# Adding Threshold Concepts to the Description Logic $\mathcal{EL}$

Franz Baader[1], Gerhard Brewka[2], and Oliver Fernández Gil[2,⋆]

[1] Theoretical Computer Science, TU Dresden, Germany
baader@tcs.inf.tu-dresden.de
[2] Department of Computer Science, University of Leipzig, Germany
{brewka,fernandez}@informatik.uni-leipzig.de

**Abstract.** We introduce an extension of the lightweight Description Logic $\mathcal{EL}$ that allows us to define concepts in an approximate way. For this purpose, we use a graded membership function, which for each individual and concept yields a number in the interval $[0,1]$ expressing the degree to which the individual belongs to the concept. Threshold concepts $C_{\sim t}$ for $\sim \in \{<, \leq, >, \geq\}$ then collect all the individuals that belong to $C$ with degree $\sim t$. We generalize a well-known characterization of membership in $\mathcal{EL}$ concepts to construct a specific graded membership function *deg*, and investigate the complexity of reasoning in the Description Logic $\tau\mathcal{EL}(deg)$, which extends $\mathcal{EL}$ by threshold concepts defined using *deg*. We also compare the instance problem for threshold concepts of the form $C_{>t}$ in $\tau\mathcal{EL}(deg)$ with the relaxed instance queries of Ecke et al.

## 1 Introduction

Description logics (DLs) [3] are a family of logic-based knowledge representation formalisms, which can be used to represent the conceptual knowledge of an application domain in a structured and formally well-understood way. They allow their users to define the important notions of the domain as concepts by stating necessary and sufficient conditions for an individual to belong to the concept. These conditions can be atomic properties required for the individual (expressed by concept names) or properties that refer to relationships with other individuals and their properties (expressed as role restrictions). The expressivity of a particular DL is determined by what sort of properties can be required and how they can be combined.

The DL $\mathcal{EL}$, in which concepts can be built using concept names as well as the concept constructors conjunction ($\sqcap$), existential restriction ($\exists r.C$), and the top concept ($\top$), has drawn considerable attention in the last decade since, on the one hand, important inference problems such as the subsumption problem are polynomial in $\mathcal{EL}$, even with respect to expressive terminological axioms [7]. On the other hand, though quite inexpressive, $\mathcal{EL}$ can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT.[1] In $\mathcal{EL}$ we can, for

⋆ Supported by DFG Graduiertenkolleg 1763 (QuantLA).
[1] see http://www.ihtsdo.org/snomed-ct/


C. Lutz and S. Ranise (Eds.): FroCoS 2015, LNAI 9322, pp. 33–48, 2015.
DOI: 10.1007/978-3-319-24246-0_3

example, define the concept of a *happy man* as a male human that is healthy and handsome, has a rich and intelligent wife, a son and a daughter, and a friend:

$$\begin{array}{c}\mathsf{Human} \sqcap \mathsf{Male} \sqcap \mathsf{Healthy} \sqcap \mathsf{Handsome} \sqcap \\ \exists \mathsf{spouse}.(\mathsf{Rich} \sqcap \mathsf{Intelligent} \sqcap \mathsf{Female}) \sqcap \\ \exists \mathsf{child}.\mathsf{Male} \sqcap \exists \mathsf{child}.\mathsf{Female} \sqcap \exists \mathsf{friend}.\top\end{array} \tag{1}$$

For an individual to belong to this concept, all the stated properties need to be satisfied. However, maybe we would still want to call a man happy if most, though not all, of the properties hold. It might be sufficient to have just a daughter without a son, or a wife that is only intelligent but not rich, or maybe an intelligent and rich spouse of a different gender. But still, not too many of the properties should be violated.

In this paper, we introduce a DL extending $\mathcal{EL}$ that allows us to define concepts in such an approximate way. The main idea is to use a *graded membership function*, which instead of a Boolean membership value 0 or 1 yields a membership degree from the interval $[0, 1]$. We can then require a happy man to belong to the $\mathcal{EL}$ concept (1) with degree at least .8. More generally, if $C$ is an $\mathcal{EL}$ concept, then the *threshold concept* $C_{\geq t}$ for $t \in [0, 1]$ collects all the individuals that belong to $C$ with degree at least $t$. In addition to such upper threshold concepts, we will also consider lower threshold concepts $C_{\leq t}$ and allow the use of strict inequalities in both. For example, an unhappy man could be required to belong to the $\mathcal{EL}$ concept (1) with a degree less than .2.

The use of membership degree functions with values in the interval $[0, 1]$ may remind the reader of fuzzy logics. However, there is no strong relationship between this work and the work on fuzzy DLs [6] for two reasons. First, in fuzzy DLs the semantics is extended to fuzzy interpretations where concept and role names are interpreted as fuzzy sets and relations, respectively. The membership degree of an individual to belong to a complex concept is then computed using fuzzy interpretations of the concept constructors (e.g., conjunction is interpreted using an appropriate triangular norm). In our setting, we consider crisp interpretations of concept and role names, and directly define membership degrees for complex concepts based on them. Second, we use membership degrees to obtain new concept constructors, but the threshold concepts obtained by applying these constructors are again crisp rather than fuzzy.

In the next section, we will formally introduce the DL $\mathcal{EL}$, and then recall the well-known characterization of element-hood in $\mathcal{EL}$ concepts via existence of homomorphisms between $\mathcal{EL}$ description graphs (which can express both $\mathcal{EL}$ concepts and interpretations in a graphical way). In Section 3, we then extend $\mathcal{EL}$ by new threshold concept constructors, which are based on an arbitrary, but fixed graded membership function. We will impose some minimal requirements on such membership functions, and show the consequences that these conditions have for our threshold logic. In Section 4, we then introduce a specific graded membership function *deg*, which satisfies the requirements from the previous sections. Its definition is a natural extension of the homomorphism characterization of crisp membership in $\mathcal{EL}$. Basically, an individual is punished (in the sense that its membership degree is lowered) for each missing property in a uniform way. More sophisticated versions of this function, which weigh the absence of different properties in a different way, may be useful in practice. However, they are easy to define and considering them would only add clutter, but no new insights, to our investigation (in Section 5) of the computational properties of the threshold logic obtained by using this function.

In Section 6 we compare our graded membership function with similarity measures on $\mathcal{EL}$ concepts. In fact, from a technical point of view, the graded membership function introduced in Section 4 is akin to the similarity measures for $\mathcal{EL}$ concepts introduced in [14,15], though only [15] directly draws its inspirations from the homomorphism characterization of subsumption in $\mathcal{EL}$. We show that a variant of the relaxed instance query approach of [10] can be used to turn a similarity measure into a graded membership function. It turns out that, applied to a simple instance $\bowtie^1$ of the framework for constructing similarity measures in [14], this approach actually yields our membership function *deg*. In addition, we can show that the relaxed instance queries of [14] can be expressed as instance queries w.r.t. threshold concepts of the form $C_{>t}$. However, the new DL introduced in this paper is considerably more expressive than just such threshold concepts since we also allow the use of comparison operators other than $>$ in threshold concepts, and the threshold concepts can be embedded in complex $\mathcal{EL}$ concepts.

