). The higher the value is mapped to, the more likely similarity of them may hold. Intuitively, the value 0 can be interpreted as *total dissimilarity* whereas the value 1 can be interpreted as *total similarity* or *equivalence*. Hence, one may regard concept similarity measure as a generalization of the classical reasoning problem of equivalence. Its idealistic objective is to imitate similarity identification performed by a human expert. It plays a major role in the discovery of similar concepts in an ontology. For example, it is employed in bio-medical ontology-based applications to discover functional similarities of gene (Ashburner et al., 2000), it is often used by ontology alignment algorithms (Euzenat and Valtchev, 2004). There is currently a significant number of measures in DLs. Prominent examples are (Janowicz and Wilkes, 2009; Racharak and Suntisrivaraporn, 2015; D'Amato et al., 2006; Fanizzi and D'Amato, 2006; D'Amato et al., 2009; D'Amato et al., 2008; Racharak et al., 2016b). However, these measures are devised based on objective factors (a notable exception is (Racharak et al., 2016b) where a concept similarity measure under an agent's preferences is discussed.). For example, they use the structure (or interpretations) of concept descriptions in question to measure. When these mea-

Racharak T. and Tojo S.
Tuning Agent's Profile for Similarity Measure in Description Logic ELH.
DOI: 10.5220/0006249602870298
In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, pages 287-298
ISBN: 978-989-758-220-2

sures are employed to characterize similar concepts in an ontology, they may lead to unintuitive results. It may even be contradictory to an intuitive understanding of an application domain. The following example illustrates that using objective-based measures may not suffice to answer an agent's request.

Example 1.1 An agent A wants to visit a place for doing some active activities. At that moment, he would like to enjoy walking. Suppose that a place ontology has been modeled as follows:

$$
\begin{array}{rcl}
\mathsf{ActivePlace} & \sqsubseteq & \mathsf{Place} \sqcap \exists \mathsf{canWalk.Trekking} \\
 & & \sqcap \exists \mathsf{canSail.Kayaking} \\
\mathsf{Mangrove} & \sqsubseteq & \mathsf{Place} \sqcap \exists \mathsf{canWalk.Trekking} \\
\mathsf{Beach} & \sqsubseteq & \mathsf{Place} \sqcap \exists \mathsf{canSail.Kayaking} \\
\mathsf{canWalk} & \sqsubseteq & \mathsf{canMoveWithLegs} \\
\mathsf{canSail} & \sqsubseteq & \mathsf{canTravelWithSails}
\end{array}
$$

Suppose that a measure used by that Agent A considers merely the objective aspects, it is reasonable to conclude that both Mangrove and Beach are equally similar to the concept ActivePlace. However, by taking into account also the agent's preferences, Mangrove appears more suitable to his perception of ActivePlace at that moment. In other words, he will not be happy if an intelligent system happens to recommend him to go for a Beach.

□

The example shows that preferences of an agent play a decisive role in the choice of alternatives. In essence, when the choices of an answer are not totally similar to a concept in question, a measure may need to *be tuned* by subjective factors, e.g. an agent's preferences. A set of preferential aspects is identified in (Racharak et al., 2016a) called *preference profile* and is extended toward (Racharak et al., 2016b) called sim^{π}. However, realistic ontologies are complex – consisting in plenty of concept names and role names. Tuning a measure to conform with the agent's preferences should be practical, i.e. it is doable in practice. Hence, our primary motivation of this work is a deeper understanding of how preference profile is configured. In this work, we investigate and formalize the task of configuring the preference functions based on the information defined in a TBox and an ABox (cf. Section 3). This work is an extension of (Racharak et al., 2016b): with respect to its predecessor, it shows how the proposed approaches can be reconciled with the measure sim^{π}, i.e. a concept similarity measure under preference profile for DL $\mathcal{ELH}$ (cf. Section 4). Preliminaries is briefly reviewed in Section 2. Finally, the paper relates the approach to others (cf. Section 5 and Section 6) and discusses future direction (cf. Section 7).

2 PRELIMINARIES

In this section, we review the basics of Description Logic $\mathcal{ELH}$ (cf. Subsection 2.1), which provides the logical underpinning for OWL 2 EL (Group, 2012; Grau et al., 2008) and our developed measure sim^{π} (originally introduced in (Racharak et al., 2016b)) is based. After that, we briefly explain the notion of preference profile in Subsection 2.2.

2.1 Description Logic $\mathcal{ELH}$

We assume countably infinite sets CN of concept names, RN of role names, and IN of individual names that are fixed and disjoint. The set of concept descriptions, or simply concepts, for a specific DL $\mathcal{L}$ is denoted by $\mathsf{Con}(\mathcal{L})$. The set $\mathsf{Con}(\mathcal{ELH})$ of all $\mathcal{ELH}$ concepts can be inductively defined by the following grammar:

$$\mathsf{Con}(\mathcal{ELH}) ::= A \mid \top \mid C \sqcap D \mid \exists r.C$$

where $\top$ denotes the *top concept*, $A \in \mathsf{CN}$, $r \in \mathsf{RN}$, and $C, D \in \mathsf{Con}(\mathcal{ELH})$. 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, all possibly with subscripts.

A *terminology* or *TBox* $\mathcal{T}$ is a finite set of (possibly primitive) concept definitions and role hierarchy axioms, whose syntax is an expression of the form $(A \sqsubseteq D)$ $A \equiv D$, and $r \sqsubseteq s$, respectively. 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 $\mathsf{CN}^{\mathsf{def}}$), all other concept names are primitive concept names (denoted by $\mathsf{CN}^{\mathsf{pri}}$). 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*. 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 $\mathsf{RN}^{\mathsf{def}}$. All others are primitive role names, collectively denoted by $\mathsf{RN}^{\mathsf{pri}}$. We also denote a set of all r's super roles by $\mathcal{R}_r = \{s \in \mathsf{RN} | r = s \text{ or } r_i \sqsubseteq r_{i+1} \in \mathcal{T} \text{ where } 1 \leq i \leq n, r_1 = r, r_n = s\}$.

An *assertion* or ABox $\mathcal{A}$ is a finite set of concept assertions and role assertions whose syntax is an expression of the form $C(a)$ and $r(a,b)$ where $a,b \in \mathsf{IN}$, respectively. An ontology O consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$, i.e. $O = \langle \mathcal{T}, \mathcal{A} \rangle$. However, some existing ontologies may omit an ABox $\mathcal{A}$ in practice.

An interpretation I is a pair $I = \langle \Delta^I, \cdot^I \rangle$ where Δ^I, is a non-empty set representing the domain of the interpretation and $\cdot^I$ is an interpretation function which assigns to every concept name A a set $A^I \subseteq \Delta^I$, and to every role name r a binary relation $r^I \subseteq \Delta^I \times \Delta^I$. The interpretation function $\cdot^I$ is inductively extended to $\mathcal{ELH}$ concepts in the usual manner:

$$\top^I = \Delta; \qquad (C \sqcap D)^I = C^I \cap D^I;$$
$$(\exists r.C)^I = \{a \in \Delta^I \mid \exists b \in \Delta^I : (a,b) \in r^I \wedge b \in C^I\}.$$

An interpretation I is said to be a *model* of a TBox $\mathcal{T}$ (in symbols, $I \models \mathcal{T}$) if it satisfies all axioms in $\mathcal{T}$. I satisfies axioms $A \sqsubseteq C$, $A \equiv C$, and $r \sqsubseteq s$, respectively, if $A^I \subseteq C^I$, $A^I = C^I$, and $r^I \subseteq s^I$. Also, an interpretation I is said to be a *model* of an ABox $\mathcal{A}$ (in symbols, $I \models \mathcal{A}$) if it satisfies all axioms in $\mathcal{A}$. I satisfies axioms $C(a)$ and $r(a,b)$ if $a^I \in C^I$ and $(a,b) \in r^I$, respectively. Furthermore, an interpretation I is said to be a *model* of an ontology O if it satisfies all axioms in $\mathcal{T}$ and $\mathcal{A}$. An interpretation $I_{\mathcal{A}}$ is called the *canonical interpretation* if:

1. $\Delta^{I_{\mathcal{A}}}$ of $I_{\mathcal{A}}$ consists of all individual names in $\mathcal{A}$;
2. $\forall A \in \mathsf{CN}$, we define $A^{I_{\mathcal{A}}} = \{x \mid A(x) \in \mathcal{A}\}$; and
3. $\forall r \in \mathsf{RN}$, we define $r^{I_{\mathcal{A}}} = \{(x,y) \mid r(x,y) \in \mathcal{A}\}$.

That is, the canonical interpretation $I_{\mathcal{A}}$ is the interpretation which takes the set of ABox as the interpretation domain.

The main inference problem for $\mathcal{ELH}$ is the subsumption problem. That is, given $C, D \in \mathsf{Con}(\mathcal{ELH})$ and an ontology O, C is *subsumed* by D w.r.t. O (in symbols, $C \sqsubseteq_O D$) if $C^I \subseteq D^I$ for every model I of O. Furthermore, C and D are *equivalent* w.r.t. O (in symbols, $C \equiv_O D$) if $C \sqsubseteq_O D$ and $D \sqsubseteq_O C$. A much more interesting inference problem, which is based on concept subsumption, is the concept hierarchy. That is, let $\mathsf{CN}(O)$ be the concept names occurring in O, the concept hierarchy of O is the most compact representation of the partial ordering $(\mathsf{CN}(O), \sqsubseteq_O)$ induced by the subsumption relation w.r.t O. When an ontology O is empty or is clear from the context, we omit to denote O, i.e. $C \sqsubseteq D$ or $C \equiv D$. Furthermore, a more generalization of the concept equivalence is a *concept similarity measure under preference profile* (cf. Definition 2.3), which is originally introduced in (Racharak et al., 2016b).

2.2 Preference Profile

We first introduced *preference profile* (denoted by π) in (Racharak et al., 2016a) as a collection of preferential elements in which the development of concept similarity measure should consider. Its first intuition is to model different forms of preferences (of an agent) based on concept names and role names. Measures adopted this notion are flexible to be tuned by an agent and can determine the similarity conformable to that agent's perception. We give a formal definition of each preferential aspect in the following definition.

Definition 2.1 (Preference Profile (Racharak et al., 2016a)). Let $\mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$, $\mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$, and $\mathsf{RN}(\mathcal{T})$ be a set of primitive concept names occurring in $\mathcal{T}$, a set of primitive role names occurring in $\mathcal{T}$, and a set of role names occurring in $\mathcal{T}$, respectively. A *preference profile*, in symbol π, is a quintuple $\langle \mathfrak{i}^c, \mathfrak{i}^r, \mathfrak{s}^c, \mathfrak{s}^r, \mathfrak{d} \rangle$[1] where

- $\mathfrak{i}^c : \mathsf{CN} \to [0,2]$ where $\mathsf{CN} \subseteq \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ is called *primitive concept importance*;
- $\mathfrak{i}^r : \mathsf{RN} \to [0,2]$ where $\mathsf{RN} \subseteq \mathsf{RN}(\mathcal{T})$ is called *role importance*;
- $\mathfrak{s}^c : \mathsf{CN} \times \mathsf{CN} \to [0,1]$ where $\mathsf{CN} \subseteq \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ is called *primitive concepts similarity*;
- $\mathfrak{s}^r : \mathsf{RN} \times \mathsf{RN} \to [0,1]$ where $\mathsf{RN} \subseteq \mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$ is called *primitive roles similarity*; and
- $\mathfrak{d} : \mathsf{RN} \to [0,1]$ where $\mathsf{RN} \subseteq \mathsf{RN}(\mathcal{T})$ is called *role discount factor*.

We discuss the interpretation of each above function in order. Firstly, for any $A \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$, $\mathfrak{i}^c(A) = 1$ captures an expression of normal importance on A, $\mathfrak{i}^c(A) > 1$ ($\mathfrak{i}^c(A) < 1$) indicates that A has higher (and lower, respectively) importance, and $\mathfrak{i}^c(A) = 0$ indicates that A has no importance to the agent. Secondly, we define the interpretation of $\mathfrak{i}^r$ in the similar fashion as $\mathfrak{i}^c$ for any $r \in \mathsf{RN}(\mathcal{T})$. Thirdly, for any $a,b \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$, $\mathfrak{s}^c(A,B) = 1$ captures an expression of total similarity between A and B and $\mathfrak{s}^c(A,B) = 0$ captures an expression of total dissimilarity between A and B. Fourthly, the interpretation of $\mathfrak{s}^r$ is defined in the similar fashion as $\mathfrak{s}^c$ for any $r,s \in \mathsf{RN}^{pri}(\mathcal{T})$. Lastly, 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).

Definition 2.2 (Default Preference Profile (Racharak et al., 2016a)). The *default preference profile*, in symbol π_0, is the quintuple $\langle \mathfrak{i}^c_0, \mathfrak{i}^r_0, \mathfrak{s}^c_0, \mathfrak{s}^r_0, \mathfrak{d}_0 \rangle$

[1]In the original definition of preference profile (Racharak et al., 2016a; Racharak et al., 2016b), both $\mathfrak{i}^c$ and $\mathfrak{i}^r$ are mapped to $\mathbb{R}_{\geq 0}$, which is a minor error.

where

$$\begin{aligned}
\mathfrak{i}^{\mathfrak{c}}_0(A) &= 1 \text{ for all } A \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}),\\
\mathfrak{i}^{\mathfrak{r}}_0(r) &= 1 \text{ for all } r \in \mathsf{RN}(\mathcal{T}),\\
\mathfrak{s}^{\mathfrak{c}}_0(A,B) &= 0 \text{ for all } (A,B) \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}) \times \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}),\\
\mathfrak{s}^{\mathfrak{r}}_0(r,s) &= 0 \text{ for all } (r,s) \in \mathsf{RN}^{\mathsf{pri}}(\mathcal{T}) \times \mathsf{RN}^{\mathsf{pri}}(\mathcal{T}),\\
\mathfrak{d}_0(r) &= 0.4 \text{ for all } r \in \mathsf{RN}(\mathcal{T}).
\end{aligned}$$

Let us also note that the value of $\mathfrak{d}_0$ determines how important the existential information should be considered by a measure in the default manner (see the interpretation of $\mathfrak{d}$). This information is indeed dependent on an application domain and might be redefined. For instance, if $\mathfrak{d}_0$ is defined as 0.3, 0.4, 0.5, then $\exists\mathsf{part.Heart} \overset{\pi}{\sim}_{\mathcal{T}} \mathsf{part.Colon}$ yields 0.3, 0.4, 0.5, respectively.[2] In the following, we give the formal definition of concept similarity measure under preference profile.

Definition 2.3 ((Racharak et al., 2016b)). Given a preference profile π, two concepts $C, D \in \mathsf{Con}(\mathcal{L})$, and a TBox $\mathcal{T}$, a *concept similarity measure under preference profile* w.r.t. a TBox $\mathcal{T}$ is a function $\overset{\pi}{\sim}_{\mathcal{T}} : \mathsf{Con}(\mathcal{L}) \times \mathsf{Con}(\mathcal{L}) \to [0,1]$. A function $\overset{\pi}{\sim}_{\mathcal{T}}$ is called *preference invariance w.r.t equivalence* if $C \equiv D \Leftrightarrow (C \overset{\pi}{\sim}_{\mathcal{T}} D = 1$ for any $\pi)$.

Taking $\pi = \pi_0$ for a concept similarity measure under preference profile, i.e. $\overset{\pi_0}{\sim}_{\mathcal{T}}$, obtains an objective similarity degree. We prove this in Theorem 3.1 of (Racharak et al., 2016b).

3 STRATEGIES OF TUNING π

3.1 Tuning $\mathfrak{i}^{\mathfrak{c}}$

This subsection exhibits a strategy for tuning primitive concept importance $\mathfrak{i}^{\mathfrak{c}}$ in practice. Realistic ontologies are generally complex – consisting in plenty of concept names. Hence, having some strategies of tuning is useful since it helps to pave the way for a more convenient use of preference profile.

As a starting point, we seek to observe characteristics of realistic ontologies whose TBox is *unfoldable*[3], e.g. a popular medical ontology SNOMED, denoted by $\mathcal{O}_{\mathsf{med}}$, (Stearns et al., 2001). Figure 1 gives an example of concept definitions in $\mathcal{O}_{\mathsf{med}}$ and Figure 2 shows the concept hierarchy w.r.t. $\mathcal{O}_{\mathsf{med}}$.

$$\begin{array}{rcl}
\mathsf{Pericardium} & \sqsubseteq & \mathsf{Tissue} \sqcap \exists\mathsf{part.Heart}\\
\mathsf{Endocardium} & \sqsubseteq & \mathsf{Tissue} \sqcap \exists\mathsf{part.Heart}\\
\mathsf{Appendicitis} & \equiv & \mathsf{Inflammation}\\
 & & \sqcap\exists\mathsf{loc.Appendix}\\
\mathsf{Pericarditis} & \equiv & \mathsf{Inflammation}\\
 & & \sqcap\exists\mathsf{loc.Pericardium}\\
\mathsf{Endocarditis} & \equiv & \mathsf{Inflammation}\\
 & & \sqcap\exists\mathsf{loc.Endocardium}\\
\mathsf{Inflammation} & \sqsubseteq & \mathsf{Disease}\\
\mathsf{HeartDisease} & \equiv & \mathsf{Disease} \sqcap \exists\mathsf{loc}.\exists\mathsf{part.Heart}
\end{array}$$

Figure 1: Example of concept definitions in $\mathcal{O}_{\mathsf{med}}$.

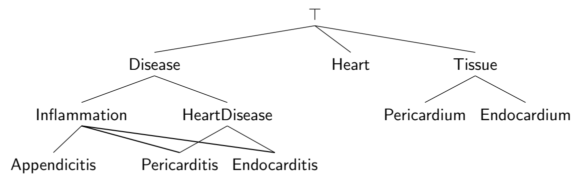

Figure 2: The concept hierarchy of $\mathcal{O}_{\mathsf{med}}$.