This paper is an extended version of [16]. Due to the space constraints, we cannot provide all technical details and proofs in this paper. They can be found in the technical report [1].

## 2 The Description Logic $\mathcal{EL}$

We start by defining syntax and semantics of $\mathcal{EL}$. Starting with finite sets of concept names $\mathsf{N_C}$ and role names $\mathsf{N_R}$, the set $\mathcal{C}_{\mathcal{EL}}$ of $\mathcal{EL}$ concept descriptions is obtained by using the concept constructors *conjunction* $(C \sqcap D)$, *existential restriction* $(\exists r.C)$ and *top* $(\top)$, in the following way:

$$C ::= \top \mid A \mid C \sqcap C \mid \exists r.C$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $C \in \mathcal{C}_{\mathcal{EL}}$.

An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty domain $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$ that assigns subsets of $\Delta^{\mathcal{I}}$ to each concept name and binary relations over $\Delta^{\mathcal{I}}$ to each role name. The interpretation function $\cdot^{\mathcal{I}}$ is inductively extended to concept descriptions in the usual way:

$$\top^{\mathcal{I}} := \Delta^{\mathcal{I}},$$
$$(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}},$$
$$(\exists r.C)^{\mathcal{I}} := \{x \in \Delta^{\mathcal{I}} \mid \exists y.\, (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}.$$

Given $C, D \in \mathcal{C}_{\mathcal{EL}}$, we say that $C$ is *subsumed by* $D$ (denoted as $C \sqsubseteq D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for every interpretation $\mathcal{I}$. These two concept descriptions are *equivalent* (denoted as $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$. Finally, $C$ is *satisfiable* iff $C^{\mathcal{I}} \neq \emptyset$ for some interpretation $\mathcal{I}$.

Our definition of graded membership will be based on graphical representations of concepts and interpretations, and on homomorphisms between such representations. For this reason, we recall these notions together with the pertinent results. They are all taken from [4,12,2].

**Definition 1 ($\mathcal{EL}$ Description Graphs).** *An $\mathcal{EL}$* description graph *is of the form $G = (V_G, E_G, \ell_G)$ where:*

- *$V_G$ is a set of nodes.*
- *$E_G \subseteq V_G \times \mathsf{N_R} \times V_G$ is a set of edges labelled by role names,*
- *$\ell_G : V_G \rightarrow 2^{\mathsf{N_C}}$ is a function that labels nodes with sets of concept names.*

An $\mathcal{EL}$ *description tree* $T$ is a description graph that is a tree with a distinguished element $v_0$ representing its root. In [4], it was shown that every $\mathcal{EL}$ concept description $C$ can be translated into a corresponding description tree $T_C$ and vice versa. Furthermore, every interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ can be translated into an $\mathcal{EL}$ description graph $G_{\mathcal{I}} = (V_{\mathcal{I}}, E_{\mathcal{I}}, \ell_{\mathcal{I}})$ in the following way [2]:

- $V_{\mathcal{I}} = \Delta^{\mathcal{I}}$,
- $E_{\mathcal{I}} = \{(vrw) \mid (v, w) \in r^{\mathcal{I}}\}$,
- $\ell_{\mathcal{I}}(v) = \{A \mid v \in A^{\mathcal{I}}\}$ for all $v \in V_{\mathcal{I}}$.

*Example 1.* The $\mathcal{EL}$ concept description

$$C := A \sqcap \exists r.(A \sqcap B \sqcap \exists r.\top) \sqcap \exists r.A$$

yields the $\mathcal{EL}$ description tree $T_C$ depicted on the left-hand side in Figure 1. The description graph on the right-hand side corresponds to the following interpretation:

- $\Delta^{\mathcal{I}} := \{a_1, a_2, a_3\}$,
- $A^{\mathcal{I}} := \{a_1, a_2\}$ and $B^{\mathcal{I}} := \{a_2, a_3\}$,
- $r^{\mathcal{I}} := \{(a_1, a_2), (a_2, a_3), (a_3, a_1)\}$.

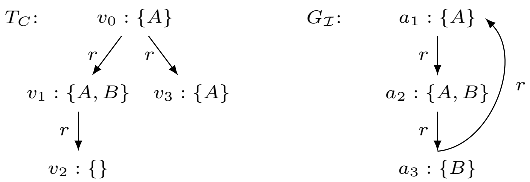


**Fig. 1.** $\mathcal{EL}$ description graphs.

Next, we generalize homomorphisms between $\mathcal{EL}$ description trees [4] to arbitrary graphs.

**Definition 2 (Homomorphisms on $\mathcal{EL}$ Description Graphs).** *Let* $G = (V_G, E_G, \ell_G)$ *and* $H = (V_H, E_H, \ell_H)$ *be two* $\mathcal{EL}$ *description graphs. A mapping* $\varphi : V_G \to V_H$ *is a* homomorphism *from* $G$ *to* $H$ *iff the following conditions are satisfied:*

1. $\ell_G(v) \subseteq \ell_H(\varphi(v))$ *for all* $v \in V_G$*, and*
2. $vrw \in E_G$ *implies* $\varphi(v)r\varphi(w) \in E_H$.

*This homomorphism is an* isomorphism *iff it is bijective, equality instead of just inclusion holds in 1., and biimplication instead of just implication holds in 2.*

In Example 1, the mapping $\varphi$ with $\varphi(v_i) = a_{i+1}$ for $i = 0, 1, 2$ and $\varphi(v_3) = a_2$ is a homomorphism. Homomorphisms between $\mathcal{EL}$ description trees can be used to characterize subsumption in $\mathcal{EL}$.

**Theorem 1 ([4]).** *Let* $C, D$ *be* $\mathcal{EL}$ *concept descriptions and* $T_C, T_D$ *the corresponding* $\mathcal{EL}$ *description trees. Then* $C \sqsubseteq D$ *iff there exists a homomorphism from* $T_D$ *to* $T_C$ *that maps the root of* $T_D$ *to the root of* $T_C$.

The proof of this result can be easily adapted to obtain a similar characterization of *element-hood* in $\mathcal{EL}$, i.e., whether $d \in C^{\mathcal{I}}$ for some $d \in \Delta^{\mathcal{I}}$.

**Theorem 2.** *Let* $\mathcal{I}$ *be an interpretation,* $d \in \Delta^{\mathcal{I}}$*, and* $C$ *an* $\mathcal{EL}$ *concept description. Then,* $d \in C^{\mathcal{I}}$ *iff there exists a homomorphism* $\varphi$ *from* $T_C$ *to* $G_{\mathcal{I}}$ *such that* $\varphi(v_0) = d$.

In Example 1, the existence of the homomorphism $\varphi$ defined above thus shows that $a_1 \in C^{\mathcal{I}}$. Equivalence of $\mathcal{EL}$ concept descriptions can be characterized via the existence of isomorphisms, but for this the concept descriptions first need to be normalized by removing redundant existential restrictions. To be more precise, the *reduced form* of an $\mathcal{EL}$ concept description is obtained by applying the rewrite rule $\exists r.C \sqcap \exists r.D \longrightarrow \exists r.C$ if $C \sqsubseteq D$ as long as possible. This rule is applied modulo associativity and commutativity of $\sqcap$, and not only on the top-level conjunction of the description, but also under the scope of existential restrictions. Since every application of the rule decreases the size of the description, it is easy to see that the reduced form can be computed in polynomial time. We say that an $\mathcal{EL}$ concept description is *reduced* iff this rule does not apply to it. In our Example 1, the reduced form of $C$ is the reduced description $A \sqcap \exists r.(A \sqcap B \sqcap \exists r.\top)$.

**Theorem 3 ([12]).** *Let* $C, D$ *be* $\mathcal{EL}$ *concept descriptions,* $C^r, D^r$ *their reduced forms, and* $T_{C^r}, T_{D^r}$ *the corresponding* $\mathcal{EL}$ *description trees. Then* $C \equiv D$ *iff there exists an isomorphism between* $T_{C^r}$ *and* $T_{D^r}$.

## 3 The Logic $\tau\mathcal{EL}(m)$

Our new logic will allow us to take an arbitrary $\mathcal{EL}$ concept $C$ and turn it into a threshold concept. To this end we introduce a family of constructors that are based on the membership degree of individuals in $C$. For instance, the threshold concept $C_{>.8}$ represents the individuals that belong to $C$ with degree $> .8$. The semantics of the new threshold concepts depends on a (graded) membership function $m$. Given an interpretation $\mathcal{I}$, this function takes a domain element $d \in \Delta^{\mathcal{I}}$ and an $\mathcal{EL}$ concept $C$ as input, and returns a value between 0 and 1, representing the extent to which $d$ belongs to $C$ in $\mathcal{I}$.

The choice of the membership function obviously has a great influence on the semantics of the threshold concepts. In Section 4 we will propose one specific such function *deg*, but we do not claim this is the only reasonable way to define such a function. Rather, the membership function is a parameter in defining the logic. To highlight this dependency, we call the logic $\tau\mathcal{EL}(m)$.

Nevertheless, membership functions are not arbitrary. There are two properties we require such functions to satisfy:

**Definition 3.** *A graded membership function $m$ is a family of functions that contains for every interpretation $\mathcal{I}$ a function $m^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \to [0,1]$ satisfying the following conditions (for $C, D \in \mathcal{C}_{\mathcal{EL}}$):*

$$\begin{aligned} &M1\ : d \in C^{\mathcal{I}} \Leftrightarrow m^{\mathcal{I}}(d, C) = 1 \quad \textit{for all } d \in \Delta^{\mathcal{I}},\\ &M2\ : C \equiv D \Leftrightarrow \textit{ for all } d \in \Delta^{\mathcal{I}} : m^{\mathcal{I}}(d, C) = m^{\mathcal{I}}(d, D). \end{aligned}$$

Property *M1* requires that the value 1 is a distinguished value reserved for proper containment in a concept. Property *M2* requires equivalence invariance. It expresses the intuition that the membership value should not depend on the syntactic form of a concept, but only on its semantics. Note that the right to left implication in *M2* already follows from *M1*.

We now turn to the syntax of $\tau\mathcal{EL}(m)$. Given finite sets of concept names $\mathsf{N_C}$ and role names $\mathsf{N_R}$, $\tau\mathcal{EL}(m)$ *concept descriptions* are defined as follows:

$$\widehat{C} ::= \top \mid A \mid \widehat{C} \sqcap \widehat{C} \mid \exists r.\widehat{C} \mid E_{\sim q}$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, $\sim\, \in \{<, \leq, >, \geq\}$, $q \in [0,1] \cap \mathbb{Q}$, $E$ is an $\mathcal{EL}$ concept description and $\widehat{C}$ is a $\tau\mathcal{EL}(m)$ concept description. Concepts of the form $E_{\sim q}$ are called *threshold concepts.*

The semantics of the new threshold concepts is defined in the following way:

$$[E_{\sim q}]^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid m^{\mathcal{I}}(d, E) \sim q\}.$$

The extension of $.^{\mathcal{I}}$ to more complex concepts is defined as in $\mathcal{EL}$ by additionally considering the semantics of the newly introduced threshold concepts.

Requiring property *M1* has the following consequences for the semantics of threshold concepts:

**Proposition 1.** *For every $\mathcal{EL}$ concept description $E$ we have*

$$E_{\geq 1} \equiv E \quad \textit{and} \quad E_{<1} \equiv \neg E,$$

*where the semantics of negation is defined as usual, i.e., $[\neg E]^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus E^{\mathcal{I}}$.*

The second equivalence basically says that $\tau\mathcal{EL}(m)$ can express negation of $\mathcal{EL}$ concept descriptions. This does not imply that $\tau\mathcal{EL}(m)$ is closed under negation since the threshold constructors can only be applied to $\mathcal{EL}$ concept descriptions. Thus, negation cannot be nested using these constructors. A formal proof that $\tau\mathcal{EL}(deg)$ for the membership function *deg* introduced in the next section cannot express full negation can be found in [1]. However, atomic negation (i.e., negation applied to concept names) can obviously be expressed. Consequently, unlike $\mathcal{EL}$ concept descriptions, not all $\tau\mathcal{EL}(m)$ concept descriptions are satisfiable. A simple example is the concept description $A_{\geq 1} \sqcap A_{<1}$, which is equivalent to $A \sqcap \neg A$.

## 4 The Membership Function *deg*

To make things more concrete, we now introduce a specific membership function, denoted *deg*. Given an interpretation $\mathcal{I}$, an element $d \in \Delta^{\mathcal{I}}$, and an $\mathcal{EL}$ concept description $C$, this function is supposed to measure to which degree $d$ satisfies the conditions for membership expressed by $C$. To come up with such a measure, we use the homomorphism characterization of membership (see Theorem 2) as starting point. Basically, we consider all partial mappings from $T_C$ to $G_{\mathcal{I}}$ that map the root of $T_C$ to $d$ and respect the edge structure of $T_C$. For each of these mappings we then calculate to which degree it satisfies the homomorphism conditions, and take the degree of the best such mapping as the membership degree $deg^{\mathcal{I}}(d, C)$. We consider partial mappings rather than total ones since one of the violations of properties demanded by $C$ could be that a required role successor does not exist at all.