According to the above figures, it is intuitive to express primitive concept importance through the concept hierarchy. For instance, an agent may say *my concept importance goes through* Disease; or *my concept importance goes through* HeartDisease. Informally investigating, let $\mathsf{CN}^{\mathsf{pri}}(\mathcal{O}_{\mathsf{med}})$ be a set of primitive concept names occurring in $\mathcal{O}_{\mathsf{med}}$. Since $\mathsf{Disease} \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{O}_{\mathsf{med}})$, the former case is simple, e.g. an agent may mean $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$. The latter case is a bit complicated since $\mathsf{HeartDisease} \notin \mathsf{CN}^{\mathsf{pri}}(\mathcal{O}_{\mathsf{med}})$. However, the agent's intention may mean $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$ and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 1.2$. This informal investigation shapes the development as follows:

Definition 3.1. Let $\mathsf{CN}(\mathcal{T})$ ($\mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ and $\mathsf{CN}^{\mathsf{def}}(\mathcal{T})$) be a set of concept names (primitive concept names and defined concept names, respectively) occurring in $\mathcal{T}$. Then, a *propagation for primitive concept importance* is a partial function $\mathfrak{I}^{\mathfrak{c}} : \mathsf{CN}' \to [0,2]$, where $\mathsf{CN}' \subseteq \mathsf{CN}(\mathcal{T}) \cup \{\top\}$, such that a mapping n of $\mathfrak{I}^{\mathfrak{c}}$ on X (i.e. $\mathfrak{I}^{\mathfrak{c}}(X) = n$) is defined inductively as follows:[4].

1. $X \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}) \Rightarrow \mathfrak{i}^{\mathfrak{c}}(X) = n$;
2. $X := \top \Rightarrow \forall x \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T}) : \mathfrak{i}^{\mathfrak{c}}(x) = n$; and
3. $X \in \mathsf{CN}^{\mathsf{def}}(\mathcal{T}) \Rightarrow \forall x \in \mathsf{RHS}(X) : \mathfrak{I}^{\mathfrak{c}}(x) = n$.

where $\mathsf{RHS}(X)$ is a set of concept names appearing on the right-hand side of X.

Its interpretation is defined in a usual way. That is, for any $A \in \mathsf{CN}(\mathcal{T})$, $\mathfrak{I}^{\mathfrak{c}}(A) = 1$ captures an expression of normal importance on A, $\mathfrak{I}^{\mathfrak{c}}(A) > 1$ (and $\mathfrak{I}^{\mathfrak{c}}(A) <$

[2] These numbers is obtained from an application of sim^{π}. Its precise definition is given in Section 4.

[3] According to this investigation, we assume an ontology $\mathcal{O}$ has an unfoldable TBox $\mathcal{T}$ in this work.

[4] Later, we discuss some restrictions the readers should take into account when the notion $\mathfrak{I}^{\mathfrak{c}}$ is employed.

1) indicates that A has higher (and lower, respectively) importance, $\mathfrak{I}^{\mathfrak{c}}(A) = 0$ indicates that A has no importance to the agent. For the special case, $\mathfrak{I}^{\mathfrak{c}}(\top) = n$ indicates that every primitive concept name occurring on $\mathcal{T}$ is of equal importance at n.

Example 3.1. From Figure 1, suppose that an agent A is using a concept similarity measure under preference profile for querying some names that expose the similar characteristics to HeartDisease. Thus, the agent can express a preference $\mathfrak{I}^{\mathfrak{c}}(\mathsf{HeartDisease}) = 1.2$ instead of individually specifying $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$ and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 1.2$. □

There are a few concerns that we should take into account, i.e. **(1)** inconsistent preferences of concepts occurring on the same branch of the concept hierarchy; **(2)** inconsistent preferences of defined concepts occurring on different branches of the concept hierarchy; and **(3)** expressing preferences when a TBox $\mathcal{T}$ contains equivalent defined concepts.

Let us take a look on our first concern through the following preference expression: *my concept importance goes through* Disease*, especially* HeartDisease. In this example, we may take $\mathfrak{I}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$ and $\mathfrak{I}^{\mathfrak{c}}(\mathsf{HeartDisease}) = 1.3$. Here, Disease is redefined on $\mathfrak{i}^{\mathfrak{c}}$ twice. This kind of scenarios is possible to happen because we are extending the aspect of primitive concept importance toward both types of concept names. There are many ways to handle this with the use of operators $\oplus : [0,2]^2 \to [0,2]$.

Definition 3.2. Let $A \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$ be a set of primitive concept names occurring in $\mathcal{T}$ and $x_0, x_1 \in [0,2]$. Also, let $\mathfrak{i}^{\mathfrak{c}}(A) = x_0$ be the previous mapping on A. We compute a new mapping $\mathfrak{i}^{\mathfrak{c}}(A) = x_1$ as follows:

$$\mathfrak{i}^{\mathfrak{c}}(A) = \begin{cases} x_1 & \text{if } \mathfrak{i}^{\mathfrak{c}} \text{ is not defined on } A \\ x_0 \oplus x_1 & \text{otherwise} \end{cases} \quad (1)$$

The notion of the operator remains abstract here as its concrete operators may vary on the context of use. In the following, we establish some of the abstract notion $\oplus$, i.e. $\oplus_{max}$, $\oplus_{first}$, and $\oplus_{last}$. Let two real numbers $x_1, x_0 \in [0,2]$. Then,

$$x_0 \oplus_{max} x_1 = \max\{x_0, x_1\} \quad (2)$$

$$x_0 \oplus_{first} x_1 = x_0 \quad (3)$$

$$x_0 \oplus_{last} x_1 = x_1 \quad (4)$$

Example 3.2. From Figure 1, an agent might say *My interest is* Disease *except* HeartDisease. That is, we may take $\mathfrak{I}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$ (i.e. $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$) and $\mathfrak{I}^{\mathfrak{c}}(\mathsf{HeartDisease}) = 0$ (i.e. $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2 \oplus 0$ and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 0$). Taking $\oplus$ as $\oplus_{max}$ yields $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1.2$ and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 0$. It also yields the same results by taking $\oplus$ as $\oplus_{first}$. □

Example 3.3. From Figure 1, an agent might say *My concern is nothing except* HeartDisease. That is, we may take $\mathfrak{I}^{\mathfrak{c}}(\top) = 0$ (i.e. $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 0$, $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Tissue}) = 0$, and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 0$) and $\mathfrak{I}^{\mathfrak{c}}(\mathsf{HeartDisease}) = 1$ (i.e. $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 0 \oplus 1$ and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 0 \oplus 1$). Taking $\oplus$ as $\oplus_{max}$ yields $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Disease}) = 1$, $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Tissue}) = 0$, and $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Heart}) = 1$. It also yields the same results by taking $\oplus$ as $\oplus_{last}$. □

Now, we discuss our second concern on the application of $\mathfrak{I}^{\mathfrak{c}}$. Let us consider the following preference expression of an agent: $\mathfrak{I}^{\mathfrak{c}}(\mathsf{Pericarditis}) = 1.2$ and $\mathfrak{I}^{\mathfrak{c}}(\mathsf{Endocarditis}) = 1.6$. One may notice that both use the primitive concept name Disease in common. Now, a natural question to ask is how a concept importance value should be propagated since a propagation may cause an inconsistency of preference values for a primitive concept name, such as Disease in this example. This requires further work to study. However, one simple way for handling this problem is to prevent a mapping leading to this situation. Instead, an agent has to tune primitive concept names via the primitive concept importance $\mathfrak{i}^{\mathfrak{c}}$ individually.

Lastly, we contemplate our third concern. That is, what happens if a defined concept name C_1 is defined on $\mathfrak{I}^{\mathfrak{c}}$ and there exists another defined concept name C_2 such that $C_1 \equiv_{\mathcal{T}} C_2$? A natural way for handling this problem is to treat C_2 in the same way as C_1 (because they are equivalent). In particular, $\forall C_1, C_2 \in \mathsf{CN}^{\mathsf{def}}(\mathcal{T}) : C_1 \equiv_{\mathcal{T}} C_2 \Rightarrow \mathfrak{I}^{\mathfrak{c}}(C_1) = \mathfrak{I}^{\mathfrak{c}}(C_2)$. Nevertheless, this also requires further work to explore other possibilities for coping with this problem and investigate desired properties the notion $\overset{\pi}{\sim}_{\mathcal{T}}$ should hold when it is used with $\mathfrak{I}^{\mathfrak{c}}$.

3.2 Tuning $\mathfrak{i}^{\mathfrak{r}}$

Let us remind that $\mathfrak{i}^{\mathfrak{r}}$ is a function which maps a role name $r \in \mathsf{RN}$ to a value $x \in [0,2]$. Its primary motivation is to define a user-identified importance value for an individual role name. A distinguished characteristic of $\mathfrak{i}^{\mathfrak{r}}$ to $\mathfrak{i}^{\mathfrak{c}}$ is that, not only restricted to primitive ones, ones may also define an importance on defined role names. We bear this understanding on the development of a strategy to tune $\mathfrak{i}^{\mathfrak{r}}$ as follows.

Our primary motivation of providing a strategy to help tuning $\mathfrak{i}^{\mathfrak{r}}$ is similar to that one of $\mathfrak{i}^{\mathfrak{c}}$. That is, realistic ontologies are complex – consisting in plenty of role names. An intuitive way to simplify the task of tuning $\mathfrak{i}^\mathfrak{r}$ is to proceed on a more general role name. We note that, suppose $r \sqsubseteq s \in \mathcal{T}$, a role s is said to be more general than a role r. This intuition shapes our development as follows:

Definition 3.3. Let $\mathsf{RN}(\mathcal{T})$ be a set of role names occurring in $\mathcal{T}$. Then, a *propagation for role importance* is a partial function $\mathfrak{I}^\mathfrak{r} : \mathsf{RN}' \to [0,2]$, where $\mathsf{RN}' \subseteq \mathsf{RN}(\mathcal{T})$, such that a mapping n of $\mathfrak{I}^\mathfrak{r}$ on X (i.e. $\mathfrak{I}^\mathfrak{r}(X) = n$) is inductively defined as follows:[5].