To formalize this idea, we first define the notion of *partial tree-to-graph homomorphisms* from description trees to description graphs. In this definition, the node labels are ignored (they will be considered in the next step).

**Definition 4.** *Let $T = (V_t, E_t, \ell_t, v_0)$ and $G = (V_g, E_g, \ell_g)$ be a description tree (with root $v_0$) and a description graph, respectively. A partial mapping $h : V_t \to V_g$ is a* partial tree-to-graph homomorphism *(ptgh) from $T$ to $G$ iff the following conditions are satisfied:*

1. *$\mathsf{dom}(h)$ is a sub-tree of $T$ with root $v_0$, i.e., $v_0 \in \mathsf{dom}(h)$ and if $(v, r, w) \in E_t$ and $w \in \mathsf{dom}(h)$, then $v \in \mathsf{dom}(h)$;*
2. *for all edges $(v, r, w) \in E_t$, $w \in \mathsf{dom}(h)$ implies $(h(v), r, h(w)) \in E_g$.*

In order to measure how far away from a homomorphism according to Definition 2 such a *ptgh* is, we define the notion of a *weighted* homomorphism between a finite $\mathcal{EL}$ description tree and an $\mathcal{EL}$ description graph.

**Definition 5.** *Let $T$ be a finite $\mathcal{EL}$ description tree, $G$ an $\mathcal{EL}$ description graph and $h : V_T \mapsto V_G$ a ptgh from $T$ to $G$. We define the* weighted *homomorphism induced by $h$ from $T$ to $G$ as a function $h_w : \mathsf{dom}(h) \to [0, 1]$ as follows. For a* *given* $v \in \mathsf{dom}(h)$*, let* $k^*(v)$ *be the number of successors of* $v$ *in* $T$*, and* $v_1, \ldots, v_k$ *the* $k$ $(0 \le k \le k^*(v))$ *children of* $v$ *in* $T$ *such that* $v_i \in \mathsf{dom}(h)$*. Then*

$$h_w(v) := \begin{cases} 1 & \text{if } |\ell_T(v)| + k^*(v) = 0 \\ \frac{|\ell_T(v) \cap \ell_G(h(v))| + \sum_{1 \le i \le k} h_w(v_i)}{|\ell_T(v)| + k^*(v)} & \text{otherwise.} \end{cases}$$

It is easy to see that $h_w$ is well-defined. In fact, $T$ is a finite tree, which ensures that the recursive definition of $h_w$ is well-founded. In addition, the first case in the definition ensures that division by zero is avoided. Using value 1 in this case is justified since then no property is required. In the second case, missing concept names and missing successors decrease the weight of a node since then the required name or successor contributes to the denominator, but not to the numerator. Required successors that are there are only counted if they are successors for the correct role, and then they do not contribute with value 1 to the numerator, but only with their weight (i.e., the degree to which they match the requirements for this successor).

When defining the value of the membership function $deg^{\mathcal{I}}(d, C)$, we do not use the concept $C$ directly, but rather its reduced form $C^r$. This will ensure that *deg* satisfies property *M2* (see Proposition 2 below).

**Definition 6.** *Let* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *be an interpretation,* $d$ *an element of* $\Delta^{\mathcal{I}}$*, and* $C$ *an* $\mathcal{EL}$ *concept description with reduced form* $C^r$*. In addition, let* $\mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$ *be the set of all ptghs from* $T_{C^r}$ *to* $G_{\mathcal{I}}$ *with* $h(v_0) = d$*. The set* $\mathcal{V}^{\mathcal{I}}(d, C^r)$ *of all relevant values is defined as*

$$\mathcal{V}^{\mathcal{I}}(d, C^r) := \{q \mid h_w(v_0) = q \text{ and } h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)\}.$$

*Then we define* $deg^{\mathcal{I}}(d, C) := \max \mathcal{V}^{\mathcal{I}}(d, C^r)$*.*

If the interpretation $\mathcal{I}$ is infinite, there may exist infinitely many *ptghs* from $T_{C^r}$ to $G_{\mathcal{I}}$ with $h(v_0) = d$. Therefore, it is not immediately clear whether the maximum in the above definition actually exists, and thus whether $deg^{\mathcal{I}}(d, C)$ is well-defined. To prove that the maximum exists also for infinite interpretations, we show that the set $\mathcal{V}^{\mathcal{I}}(d, C^r)$ is actually a finite set. For this purpose, we introduce canonical interpretations induced by *ptghs.*

**Definition 7 (Canonical Interpretation).** *Let* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *be an interpretation,* $C$ *an* $\mathcal{EL}$ *concept description and* $h$ *be a ptgh from* $T_{C^r}$ *to* $G_{\mathcal{I}}$*. The* canonical interpretation $\mathcal{I}_h$ *induced by* $h$ *is the one having the description tree* $T_{\mathcal{I}_h} = (V_{\mathcal{I}_h}, E_{\mathcal{I}_h}, v_0, \ell_{\mathcal{I}_h})$ *with*

$$\begin{aligned} V_{\mathcal{I}_h} &:= \mathsf{dom}(h), \\ E_{\mathcal{I}_h} &:= \{vrw \in E_{T_{C^r}} \mid v, w \in \mathsf{dom}(h)\} \\ \ell_{\mathcal{I}_h}(v) &:= \ell_{T_{C^r}}(v) \cap \ell_{\mathcal{I}}(h(v)) \text{ for all } v \in \mathsf{dom}(h). \end{aligned}$$

**Lemma 1.** *Let* $\mathcal{I}$ *be an interpretation,* $d \in \Delta^{\mathcal{I}}$ *and* $C$ *an* $\mathcal{EL}$ *concept description. Then the following two properties hold:*

1. *there are only finitely many different canonical interpretations induced by ptghs* $h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$*;*
2. *for every* $h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$*, the identity mapping* $i^{\mathcal{I}_h} : dom(h) \rightarrow V_{\mathcal{I}_h}$ *with* $i^{\mathcal{I}_h}(v) = v$ *for all* $v \in \mathsf{dom}(h)$ *is a ptgh from* $T_C$ *to* $T_{\mathcal{I}_h}$ *that satisfies* $h_w(v_0) = i_w^{\mathcal{I}_h}(v_0)$*.*

The first statement is an easy consequence of the fact that the description tree for a canonical interpretation has nodes from the finite set of nodes of $T_{C^r}$ and labels from the finite set of concept and role names. The second fact is not hard to show, and it obviously implies that the set $\mathcal{V}^{\mathcal{I}}(d, C^r)$ is finite. Consequently, $deg^{\mathcal{I}}(d, C)$ is well-defined. Moreover, as an easy consequence of the proof of Lemma 1 we can show that the same value can be obtained by considering the corresponding canonical interpretation. To be more precise:

**Lemma 2.** *Let* $\mathcal{I} = (\Delta^{\mathcal{I}}, .^{\mathcal{I}})$ *be an interpretation,* $d \in \Delta^{\mathcal{I}}$ *and* $C$ *an* $\mathcal{EL}$ *concept description. Let* $h$ *be a ptgh from* $T_{C^r}$ *to* $G_{\mathcal{I}}$ *such that* $h(v_0) = d$ *and* $deg^{\mathcal{I}}(d, C) = h_w(v_0)$*. In addition, let* $\mathcal{I}_h$ *be the corresponding canonical interpretation. Then,* $deg^{\mathcal{I}_h}(v_0, C) = deg^{\mathcal{I}}(d, C)$*.*

If the interpretation $\mathcal{I}$ is finite, $deg^{\mathcal{I}}(d, C)$ for $d \in \Delta^{\mathcal{I}}$ and an $\mathcal{EL}$ concept description $C$ can actually be computed in polynomial time. The polynomial time algorithm described in [1] is inspired by the polynomial time algorithm for checking the existence of a homomorphism between $\mathcal{EL}$ description trees [5,4], and similar to the algorithm for computing the similarity degree between $\mathcal{EL}$ concept descriptions introduced in [15]. Finally, it remains to show that *deg* satisfies the properties required for a membership function.

**Proposition 2.** *The function deg satisfies M1 and M2.*

In fact, *M1* is easy to show and *M2* follows from the fact that we use the reduced form of a concept description rather than the description itself. Otherwise, *M2* would not hold. For example, consider the concept description $C := \exists r.A \sqcap \exists r.(A \sqcap B)$, which is equivalent to its reduced form $C^r = \exists r.(A \sqcap B)$. Let $d$ be an individual that has a single $r$-successor belonging to $A$, but not to $B$. Then using $C$ instead of $C^r$ would yield membership degree 3/4, whereas the use of $C^r$ yields the degree 1/2.

## 5 Reasoning in $\tau\mathcal{EL}(deg)$

We start with investigating the complexity of terminological reasoning (subsumption, satisfiability) in $\tau\mathcal{EL}(deg)$, and then turn to assertional reasoning (consistency, instance). In the following, we assume that all concept descriptions $E$ occurring in threshold concepts $E_{\sim q}$ are reduced (i.e., $E^r = E$), and thus we can directly use $E$ when computing membership degrees. This is without loss of generality since the reduced form of an $\mathcal{EL}$ concept description can be computed in polynomial time.

**Terminological Reasoning.** In contrast to $\mathcal{EL}$, where every concept description is satisfiable, we have seen in Section 3 that there are unsatisfiable $\tau\mathcal{EL}(deg)$ concept descriptions, such as $A_{\geq 1} \sqcap A_{<1}$. Thus, the satisfiability problem is non-trivial in $\tau\mathcal{EL}(deg)$. In fact, by a simple reduction from the well-known NP-complete problem ALL-POS ONE-IN-THREE 3SAT [11], we can show that testing $\tau\mathcal{EL}(deg)$ concept descriptions for satisfiability is actually NP-hard. The main idea underlying this reduction is that, for any three distinct concept names $A_i, A_j, A_k$, an individual belongs to $(A_i \sqcap A_j \sqcap A_k)_{\leq 1/3} \sqcap (A_i \sqcap A_j \sqcap A_k)_{\geq 1/3}$ iff it belongs to exactly one of these three concepts. This also yields coNP-hardness of subsumption in $\tau\mathcal{EL}(deg)$ since unsatisfiability can be reduced to subsumption: $\widehat{C}$ is not satisfiable iff $\widehat{C} \sqsubseteq A_{\geq 1} \sqcap A_{<1}$.

**Lemma 3.** *In $\tau\mathcal{EL}(deg)$, satisfiability is NP-hard and subsumption is coNP-hard.*

Before proving an NP upper bound for satisfiability, we show that the homomorphism characterization of membership in an $\mathcal{EL}$ concept can be extended to $\tau\mathcal{EL}(deg)$. For this, we first extend $\mathcal{EL}$ description graphs to $\tau\mathcal{EL}(deg)$ description graphs. This is done by allowing the node labelling function to assign, in addition, threshold concepts as labels.

**Definition 8.** *Let $\widehat{H} = (V_H, E_H, \widehat{\ell}_H)$ be a $\tau\mathcal{EL}(deg)$ description graph and $\mathcal{I}$ an interpretation with associated $\mathcal{EL}$ description graph $G_\mathcal{I} = (V_\mathcal{I}, E_\mathcal{I}, \ell_\mathcal{I})$. The mapping $\phi : V_H \rightarrow V_\mathcal{I}$ is a $\tau$-homomorphism from $\widehat{H}$ to $G_\mathcal{I}$ iff*

1. *$\phi$ is a homomorphism from $\widehat{H}$ to $G_\mathcal{I}$ according to Definition 2, where threshold concepts in labels are ignored,*
2. *for all $v \in V_H$: if $E_{\sim q} \in \widehat{\ell}_H(v)$, then $\phi(v) \in [E_{\sim q}]^\mathcal{I}$.*

If the interpretation $\mathcal{I}$ is finite, then the existence of a $\tau$-homomorphism can be checked in polynomial time. Intuitively, for the first condition one just needs to check for the existence of a classical homomorphism, and for the second one needs to compute membership degrees. Both can be done in polynomial time. Similar to $\mathcal{EL}$, the existence of a $\tau$-homomorphism characterizes membership in $\tau\mathcal{EL}(deg)$ concept descriptions.

**Theorem 4.** *Let $\mathcal{I}$ be an interpretation, $d \in \Delta^\mathcal{I}$, and $\widehat{C}$ a $\tau\mathcal{EL}(deg)$ concept description. Then, $d \in \widehat{C}^\mathcal{I}$ iff there exists a $\tau$-homomorphism $\phi$ from $T_{\widehat{C}}$ to $G_\mathcal{I}$ such that $\phi(v_0) = d$.*

This theorem can be used to prove a bounded model property for $\tau\mathcal{EL}(deg)$ concept descriptions.

**Lemma 4.** *Let $\widehat{C}$ be a $\tau\mathcal{EL}(deg)$ concept description of size $\mathsf{s}(\widehat{C})$. If $\widehat{C}$ is satisfiable, then there exists an interpretation $\mathcal{J}$ such that $\widehat{C}^\mathcal{J} \neq \emptyset$ and $|\Delta^\mathcal{J}| \leq \mathsf{s}(\widehat{C})$.*

*Proof sketch.* Since $\widehat{C}$ is satisfiable, there is an interpretation $\mathcal{I}$ and some $d \in \Delta^\mathcal{I}$ such that $d \in \widehat{C}^\mathcal{I}$. Therefore, there exists a $\tau$-homomorphism $\phi$ from $T_{\widehat{C}}$ to $G_\mathcal{I}$

with $\phi(v_0) = d$. The idea is to use $\phi$ and small fragments of $\mathcal{I}$ to build $\mathcal{J}$ and a $\tau$-homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$, and then apply Theorem 4 to $\widehat{C}$ and $\mathcal{J}$.