1. $\forall X' \in \mathsf{RN}(\mathcal{T})$: $(X' \sqsubseteq X \in \mathcal{T} \Rightarrow \mathfrak{i}^\mathfrak{r}(X) = n$ and $\mathfrak{I}^\mathfrak{r}(X') = n)$; and
2. $\forall X' \in \mathsf{RN}(\mathcal{T})$: $(X' \sqsubseteq X \notin \mathcal{T} \Rightarrow \mathfrak{i}^\mathfrak{r}(X) = n)$.

There are a few concerns that we should take into account, i.e. **(1)** inconsistent preferences of roles occurring on the same branch of the role hierarchy; and **(2)** expressing preferences when a TBox $\mathcal{T}$ contains equivalent role names. We discuss these in order.

Let us take a look on our first concern through the following preference expression (according to an ontology given in Example 1.1): *my role importance goes through* $\mathsf{canMoveWithLegs}$*, especially* $\mathsf{canWalk}$. Suppose we take $\mathfrak{I}^\mathfrak{c}(\mathsf{canMoveWithLegs}) = 1.2$ and $\mathfrak{I}^\mathfrak{c}(\mathsf{canWalk}) = 1.3$. Here, $\mathsf{canWalk}$ is redefined on $\mathfrak{i}^\mathfrak{r}$ twice. Similar to $\mathfrak{I}^\mathfrak{c}$, we handle this problem with the use of operators $\oplus : [0,2]^2 \to [0,2]$.

Definition 3.4. Let $r \in \mathsf{RN}(\mathcal{T})$ be a set of role names occurring in $\mathcal{T}$ and $x_0, x_1 \in [0,2]$. Also, let $\mathfrak{i}^\mathfrak{r}(r) = x_0$ be the previous mapping on r. We compute a new mapping $\mathfrak{i}^\mathfrak{r}(r) = x_1$ as follows:

$$\mathfrak{i}^\mathfrak{r}(r) = \begin{cases} x_1 & \text{if } \mathfrak{i}^\mathfrak{r} \text{ is not defined on } r \\ x_0 \oplus x_1 & \text{otherwise} \end{cases} \tag{5}$$

The operator remains abstract here as its concrete operators may vary on the context of use and may be defined in the same sense as $\mathfrak{I}^\mathfrak{c}$ (e.g. Equation 2 to 4). For example, an agent may prefer to take the last mapping when that agent says exceptional cases (e.g. r except s where $r \in \mathcal{R}_s$) in order to suppress the previously propagated value. Also, an agent may prefer to take the last mapping when that agent would like to emphasize some special circumstances (e.g. r especially s where $r \in \mathcal{R}_s$) in order to suppress the previously propagated value.

Example 3.4. From Example 1.1, an agent might say *My interest is* $\mathsf{canMoveWithLegs}$ *except* $\mathsf{canWalk}$. That is, we may take $\mathfrak{I}^\mathfrak{r}(\mathsf{canMoveWithLegs}) = 1.2$ (i.e. $\mathfrak{i}^\mathfrak{r}(\mathsf{canMoveWithLegs}) = 1.2$ and $\mathfrak{i}^\mathfrak{r}(\mathsf{canWalk}) = 1.2$) and $\mathfrak{I}^\mathfrak{r}(\mathsf{canWalk}) = 0$ (i.e. $\mathfrak{i}^\mathfrak{r}(canWalk) = 1.2 \oplus 2$). Using $\oplus_{first}$ for $\oplus$ yields $\mathfrak{i}^\mathfrak{r}(\mathsf{canMoveWithLegs}) = 1.2$ and $\mathfrak{i}^\mathfrak{r}(\mathsf{canWalk}) = 0$. □

Example 3.5. From Example 1.1, an agent might say *My interest is* $\mathsf{canMoveWithLegs}$*, especially* $\mathsf{canWalk}$. Let us take $\mathfrak{I}^\mathfrak{c}(\mathsf{canMoveWithLegs}) = 1.2$ (i.e. $\mathfrak{i}^\mathfrak{r}(\mathsf{canMoveWithLegs}) = 1.2$ and $\mathfrak{i}^\mathfrak{r}(\mathsf{canWalk}) = 1.2$) and $\mathfrak{I}^\mathfrak{c}(\mathsf{canWalk}) = 1.3$ (i.e. $\mathfrak{i}^\mathfrak{r}(\mathsf{canWalk}) = 1.2 \oplus 1.3$). Using $\oplus_{last}$ for $\oplus$ yields $\mathfrak{i}^\mathfrak{r}(\mathsf{canMoveWithLegs}) = 1.2$ and $\mathfrak{i}^\mathfrak{r}(\mathsf{canWalk}) = 1.3$. □

Lastly, we discuss the second concern. That is, what happens if a defined role name r_1 is defined on $\mathfrak{I}^\mathfrak{r}$ and there exists another defined role name r_2 such that $r_1 \sqsubseteq_\mathcal{T} r_2$ and $r_2 \sqsubseteq_\mathcal{T} r_1$? Similar to our basic handling of this case in $\mathfrak{I}^\mathfrak{c}$, we recommend to treat r_2 in the same way as r_1 (because they are equivalent). Nevertheless, this also requires further work to explore other possibilities for coping with this problem and investigate desired properties the notion $\overset{\pi}{\sim}_\mathcal{T}$ should hold when it is used with $\mathfrak{I}^\mathfrak{r}$.

3.3 Tuning $\mathfrak{s}^\mathfrak{c}$

In this subsection, we present a strategy for tuning primitive concepts similarity. If an ABox $\mathcal{A}$ is presented, then we can induce the canonical interpretation $I_\mathcal{A}$ from $\mathcal{A}$ to calculate primitive concepts similarity for all possible primitive concept pairs. Suppose that $I_\mathcal{A}$ is constructed and let $A, B \in \mathsf{CN}^{\mathsf{pri}}(\mathcal{T})$, we establish the following calculation for the function $\mathfrak{s}^\mathfrak{c}$.

$$\mathfrak{s}^\mathfrak{c}(A,B) = \begin{cases} 1 & \text{if } A^{I_\mathcal{A}} = B^{I_\mathcal{A}} = \emptyset \\ \frac{|A^{I_\mathcal{A}} \cap B^{I_\mathcal{A}}|}{|A^{I_\mathcal{A}} \cup B^{I_\mathcal{A}}|} & \text{otherwise} \end{cases} \tag{6}$$

where $|\cdot|$ represents the set cardinality.

Intuitively, Equation 6 computes the commonality of both primitive concept names. Since O_{med} does not contain an ABox $\mathcal{A}$, let us use a handcraft ontology to exemplify the calculation.

Example 3.6. Let a family ontology $O = \langle \mathcal{T}, \mathcal{A} \rangle$ in which $\mathcal{T}$ is defined as follows:

$$\begin{aligned} \mathsf{Grandfather} &\equiv \mathsf{Man} \sqcap \exists \mathsf{child.Parent} \\ \mathsf{Parent} &\equiv \mathsf{Person} \sqcap \exists \mathsf{child.Person} \\ \mathsf{Man} &\equiv \mathsf{Male} \sqcap \mathsf{Person} \end{aligned}$$

Let an ABox $\mathcal{A}$ is defined as follows:

$$\begin{array}{rl} \mathsf{child}(john, elise) & \mathsf{child}(emma, watson) \\ \mathsf{Person}(john) & \mathsf{Person}(elise) \\ \mathsf{Person}(emma) & \mathsf{Person}(watson) \\ \mathsf{Male}(john) & \mathsf{Male}(watson) \end{array}$$

Thus, $\Delta^{I_\mathcal{A}} = \{elise,\ john,\ emma,\ watson\}$, $\mathsf{Person}^{I_\mathcal{A}} = \{john,\ elise,\ emma,\ watson\}$, $\mathsf{Male}^{I_\mathcal{A}}$

[5] Later, we discuss some restrictions the readers should take into account when the notion $\mathfrak{I}^\mathfrak{r}$ is employed.

$= \{john, \ watson\}$, and $\mathfrak{s}^{\mathfrak{c}}(\mathsf{Person}, \mathsf{Male}) = \frac{|john, watson|}{|john, elise, emma, watson|} = \frac{1}{2} = 0.5$.