The interpretation $\mathcal{J}$ is built in two steps. We first use as base interpretation $\mathcal{I}_0$ the interpretation associated to the description tree $T_{\widehat{C}}$, where we ignore labels of the form $E_{\sim q}$. Then the identity mapping $\phi_{id}$ is a homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}_0}$. However, this interpretation and homomorphism need not satisfy Condition 2 of Definition 8. To repair this, we extend $\mathcal{I}_0$ to $\mathcal{J}$ by adding appropriate canonical interpretations. To be more precise, for a given node $v$ in $\mathcal{I}_0$ that has $E_{\sim q}$ in its label, we know that $\phi(v) \in [E_{\sim q}]^{\mathcal{I}}$, i.e. $deg^{\mathcal{I}}(\phi(v), E) \sim q$. By Lemma 2, we do not need all of $\mathcal{I}$ to obtain the degree $deg^{\mathcal{I}}(\phi(v), E)$. It is sufficient to use the fragment corresponding to the canonical interpretation. The interpretation $\mathcal{J}$ satisfying $\widehat{C}$ is obtained from $\mathcal{I}_0$ by plugging in such canonical interpretations where ever it is required by threshold concepts in labels of nodes (see [1] for details).

Since the size of $\mathcal{I}_0$ is bounded by the size of $\widehat{C}$ (without counting the threshold concepts) and since the size of a canonical interpretation added to satisfy a threshold concept $E_{\sim q}$ in $\widehat{C}$ is bounded by the size of $E$, this yields the required bound for the size of $\mathcal{J}$. $\square$

This lemma yields a standard guess-and-check NP-algorithm to decide satisfiability of $\widehat{C}$: first guess an interpretation $\mathcal{J}$ of size at most $\mathsf{s}(\widehat{C})$, and then check (in polynomial time) whether there exists a $\tau$-homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$.

A coNP-upper bound for subsumption cannot directly be obtained from the fact that satisfiability is in NP. In fact, though we have $\widehat{C} \sqsubseteq \widehat{D}$ iff $\widehat{C} \sqcap \neg\widehat{D}$ is unsatisfiable, this equivalence cannot be used directly since $\neg\widehat{D}$ need not be a $\tau\mathcal{EL}(deg)$ concept description. Nevertheless, we can extend the ideas used in the proof of Lemma 4 to obtain a bounded model property for satisfiability of concepts of the form $\widehat{C} \sqcap \neg\widehat{D}$.

**Lemma 5.** *Let $\widehat{C}$ and $\widehat{D}$ be $\tau\mathcal{EL}(deg)$ concept descriptions of respective sizes $\mathsf{s}(\widehat{C})$ and $\mathsf{s}(\widehat{D})$. If $\widehat{C} \sqcap \neg\widehat{D}$ is satisfiable, then there exists an interpretation $\mathcal{J}$ such that $\widehat{C}^{\mathcal{J}} \setminus \widehat{D}^{\mathcal{J}} \neq \emptyset$ and $|\Delta^{\mathcal{J}}| \leq \mathsf{s}(\widehat{C}) \times \mathsf{s}(\widehat{D})$.*

*Proof sketch.* We first apply the construction used in the proof of Lemma 4 to construct, for a given interpretation $\mathcal{I}$ with $\widehat{C}^{\mathcal{I}} \setminus \widehat{D}^{\mathcal{I}} \neq \emptyset$, an interpretation $\mathcal{J}_0$ such that $\widehat{C}^{\mathcal{J}_0} \neq \emptyset$ and $|\Delta^{\mathcal{J}_0}| \leq \mathsf{s}(\widehat{C})$. This construction is such that $G_{\mathcal{J}_0}$ is tree-shaped and there is a homomorphism $\varphi$ from $G_{\mathcal{J}_0}$ to $G_{\mathcal{I}}$ with $\varphi(v_0) = d$. We then use $\varphi$ to extend $\mathcal{J}_0$ to $\mathcal{J}$ such that $v_0 \notin \widehat{D}^{\mathcal{J}}$ holds. Starting with the root $v_0$, we consider all the nodes in $\Delta^{\mathcal{J}_0}$ in a top-down manner.

First, assume that $\widehat{D}$ contains a top-level conjunct of the form $E_{\leq q}$ such that $d = \varphi(v_0) \notin [E_{\leq q}]^{\mathcal{I}}$, but $v_0 \in [E_{\leq q}]^{\mathcal{J}_0}$. Then we attach to $v_0$ a canonical interpretation that yields for $d$ the same membership degree as $\mathcal{I}$ to ensure that, in the extended interpretation, $v_0$ no longer belongs to $E_{\leq q}$.

Now, consider the case where $\widehat{D}$ contains a top-level conjunct $\widehat{F} = \exists r.\widehat{F}'$ such that $d = \varphi(v_0) \notin \widehat{F}^{\mathcal{I}}$, but $v_0 \in \widehat{F}^{\mathcal{J}_0}$. Then there is an $r$-successor $v$ of $v_0$ that satisfies $v \in [\widehat{F}']^{\mathcal{J}_0}$, but since $\varphi(v)$ is an $r$-successor of $\varphi(v_0)$ in $\mathcal{I}$, we also have

$\varphi(v) \notin [\widehat{F}']^{\mathcal{I}}$. We can now recursively apply the construction to $v$. Overall, the construction terminates and considers every node in $\Delta^{\mathcal{J}_0}$ only once since $G_{\mathcal{J}_0}$ is tree-shaped. Since the number of nodes in $\Delta^{\mathcal{J}_0}$ is bounded by $\mathsf{s}(\widehat{C})$ and the size of each of the added canonical interpretations is bounded by $\mathsf{s}(\widehat{D})$, we obtain the desired bound on the size of $\mathcal{J}$. □

The lemma yields an obvious guess-and-check NP-algorithm for non-subsumption, which shows that subsumption is in co-NP. Overall, we thus have shown:

**Theorem 5.** *In $\tau\mathcal{EL}(deg)$, satisfiability is NP-complete and subsumption coNP-complete.*

**Assertional Reasoning.** Information about specific individuals can be expressed in an ABox. An *ABox* $\mathcal{A}$ is a finite set of *assertions* of the form $\widehat{C}(a)$ or $r(a,b)$, where $\widehat{C}$ is a $\tau\mathcal{EL}(deg)$ concept description, $r \in \mathsf{N_R}$, and $a, b$ are individual names. In addition to concept and role names, an interpretation $\mathcal{I}$ now assigns domain elements $a^{\mathcal{I}}$ to individual names $a$. The assertion $\widehat{C}(a)$ is satisfied by $\mathcal{I}$ iff $a^{\mathcal{I}} \in \widehat{C}^{\mathcal{I}}$, and $r(a,b)$ is satisfied by $\mathcal{I}$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. The interpretation $\mathcal{I}$ is a *model* of $\mathcal{A}$ iff $\mathcal{I}$ satisfies all assertions in $\mathcal{A}$. The ABox $\mathcal{A}$ is *consistent* iff it has a model, and the individual $a$ is an *instance* of the concept $\widehat{C}$ in $\mathcal{A}$ (written as $\mathcal{A} \models \widehat{C}(a)$) iff $a^{\mathcal{I}} \in \widehat{C}^{\mathcal{I}}$ holds in all models of $\mathcal{A}$.