□

3.4 Tuning $\mathfrak{s}^{\mathfrak{r}}$

This subsection presents a strategy for tuning primitive roles similarity. Indeed, we attempt in the similar fashion as what we do for $\mathfrak{s}^{\mathfrak{c}}$. That is, we use the canonical interpretation $I_{\mathcal{A}}$ to obtain primitive roles similarity. Let $r, s \in \mathsf{RN}^{\mathsf{pri}}(\mathcal{T})$ and define operators $\cdot^f$ and $\cdot^s$ for any primitive role r as $r^f = \{x \mid (x,y) \in r^{I_{\mathcal{A}}}\}$ and $r^s = \{y \mid (x,y) \in r^{I_{\mathcal{A}}}\}$, respectively, then:

$$\mathfrak{s}^{\mathfrak{r}}(r,s) = \begin{cases} 1 & \text{if } r^{I_{\mathcal{A}}} = s^{I_{\mathcal{A}}} = \emptyset \\ \lambda \cdot \frac{|r^f \cap s^f|}{|r^f \cup s^f|} + (1-\lambda) \cdot \frac{|r^s \cap s^s|}{|r^s \cup s^s|} & \text{otherwise} \end{cases} \tag{7}$$

where $0 < \lambda < 1$ and $|\cdot|$ represents the set cardinality.

Intuitively, Equation 7 is defined as the weighted sum of the commonality on the first arguments of roles and the commonality on the second arguments of roles. It is recommended to set the weight $\lambda = \frac{|r^f \cup s^f|}{|r^f \cup s^f| + |r^s \cup s^s|}$, i.e. the proportion of all individuals appearing on the first arguments to all individuals appearing on both arguments. The following example exemplifies the calculation.

Example 3.7. Let a family ontology $O = \langle \mathcal{T}, \mathcal{A} \rangle$ in which $\mathcal{T}$ is defined as follows:

$$\begin{aligned} \mathsf{Parent} &\equiv \mathsf{Person} \sqcap \exists \mathsf{child}.\mathsf{Person} \\ \mathsf{BrotherSister} &\equiv \mathsf{Person} \sqcap \exists \mathsf{sibling}.\mathsf{Person} \end{aligned}$$

Let an ABox $\mathcal{A}$ is defined as follows:

$$\begin{array}{rl} \mathsf{sibling}(john, max) & \mathsf{sibling}(yok, watson) \\ \mathsf{child}(emma, yok) & \mathsf{child}(john, elise) \\ \mathsf{child}(emma, watson) & \mathsf{Person}(john) \\ \mathsf{Person}(elise) & \mathsf{Person}(emma) \\ \mathsf{Person}(watson) & \mathsf{Person}(max) \\ \mathsf{Person}(yok) & \end{array}$$

Thus, $\Delta^{I_{\mathcal{A}}} = \{elise, \ john, \ emma, \ watson, \ max, \ yok\}$, $\mathsf{child}^f = \{emma, john\}$, $\mathsf{sibling}^f = \{yok, john\}$, $\mathsf{child}^s = \{yok, watson, elise\}$, $\mathsf{sibling}^s = \{watson, max\}$, and $\mathfrak{s}^{\mathfrak{r}}(\mathsf{child}, \mathsf{sibling}) = \frac{3}{7} \cdot \frac{1}{3} + \frac{4}{7} \cdot \frac{1}{4} \approx 0.48$.

□

3.5 Tuning $\mathfrak{d}$

The primary motivation of this aspect is to capture an expression of total expression on a role beyond a corresponding nested concept (Racharak et al., 2016a). Hence, tuning this aspect may requires skilled domain expertise. For example, SNOMED ontology engineers realize that roleGroup is used to nestedly group existential restrictions; hence, it can unintentionally increase the degree of similarity due to role commonality. Considering this fact, they may set $\mathfrak{d}(\mathsf{roleGroup}) = 0$. This shows that role discount factor of different role names may be independent. However, the same strategy as $\mathfrak{I}^{\mathfrak{r}}$ can be employed to comfort on configuring this aspect, i.e. a propagation for role discount factor via a more general role name.

4 APPLYING PROPOSED STRATEGIES TO MEASURE $\mathcal{ELH}$ CONCEPTS

In this section, we show an applicability of the proposed strategies for tuning preference profile π to be used with the measure sim^π (Racharak et al., 2016b). We note that sim^π is an instance of $\overset{\pi}{\sim}_{\mathcal{T}}$ (Definition 2.3) for DL $\mathcal{ELH}$. Using sim^π requires that concept definitions in a TBox $\mathcal{T}$ must be fully expanded, i.e. for each defined concept name $A \in \mathsf{CN}^{\mathsf{def}}(\mathcal{T})$, such that $A \equiv D$, we simply replace A with D wherever it occurs in C and continue to recursively expand D. If A is of the form $A \sqsubseteq D$, then we replace A with $X \sqcap D$ such that X is a fresh concept wherever A occurs in C and recursively expand D. We note that X represents the primitiveness of A, i.e. the unspecified characteristics that differentiate it from D.

For the purpose of self-containment, we include the original definition of homomorphism degree under preference profile hd^π and the similarity degree under preference profile sim^π here.

In order to consider all aspects of preference profile, we have presented a *total importance function* as $\hat{\mathfrak{i}} : \mathsf{CN}^{\mathsf{pri}} \cup \mathsf{RN} \to [0,2]$

$$\hat{\mathfrak{i}}(x) = \begin{cases} \mathfrak{i}^{\mathfrak{c}}(x) & \text{if } x \in \mathsf{CN}^{\mathsf{pri}} \text{ and } \mathfrak{i}^{\mathfrak{c}} \text{ is defined on } x \\ \mathfrak{i}^{\mathfrak{r}}(x) & \text{if } x \in \mathsf{RN} \text{ and } \mathfrak{i}^{\mathfrak{r}} \text{ is defined on } x \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

A *total similarity function* is also presented as $\hat{\mathfrak{s}} : (\mathsf{CN}^{\mathsf{pri}} \times \mathsf{CN}^{\mathsf{pri}}) \cup (\mathsf{RN}^{\mathsf{pri}} \times \mathsf{RN}^{\mathsf{pri}}) \to [0,1]$ using primitive concepts similarity and primitive roles similarity.

$$\hat{\mathfrak{s}}(x,y) = \begin{cases} 1 & \text{if } x = y \\ \mathfrak{s}^{\mathfrak{c}}(x,y) & \text{if } (x,y) \in \mathsf{CN}^{\mathsf{pri}} \times \mathsf{CN}^{\mathsf{pri}} \\ & \text{and } \mathfrak{s}^{\mathfrak{c}} \text{ is defined on } (x,y) \\ \mathfrak{s}^{\mathfrak{r}}(x,y) & \text{if } (x,y) \in \mathsf{RN}^{\mathsf{pri}} \times \mathsf{RN}^{\mathsf{pri}} \\ & \text{and } \mathfrak{s}^{\mathfrak{r}} \text{ is defined on } (x,y) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Similarly, a *total role discount factor function* is presented in the following in term of a function $\hat{\mathfrak{d}} : \mathsf{RN} \to [0,1]$ based on role discount factor.

$$\hat{\mathfrak{d}}(x) = \begin{cases} \mathfrak{d}(x) & \text{if } \mathfrak{d} \text{ is defined on } x \\ 0.4 & \text{otherwise} \end{cases} \tag{10}$$

Let us note that the default value of Equation 8 - 10 is set according to the default preference profile π_0 (Definition 2.2).

Let $C \in \mathsf{Con}(\mathcal{ELH})$ be a fully expanded concept to the form:

$$P_1 \sqcap \cdots \sqcap P_m \sqcap \exists r_1.C_1 \sqcap \cdots \sqcap \exists r_n.C_n$$

where $P_i \in \mathsf{CN}^{\mathsf{pri}}$, $r_j \in \mathsf{RN}$, $C_j \in \mathsf{Con}(\mathcal{ELH})$ in the same format, $1 \leq i \leq m$, and $1 \leq j \leq n$. The set $P_1, \ldots, P_m$ and the set $\exists r_1.C_1, \ldots, \exists r_n.C_n$ are denoted by $\mathcal{P}_C$ and $\mathcal{E}_C$, respectively. An $\mathcal{ELH}$ concept description can be structurally transformed into the corresponding $\mathcal{ELH}$ description tree. The root v_0 of the $\mathcal{ELH}$ description tree $\mathcal{T}_C$ has $\{P_1, \ldots, P_m\}$ as its label and has n outgoing edges, each labeled with r_j to a vertex v_j for $1 \leq j \leq n$. Then, a subtree with the root v_j is defined recursively relative to the concept C_j. Let $\pi = \langle \mathfrak{i}^{\mathfrak{c}}, \mathfrak{i}^{\mathfrak{r}}, \mathfrak{s}^{\mathfrak{c}}, \mathfrak{s}^{\mathfrak{r}}, \mathfrak{d} \rangle$ be a preference profile. The homomorphism degree under preference profile π can be formally defined as follows:

Definition 4.1 ((Racharak et al., 2016b)). Let $\mathbf{T}^{\mathcal{ELH}}$ be a set of all $\mathcal{ELH}$ description trees and $\mathcal{T}_C, \mathcal{T}_D \in \mathbf{T}^{\mathcal{ELH}}$ corresponds to two $\mathcal{ELH}$ concept names C and D, respectively. The *homomorphism degree under preference profile* π is a function $\mathsf{hd}^\pi : \mathbf{T}^{\mathcal{ELH}} \times \mathbf{T}^{\mathcal{ELH}} \to [0,1]$ defined inductively as follows:

$$\begin{aligned} \mathsf{hd}^\pi(\mathcal{T}_D, \mathcal{T}_C) = {} & \mu^\pi \cdot \mathsf{p\text{-}hd}^\pi(\mathcal{P}_D, \mathcal{P}_C) \\ & + (1 - \mu^\pi) \cdot \mathsf{e\text{-}set\text{-}hd}^\pi(\mathcal{E}_D, \mathcal{E}_C), \end{aligned} \tag{11}$$

where

$$\mu^\pi = \begin{cases} 1 & \text{if } \sum\limits_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A) \\ & \text{and } \sum\limits_{\exists r.X \in \mathcal{E}_D} \hat{\mathfrak{i}}(r) = 0 \\ \dfrac{\sum_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A)}{\sum_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A) + \sum_{\exists r.X \in \mathcal{E}_D} \hat{\mathfrak{i}}(r)} & \text{otherwise;} \end{cases} \tag{12}$$

$$\mathsf{p\text{-}hd}^\pi(\mathcal{P}_D, \mathcal{P}_C) = \begin{cases} 1 & \text{if } \sum\limits_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A) = 0 \\ 0 & \text{if } \sum\limits_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A) \neq 0 \text{ and } \sum\limits_{B \in \mathcal{P}_C} \hat{\mathfrak{i}}(B) = 0 \\ \mathsf{p}^{\pi *}(\mathcal{P}_D, \mathcal{P}_C) & \text{otherwise,} \end{cases} \tag{13}$$

where

$$\mathsf{p}^{\pi *}(\mathcal{P}_D, \mathcal{P}_C) = \frac{\sum_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A) \cdot \max\{\hat{\mathfrak{s}}(A,B) : B \in \mathcal{P}_C\}}{\sum_{A \in \mathcal{P}_D} \hat{\mathfrak{i}}(A)}; \tag{14}$$

$$\mathsf{e\text{-}set\text{-}hd}^\pi(\mathcal{E}_D, \mathcal{E}_C) = \begin{cases} 1 & \text{if } \sum\limits_{\exists r.X \in \mathcal{E}_D} \hat{\mathfrak{i}}(r) = 0 \\ 0 & \text{if } \sum\limits_{\exists r.X \in \mathcal{E}_D} \hat{\mathfrak{i}}(r) \neq 0 \\ & \text{and } \sum\limits_{\exists s.Y \in \mathcal{E}_C} \hat{\mathfrak{i}}(s) = 0 \\ \mathsf{e}^{\pi *}(\mathcal{E}_D, \mathcal{E}_C) & \text{otherwise,} \end{cases} \tag{15}$$

where

$$\mathsf{e}^{\pi *}(\mathcal{E}_D, \mathcal{E}_C) = \frac{\sum_{\exists r.X \in \mathcal{E}_D} \hat{\mathfrak{i}}(r) \cdot \max\{\mathsf{e\text{-}hd}^\pi(\exists r.X, \varepsilon_j) : \varepsilon_j \in \mathcal{E}_C\}}{\sum_{\exists r.X \in \mathcal{E}_D} \hat{\mathfrak{i}}(r)} \tag{16}$$

with ε_j existential restriction; and

$$\mathsf{e\text{-}hd}^\pi(\exists r.X, \exists s.Y) = \gamma^\pi(\hat{\mathfrak{d}}(r) + (1 - \hat{\mathfrak{d}}(r)) \cdot \mathsf{hd}^\pi(\mathcal{T}_X, \mathcal{T}_Y)) \tag{17}$$

where $\gamma^\pi =$

$$\begin{cases} 1 & \text{if } \sum\limits_{r' \in \mathcal{R}_r} \hat{\mathfrak{i}}(r') = 0 \\ \dfrac{\sum_{r' \in \mathcal{R}_r} \hat{\mathfrak{i}}(r') \cdot \max\{\hat{\mathfrak{s}}(r',s') : s' \in \mathcal{R}_s\}}{\sum_{r' \in \mathcal{R}_r} \hat{\mathfrak{i}}(r')}, & \text{otherwise.} \end{cases} \tag{18}$$

Intuitively, Equation 11 is defined as the weighted sum of the degree under π of primitive concepts and the degree under π of matching edges. Equation 12 indicates the weight of primitive concept names w.r.t. the importance function. Equation 13 calculates the proportion of best similarity between primitive concept names. Similarly, Equation 15 calculates the proportion of best similarity between existential information from Equation 17 and Equation 18. Equation 17 calculates the degree of similarity between matching edges. Finally, Equation 18 calculates the proportion of best similarity between role names.

Let C and D be fully expanded $\mathcal{ELH}$ concept names, $\mathcal{T}_C$ and $\mathcal{T}_D$ be the corresponding description trees, and $\pi = \langle \mathfrak{i}^c, \mathfrak{i}^r, \mathfrak{s}^c, \mathfrak{s}^r, \mathfrak{d} \rangle$ be a preference profile. The following definition formally describes the *$\mathcal{ELH}$ similarity degree under preference profile* π.

Definition 4.2 ((Racharak et al., 2016b)). The *$\mathcal{ELH}$ similarity degree under preference profile* π between C and D (denoted by $\mathsf{sim}^\pi(C,D)$) is defined as follows:

$$\mathsf{sim}^\pi(C,D) = \frac{\mathsf{hd}^\pi(\mathcal{T}_C, \mathcal{T}_D) + \mathsf{hd}^\pi(\mathcal{T}_D, \mathcal{T}_C)}{2} \qquad (19)$$

Intuitively, the degree of similarity under preference profile of two concepts is the average of the degree of having homomorphisms under preference profile in both directions. We note that $\mathsf{sim}^\pi(C,D) = 1$ indicates that C and D are total similarity under a particular π and $\mathsf{sim}^\pi(C,D) = 0$ indicates total dissimilarity between C and D under a particular π.

4.1 A Simple Methodology

Now, we are ready to discuss a simple methodology for reconciling the previously proposed strategies with the measure sim^π (Definition 4.2). Section 3 shows that each aspect of preference profile may need different strategies for tuning. For example, $\mathfrak{i}^c$ may be tuned via a defined concept name using a propagation for primitive concept importance $\mathfrak{I}^c$ whereas $\mathfrak{i}^r$ may be tuned via a more general role name using a propagation for role importance $\mathfrak{I}^r$. Also, both $\mathfrak{s}^c$ and $\mathfrak{s}^r$ may employ the canonical interpretation $I_\mathcal{A}$ to initialize values. In the following, we explain procedural steps that the readers may follow to tune preference profile for their use. These steps also hint a system flow of similarity-based under the agent's profile applications, such as the best matching concept under the agent's profile application.

1. An agent may start with tuning each aspect of preference profile individually;
2. To help tuning $\mathfrak{i}^c$, a system may present the concept hierarchy w.r.t. an ontology. Then, an agent indirectly specify primitive concept names via a defined concept name depicted on the hierarchy with the notion $\mathfrak{I}^c$ (cf. Definition 3.1 and Definition 3.2). Some patterns of an agent's utterance may be associated with certain operators, e.g.
   (a) We may associate *A especially B* with $\oplus_{first}$, where $depth(A) < depth(B)$ and $depth(X)$ is the depth of X on the concept hierarchy;
   (b) We may associate *A except B* with $\oplus_{first}$, where $depth(A) < depth(B)$ and $C_1 \neq \top$;
   (c) We may associate $\top$ *except B* with $\oplus_{last}$; and
   (d) Otherwise, the agent-defined default concrete operator is used;
3. To help tuning $\mathfrak{i}^r$, a system may present the role hierarchy w.r.t. an ontology. Then, an agent indirectly specify role names via a more general role name name depicted on the hierarchy with the noion $\mathfrak{I}^r$ (cf. Definition 3.3 and Definition 3.4). Similarly, some patterns of an agent's utterance may be associated with certain operators, e.g.
   (a) We may associate *r especially s* with $\oplus_{last}$, where $r \in \mathcal{R}_s$;
   (b) We may associate *r except s* with $\oplus_{last}$, where $r \in \mathcal{R}_s$; and
   (c) Otherwise, the agent-defined default concrete operator is used;
4. To help tuning $\mathfrak{s}^c$ and $\mathfrak{s}^r$, a system may construct the canonical interpretation, which is induced from $\mathcal{A}$. Then, each initial value for all possible primitive concept pairs and primitive role pairs is calculated according to Equation 6 and Equation 7, respectively;
5. An agent may refine the agent's preference profile if that agent wishes.

We exemplify the methodology in its applicable use cases, such as trip planning (Example 4.1).

Example 4.1. (Continuation from Example 1.1) We expand each definition in $\mathcal{T}$ as follows:

$$\begin{array}{rcl} \mathsf{ActivePlace} & \equiv & X \sqcap \mathsf{Place} \sqcap \exists \mathsf{canWalk.Trekking} \\ & & \sqcap \exists \mathsf{canSail.Kayaking} \\ \mathsf{Mangrove} & \equiv & Y \sqcap \mathsf{Place} \sqcap \exists \mathsf{canWalk.Trekking} \\ \mathsf{Beach} & \equiv & Z \sqcap \mathsf{Place} \sqcap \exists \mathsf{canSail.Kayaking} \end{array}$$

where X, Y, and Z are fresh primitive concept names. Furthermore, $\mathcal{R}_{\mathsf{canWalk}} = \{t, \mathsf{cMWL}\}$[6] and $\mathcal{R}_{\mathsf{canSail}} = \{t, \mathsf{cTWS}\}$ where t and u are also fresh primitive role names.