Since satisfiability can obviously be reduced to consistency, and subsumption to the instance problem, the lower bounds shown above also hold for assertional reasoning. Regarding upper bounds, the consistency problem can be tackled in a similar way as the satisfiability problem. As shown in [13], $\mathcal{EL}$ ABoxes can be translated into $\mathcal{EL}$ description graphs and consistency can be characterized using homomorphisms between description graphs. Again, this characterization can be extended to $\tau\mathcal{EL}(deg)$, and can be used to prove an appropriate bounded model property with a polynomial bound. Similar to our treatment of subsumption, this can then be used to obtain a bounded model property for non-instance ($\mathcal{A} \not\models \widehat{C}(a)$)). However, there the bound on the model has the size of $\widehat{C}$ in the exponent. For this reason, we obtain a coNP upper bound for the instance problem only if we consider data complexity [8], where the size of the query concept $\widehat{C}$ is assumed to be constant.

**Theorem 6.** *In $\tau\mathcal{EL}(deg)$, consistency is NP-complete, and instance checking is coNP-complete w.r.t. data complexity.*

The instance problem becomes simpler if we consider only $\mathcal{EL}$ ABoxes and *positive $\tau\mathcal{EL}(deg)$ concept descriptions*, i.e., concept descriptions $\widehat{C}$ that only contain threshold concepts of the form $E_{\geq t}$ or $E_{>t}$. Basically, given an $\mathcal{EL}$ ABox, a positive $\tau\mathcal{EL}(deg)$ concept description $\widehat{C}$, and an individual $a$, one considers the interpretation $\mathcal{I}$ corresponding to the description graph of $\mathcal{A}$, and then checks whether there is a $\tau$-homomorphism $\phi$ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = a$ (see [1] for details).

**Proposition 3.** *For positive $\tau\mathcal{EL}(deg)$ concept descriptions and $\mathcal{EL}$ ABoxes, the instance problem can be decided in polynomial time.*

## 6 Concept Similarity and Relaxed Instance Queries

In its most general form, a concept similarity measure (CSM) $\bowtie$ is a function that maps each pair of concepts $C, D$ (of a given DL) to a value $C \bowtie D \in [0,1]$ such that $C \bowtie C = 1$. Intuitively, the higher the value of $C \bowtie D$ is, the more similar the two concepts are supposed to be. Such measures can in principle be defined for arbitrary DLs, but here we restrict the attention to CSMs between $\mathcal{EL}$ concepts, i.e., a CSM is a mapping $\bowtie : \mathcal{C}_{\mathcal{EL}} \times \mathcal{C}_{\mathcal{EL}} \to [0,1]$.

Ecke et al. [10,9] use CSMs to relax instance queries, i.e., instead of requiring that an individual is an instance of the query concept, they only require that it is an instance of a concept that is "similar enough" to the query concept.

**Definition 9 ([10,9]).** *Let $\bowtie$ be a CSM, $\mathcal{A}$ an $\mathcal{EL}$ ABox, and $t \in [0,1)$. The individual $a \in \mathsf{N_I}$ is a* relaxed instance *of the $\mathcal{EL}$ query concept $Q$ w.r.t. $\mathcal{A}$, $\bowtie$, and the threshold $t$ iff there exists an $\mathcal{EL}$ concept description $X$ such that $Q \bowtie X > t$ and $\mathcal{A} \models X(a)$. The set of all individuals occurring in $\mathcal{A}$ that are relaxed instances of $Q$ w.r.t. $\mathcal{A}$, $\bowtie$, and $t$ is denoted by $\mathsf{Relax}_t^{\bowtie}(Q, \mathcal{A})$.*

We apply the same idea on the semantic level of an interpretation rather than the ABox level to obtain graded membership functions from similarity measures.

**Definition 10.** *Let $\bowtie$ be a CSM. Then, for each interpretation $\mathcal{I}$, we define the function $m_{\bowtie}^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \to [0,1]$ as*

$$m_{\bowtie}^{\mathcal{I}}(d, C) := \max\{C \bowtie D \mid D \in \mathcal{C}_{\mathcal{EL}} \text{ and } d \in D^{\mathcal{I}}\}.$$

For an arbitrary CSM $\bowtie$, the maximum in this definition need not exist since $D$ ranges over infinitely many concept descriptions. However, two properties that are satisfied by many similarity measures considered in the literature are sufficient to obtain well-definedness for $m_{\bowtie}$. The first is equivalence invariance:

– The CSM $\bowtie$ is *equivalence invariant* iff $C \equiv C'$ and $D \equiv D'$ implies $C \bowtie D = C' \bowtie D'$ for all $C, C', D, D' \in \mathcal{C}_{\mathcal{EL}}$.

To formulate the second property, we need to recall that the *role depth* of an $\mathcal{EL}$ concept description $C$ is the maximal nesting of existential restrictions in $C$; equivalently, it is the height of the description tree $T_C$. The *restriction $C_k$ of $C$ to role depth $k$* is the concept description whose description tree is obtained from $T_C$ by removing all the nodes (and edges leading to them) whose distance from the root is larger than $k$.

– The CSM $\bowtie$ is *role-depth bounded* iff $C \bowtie D = C_k \bowtie D_k$ for all $C, D \in \mathcal{C}_{\mathcal{EL}}$ and any $k$ that is larger than the minimal role depth of $C, D$.

Role-depth boundedness implies that, in Definition 10, we can restrict the maximum computation to concepts $D$ whose role depth is at most $d+1$, where $d$ is the role depth of $C$. Since it is well-known that, up to equivalence, $\mathcal{C}_{\mathcal{EL}}$ contains only finitely many concept descriptions of any fixed role depth, these two properties yield well-definedness for $m_{\bowtie}$. For $m_{\bowtie}$ to be a graded membership function, it also needs to satisfy the properties *M1* and *M2*. To obtain these two properties for $m_{\bowtie}$, we must require that $\bowtie$ satisfies the following additional property:

– The CSM $\bowtie$ is *equivalence closed* iff the following equivalence holds: $C \equiv D$ iff $C \bowtie D = 1$.