Let an ABox $\mathcal{A}$ be defined as follows:

$$\begin{array}{rl} \mathsf{cMWL}(p_2,t_1) & \mathsf{cTWS}(p_3,k_1) \\ \mathsf{canWalk}(p_2,t_1) & \mathsf{canSail}(p_3,k_1) \\ \mathsf{Trekking}(t_1) & \mathsf{Kayaking}(k_1) \\ \mathsf{Place}(p_1) & \mathsf{Place}(p_2) \\ \mathsf{Place}(p_3) & \mathsf{Mangrove}(p_2) \\ \mathsf{Beach}(p_3) & \end{array}$$

To query for a desired place, an agent needs to express his preferences. Suppose the agent says

[6]Obvious abbreviations are used here for the sake of succinctness.

My interest is a place where I can travel with by feet, especially walking, i.e. $\mathfrak{i}^{\mathfrak{c}}(\mathsf{Place}) = 1.5$, $\mathfrak{I}^{\mathfrak{r}}(\mathsf{canMoveWithLegs}) = 1.5$, and $\mathfrak{I}^{\mathfrak{r}}(\mathsf{canWalk}) = 1.8$. Also, it yields $\mathfrak{i}^{\mathfrak{r}}(\mathit{canMoveWithLegs}) = 1.5$ and $\mathfrak{i}^{\mathfrak{r}}(\mathsf{canWalk}) = 1.5 \oplus_{last} 1.8 = 1.8$.

Constructing the canonical interpretation from $\mathcal{A}$, we obtain $\Delta^{I_{\mathcal{A}}} = \{p_1, p_2, p_3, t_1, k_1\}$, $\mathfrak{s}^{\mathfrak{c}}(\mathsf{Trekking}, \mathsf{Kayaking}) = 0$, and $\mathfrak{s}^{\mathfrak{r}}(\mathsf{canMoveWithLegs}, \mathsf{canTravelWithSails}) = 0$.

Let ActivePlace, Mangrove, Place, Trekking, Kayaking, canWalk, and canSail are rewritten shortly as AP, M, P, T, K, cW, and cS, respectively. Using Definition 4.1, $\mathsf{hd}^{\pi}(\mathcal{T}_{\mathsf{AP}}, \mathcal{T}_{\mathsf{M}})$

$$
\begin{aligned}
&= (\tfrac{2.5}{5.3}) \cdot \mathsf{p\text{-}hd}^{\pi}(\mathcal{P}_{\mathsf{AP}}, \mathcal{P}_{\mathsf{M}}) + (\tfrac{2.8}{5.3}) \cdot \mathsf{e\text{-}set\text{-}hd}^{\pi}(\mathcal{E}_{\mathsf{AP}}, \mathcal{E}_{\mathsf{M}}) \\
&= (\tfrac{2.5}{5.3}) \cdot (\tfrac{\mathfrak{i}(X)\cdot\max\{\mathfrak{s}(X,Y),\mathfrak{s}(X,\mathsf{P})\}+\mathfrak{i}(\mathsf{P})\cdot\max\{\mathfrak{s}(\mathsf{P},\mathsf{Y}),\mathfrak{s}(\mathsf{P},\mathsf{P})\})}{\mathfrak{i}(X)+\mathfrak{i}(\mathsf{P})}) \\
&+ (\tfrac{2.8}{5.3}) \cdot \mathsf{e\text{-}set\text{-}hd}^{\pi}(\mathcal{E}_{\mathsf{AP}}, \mathcal{E}_{\mathsf{M}}) \\
&= (\tfrac{2.5}{5.3})(\tfrac{1\cdot\max\{0,0\}+1.5\cdot\max\{0,1\}}{1+1.5}) \\
&+ (\tfrac{2.8}{5.3}) \cdot \mathsf{e\text{-}set\text{-}hd}^{\pi}(\mathcal{E}_{\mathsf{AP}}, \mathcal{E}_{\mathsf{M}}) \\
&= (\tfrac{2.5}{5.3})(\tfrac{1.5}{2.5}) + (\tfrac{2.8}{5.3}) \left[\tfrac{\mathfrak{i}(\mathsf{cW})\cdot\max\{\mathsf{e\text{-}hd}^{\pi}(\exists\mathsf{cW}.\mathsf{T},\exists\mathsf{cW},\mathsf{T})\}+1\cdot 0}{\mathfrak{i}(\mathsf{cW})+\mathfrak{i}(\mathsf{cS})}\right] \\
&= (\tfrac{2.5}{5.3})(\tfrac{1.5}{5.3}) + (\tfrac{2.8}{5.3}) \left[\tfrac{1.8\cdot 1+1\cdot 0}{1+1.8}\right] \approx 0.623
\end{aligned}
$$

Following the same step, we obtain $\mathsf{hd}^{\pi}(\mathcal{T}_{\mathsf{M}}, \mathcal{T}_{\mathsf{AP}}) \approx 0.767$. Hence, $\mathsf{sim}^{\pi}(\mathsf{M}, \mathsf{AP}) \approx 0.695$ by using Definition 4.2. Also, we obtain $\mathsf{hd}^{\pi}(\mathcal{T}_{\mathsf{AP}}, \mathcal{T}_{\mathsf{B}}) \approx 0.472$ and $\mathsf{hd}^{\pi}(\mathcal{T}_{\mathsf{B}}, \mathcal{T}_{\mathsf{AP}}) \approx 0.714$. Hence, $\mathsf{sim}^{\pi}(\mathsf{B}, \mathsf{AP}) \approx 0.593$.

The fact that $\mathsf{sim}^{\pi}(\mathsf{M}, \mathsf{AP}) > \mathsf{sim}^{\pi}(\mathsf{B}, \mathsf{AP})$ corresponds to the agent's perception.

□

5 RELATIONSHIP TO LEARNING-BASED APPROACH

Our proposed development uses the canonical interpretation $I_{\mathcal{A}}$ to compute numerical values for mappings on $\mathfrak{s}^{\mathfrak{c}}$ (cf. Subsection 3.3) and $\mathfrak{s}^{\mathfrak{r}}$ (cf. Subsection 3.4). Its drawback is that an existence of the canonical interpretation $I_{\mathcal{A}}$ is required. This section rather discusses an alternative approach to obtain values for mappings on $\mathfrak{s}^{\mathfrak{c}}$ and $\mathfrak{s}^{\mathfrak{r}}$.

In addition to our proposed logic-based approach, another natural way to configure both $\mathfrak{s}^{\mathfrak{c}}$ and $\mathfrak{s}^{\mathfrak{r}}$ is to employ existing machine learning techniques on a large corpus. For example, one may use Word2vec (Mikolov et al., 2013) with a large corpus of text to produce a vector space. Each word in the corpus will be assigned by a corresponding vector in the space. Word vectors are positioned in the vector space such that words sharing common features in the corpus are located in close proximity to one another in the space. This characterization can later be converted into elements of the mapping $\mathfrak{s}^{\mathfrak{c}}$ and $\mathfrak{s}^{\mathfrak{r}}$. Reconciling an ontology with machine learning techniques to improve an application of $\overset{\pi}{\sim}_{\mathcal{T}}$ is interesting but is outside the scope of this paper. We leave this as a future task.

6 RELATED WORK

While there has been substantial work on concept similarity measures in the context of DLs, the topic of concept similarity measure under an agent's preferences remains relatively unaddressed. Notable exceptions include (Tongphu and Suntisrivaraporn, 2015; Lehmann and Turhan, 2012); however, measures presented in these papers may not include preferential aspects evidently in the formal definition. We discuss the differences of ours to others in the following.

As similarity may be subjective, the techniques involved in concept similarity measure can be classified into two main classes: ones which address the problem under an agent's preferences, e.g. sim^{π} (originally introduced in (Racharak et al., 2016b)), and ones which do not, e.g. (Janowicz and Wilkes, 2009; Racharak and Suntisrivaraporn, 2015; D'Amato et al., 2006; Fanizzi and D'Amato, 2006; D'Amato et al., 2009; D'Amato et al., 2008). The measure sim^{π} generalizes the notion of homomorphism structural subsumption with an aim to develop a similarity measure under preference profile for DL $\mathcal{ELH}$. As previously mentioned, (Tongphu and Suntisrivaraporn, 2015; Lehmann and Turhan, 2012) may not include preferential aspects evidently in the formal definition; however, their approaches share some viewpoints in common to preference profile. For instance, (Tongphu and Suntisrivaraporn, 2015) provides some facilities similar to $\mathfrak{i}^{\mathfrak{c}}$ and $\mathfrak{d}$ whereas (Lehmann and Turhan, 2012) provides some facilities similar to $\mathfrak{i}^{\mathfrak{c}}$, $\mathfrak{s}^{\mathfrak{c}}$, and $\mathfrak{s}^{\mathfrak{r}}$. We refer the readers to (Racharak et al., 2016a; Racharak et al., 2016b) for the detailed discussion.

Speaking out in the context of DLs, we may classify techniques in the other way round, i.e. structure-based measure and interpretation-based measure. Structure-based measure, e.g. (Janowicz and Wilkes, 2009; Racharak and Suntisrivaraporn, 2015; D'Amato et al., 2006; Fanizzi and D'Amato, 2006; Tongphu and Suntisrivaraporn, 2015; Lehmann and Turhan, 2012; Racharak et al., 2016b), is defined using the syntax of concepts being measured. On the other hand, interpretation-based measure, e.g. (D'Amato et al., 2009; D'Amato et al., 2008), is defined using interpretations and cardinality. Some measures also include elements of both, e.g. (D'Amato et al., 2006; Fanizzi and D'Amato, 2006) use structure to measure concepts but use the canonical interpretation $I_{\mathcal{A}}$ to measure similarity of

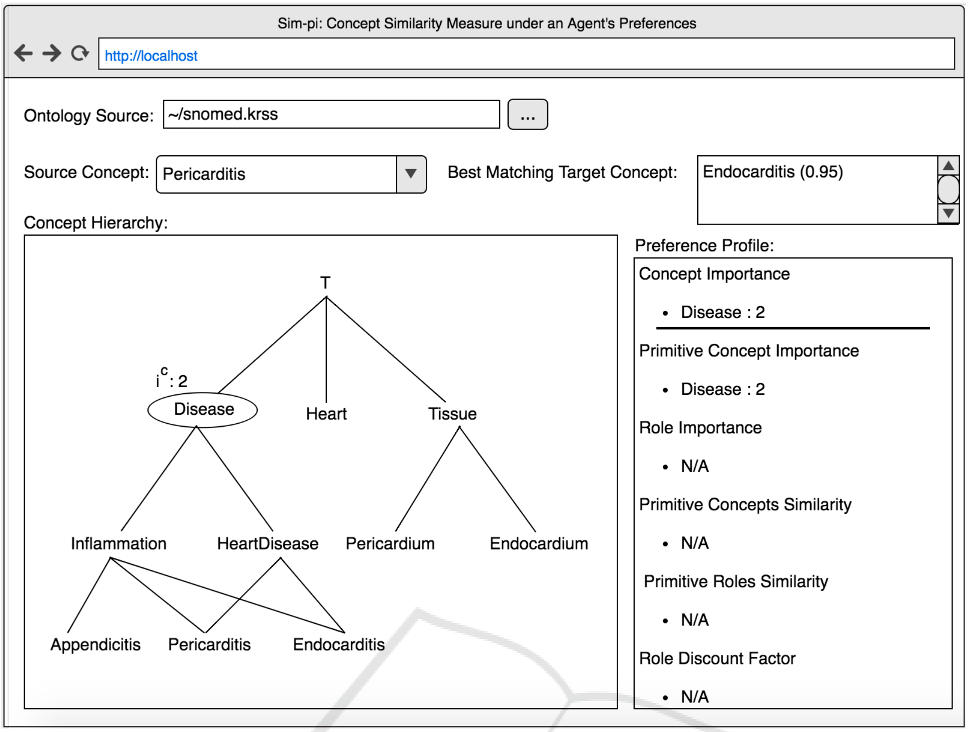

Figure 3: A mock-up of the best-matching concepts under the agent's profile application.

primitive concept names. The measure sim^π per se is categorized as a structure-based measure; however, reconciling sim^π with our strategies to help tuning $\mathfrak{s}^{\mathfrak{c}}$ and $\mathfrak{s}^{\mathfrak{r}}$ may be considered as the last category as it uses the canonical interpretation $I_{\mathcal{A}}$ for measuring similarity of primitive concept names and similarity of primitive role names.

This work is an extension of the original sim^π (Racharak et al., 2016b) in which we investigate and formalize the task of configuring the preference functions based on the information defined in a TBox and an ABox (cf. Section 3). We also show that the proposed strategies can be reconciled with the original development, i.e. sim^π (cf. Section 4).

7 DISCUSSION AND FUTURE RESEARCH

As realistic ontologies are generally complex – consisting in plenty of concept names and role names, having some strategies to tune a measure helps ontology engineers, researchers, and application users to use a measure for similarity-based under the agent's profile applications. That is, instead of specifying each aspect of preference profile individually and manually, an agent may automatically assign an importance of each primitive concept name through a defined concept name with $\mathfrak{I}^{\mathfrak{c}}$. Similarly, an agent may automatically assign an importance of each role name through a more general role name with $\mathfrak{I}^{\mathfrak{r}}$. However, these notions have some restrictions and we discuss these in Subsection 3.1 and Subsection 3.2, respectively. If an ABox is presented, the canonical interpretation can be induced and be used to further compute $\mathfrak{s}^{\mathfrak{c}}$ and $\mathfrak{s}^{\mathfrak{r}}$ for each primitive concept pair and primitive role pair (cf. Subsection 3.3 and Subsection 3.4). In any cases, these strategies are recommended to use for the initial preference tuning and these values may be refined wherever the agent wishes.

We also present a simple guideline to implement similarity-based under the agent's profile applications (cf. Subsection 4.1). As exhibited by the guideline, a system may depict the concept hierarchy and the role hierarchy w.r.t. an ontology to let an agent fine tune via $\mathfrak{I}^{\mathfrak{c}}$ and $\mathfrak{I}^{\mathfrak{r}}$, respectively. Figure 3 shows a mock-up of the best-matching concept under the agent's profile application. In this mock-up, we permit an agent to identify his preferences through the notion $\mathfrak{I}^{\mathfrak{c}}$ by highlighting concept names occurring on the hierarchy. We note that this mock-up does not show all the strategies we have discussed in this work. Also, it uses $\oplus_{max}$ to handle conflicting values on $\mathfrak{i}^{\mathfrak{c}}$.

Currently, we are under an implementation of the best-matching application under the agent's profile using our extended measure sim^π. There are several possible directions for the theoretical future research.

Firstly, our current strategies cannot be used to express complex preferences, such as multi-dimensional preferences. Hence, it appears to be a natural step to develop a high-level language for the specification of an agent's preferences in the context of similarity-based problems. Secondly, we are interested to extend the notion of preference profile to support a more expressive DL family, e.g. universal restriction, concept negation, and also, to support an ABox. Thirdly, we intend to devise a concept similarity measure under preference profile which can handle more expressive DLs. Finally, we intend to employ our developed notion of concept similarity measure under preference profile toward a system of analogical reasoning. As we have developed an argument-based logic programming for analogical reasoning in (Racharak et al., 2016c), it would be interesting to connect these two research studies.

ACKNOWLEDGEMENTS

The authors would like to thank Prachya Boonkwan for his proofreading and useful comments. This research is part of the JAIST-NECTEC-SIIT dual doctoral degree program.

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29.

Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F. (2010). *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, New York, NY, USA, 2nd edition.

D'Amato, C., Fanizzi, N., and Esposito, F. (2006). A dissimilarity measure for alc concept descriptions. *In Proceedings of the 2006 ACM Symposium on Applied Computing*, pages 1695–1699.

D'Amato, C., Fanizzi, N., and Esposito, F. (2009). A semantic similarity measure for expressive description logics. *In CoRR*, abs/0911.5043.

D'Amato, C., Staab, S., and Fanizzi, N. (2008). On the influence of description logics ontologies on conceptual similarity. *In Proceedings of Knowledge Engineering: Practice and Patterns*, pages 48–63.

Euzenat, J. and Valtchev, P. (2004). Similarity-based ontology alignment in OWL-lite. In de Mántaras, R. L. and Saitta, L., editors, *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-04)*, pages 333–337. IOS Press.

Fanizzi, N. and D'Amato, C. (2006). A similarity measure for the aln description logic. *In Proceedings of CILC 2006 - Italian Conference on Computational Logic*, pages 26–27.

Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., and Sattler, U. (2008). Owl 2: The next step for owl. *Web Semant.*, 6(4):309–322.

Group, W. O. W. (2012). OWL 2 web ontology language. document overview (second edition). W3C recommendation, W3C.

Janowicz, K. and Wilkes, M. (2009). Sim-dla: A novel semantic similarity measure for description logics reducing inter-concept to inter-instance similarity. *In Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, pages 353–367.

Lehmann, K. and Turhan, A.-Y. (2012). A framework for semantic-based similarity measures for elh-concepts. In del Cerro, L. F., Herzig, A., and Mengin, J., editors, *JELIA*, volume 7519 of *Lecture Notes in Computer Science*, pages 307–319. Springer.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Racharak, T. and Suntisrivaraporn, B. (2015). Similarity measures for $\mathcal{FL}_0$ concept descriptions from an automata-theoretic point of view. In *Information and Communication Technology for Embedded Systems (IC-ICTES), 2015 6th International Conference of*, pages 1–6.

Racharak, T., Suntisrivaraporn, B., and Tojo, S. (2016a). *Identifying an Agent's Preferences Toward Similarity Measures in Description Logics*, pages 201–208. Springer International Publishing, Cham.

Racharak, T., Suntisrivaraporn, B., and Tojo, S. (2016b). sim^{π}: A concept similarity measure under an agent's preferences in description logic $\mathcal{ELH}$. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, pages 480–487.

Racharak, T., Tojo, S., Hung, N. D., and Boonkwan, P. (2016c). Argument-based logic programming for analogical reasoning. In *Proceeding of Tenth International Workshop on Juris-informatics (JURISIN 2016)*.

Stearns, M. Q., Price, C., Spackman, K. A., and Wang, A. Y. (2001). SNOMED clinical terms: overview of the development process and project status. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 662–666.

Tongphu, S. and Suntisrivaraporn, B. (2015). 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)*.