**Proposition 4.** *Let $\bowtie$ be an equivalence invariant, role-depth bounded, and equivalence closed CSM. Then $m_{\bowtie}$ is a well-defined graded membership function.*

Consequently, an equivalence invariant, role-depth bounded, and equivalence closed CSM $\bowtie$ induces a DL $\tau\mathcal{EL}(m_{\bowtie})$. Computing instances of threshold concepts of the form $Q_{>t}$ in this logic corresponds to answering relaxed instance queries w.r.t. $\bowtie$.

**Proposition 5.** *Let $\bowtie$ be an equivalence invariant, role-depth bounded, and equivalence closed CSM, $\mathcal{A}$ an $\mathcal{EL}$ ABox, and $t \in [0,1)$. Then*

$$\mathsf{Relax}_t^{\bowtie}(Q, \mathcal{A}) = \{a \mid \mathcal{A} \models Q_{>t}(a) \text{ and } a \text{ occurs in } \mathcal{A}\},$$

*where the semantics of the threshold concept $Q_{>t}$ is defined as in $\tau\mathcal{EL}(m_{\bowtie})$.*

Lehman and Turhan [14] introduce a framework (called *simi framework*) that can be used to define a variety of similarity measures between $\mathcal{EL}$ concepts satisfying the properties required by our Propositions 4 and 5. Here, we consider only one instance of this framework and show that the similarity measure obtained this way induces our graded membership function *deg*.

Lehman and Turhan first define a directional measure and then combine the values obtained by comparing the concepts in both directions with this directional measure.

**Definition 11 ([14]).** *Let $C, D$ be two $\mathcal{EL}$ concept descriptions. If one of these two concepts is equivalent to $\top$, then we define $simi_d(\top, D) := 1$ for all $D$ and $simi_d(D, \top) := 0$ for $D \not\equiv \top$. Otherwise, let $top(C)$, $top(D)$ respectively be the set of concept names and existential restrictions in the top-level conjunction of $C, D$. We define*

$$simi_d(C, D) := \frac{\sum\limits_{C' \in top(C)} \max\{simi_a(C', D') \mid D' \in top(D)\}}{|top(C)|}, \quad \text{where}$$

$$simi_a(A, A) := 1, \quad simi_a(A, B) := 0 \text{ for } A, B \in \mathsf{N_C}, A \neq B,$$
$$simi_a(\exists r.E, A) := simi_a(A, \exists r.E) := 0 \text{ for } A \in \mathsf{N_C},$$
$$simi_a(\exists r.E, \exists r.F) := simi_d(E, F), \quad simi_a(\exists r.E, \exists s.F) := 0 \text{ for } r, s \in \mathsf{N_R}, r \neq s.$$

*The bidirectional similarity measure $\bowtie^1$ is then defined as*

$$C \bowtie^1 D := \min\{simi_d(C^r, D^r), simi_d(D^r, C^r)\}.$$

It is easy to show that $\bowtie^1$ is equivalence invariant, role-depth bounded, and equivalence closed. Note that equivalence invariance depends on the fact that we apply $simi_d$ to the reduced forms of $C, D$. Since $\bowtie^1$ satisfies the properties required by Propositions 4, it induces a graded membership function $m_{\bowtie^1}$. We can show that this function coincides with the graded membership function introduced in Section 4 (see [1] for the proof).

**Theorem 7.** *For all interpretations $\mathcal{I}$, $d \in \Delta^{\mathcal{I}}$, and $\mathcal{EL}$ concept descriptions $Q$ we have $m^{\mathcal{I}}_{\bowtie^1}(d, Q) = deg^{\mathcal{I}}(d, Q)$.*

Proposition 5 thus implies that answering of relaxed instance queries w.r.t. $\bowtie^1$ is the same as computing instances for threshold concepts of the form $Q_{>}t$ in $\tau\mathcal{EL}(deg)$. Since such concepts are positive, Proposition 3 yields the following corollary.

**Corollary 1.** *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox, $Q$ an $\mathcal{EL}$ query concept, $a$ an individual name, and $t \in [0, 1)$. Then it can be decided in polynomial time whether $a \in \mathsf{Relax}_t^{\bowtie^1}(Q, \mathcal{A})$ or not.*

Note that Ecke et al. [10,9] show only an NP upper bound w.r.t. data complexity for this problem, albeit for a larger class of instances of the *simi* framework.

## 7 Conclusion

We have introduced a family of DLs $\tau\mathcal{EL}(m)$ parameterized with a graded membership function $m$, which extends the popular lightweight DL $\mathcal{EL}$ by threshold concepts that can be used to approximate classical concepts. Inspired by the homomorphism characterization of membership in $\mathcal{EL}$ concepts, we have defined a particular membership function *deg* and have investigated the complexity of reasoning in $\tau\mathcal{EL}(deg)$. It turns out that the higher expressiveness takes its toll: whereas reasoning in $\mathcal{EL}$ can be done in polynomial time, it is NP- or coNP-complete in $\tau\mathcal{EL}(deg)$, depending on which inference problem is considered. We have also shown that concept similarity measures satisfying certain properties can be used to define graded membership functions. In particular, the function *deg* can be constructed in this way from a particular instance of the *simi* framework of Lehmann and Turhan [14]. Nevertheless, our direct definition of *deg* based on homomorphisms is important since the partial tree-to-graph homomorphisms used there are the main technical tool for showing our decidability and complexity results.

While introduced as formalism for defining concepts by approximation, a possible use-case for $\tau\mathcal{EL}(deg)$ is relaxation of instance queries, as motivated and investigated in [10,9]. Compared to the setting considered in [10,9], $\tau\mathcal{EL}(deg)$ yields a considerably more expressive query language since we can combine threshold concepts using the constructors of $\mathcal{EL}$ and can also forbid that thresholds are reached. Restricted to the setting of relaxed instance queries, our approach actually allows relaxed instance checking in polynomial time. On the other hand, [10,9] can also deal with other instances of the *simi* framework.

An important topic for future research is to consider graded membership functions $m_{\bowtie}$ that are induced by other instances of *simi*. We conjecture that these instances can also be defined directly by an appropriate adaptation of our homomorphism-based definition. The hope is then that our decidability and complexity results can be generalized to these functions. Another important topic for future research is to add TBoxes. While acyclic TBoxes can already be handled by our approach through unfolding, we would like to treat them directly by an adaptation of the homomorphism-based approach to avoid a possible exponential blowup due to unfolding. For cyclic and general TBoxes, homomorphisms probably need to be replaced by simulations [2,9].

## References


1. Baader, F., Brewka, G., Fernández Gil, O.: Adding threshold concepts to the description logic $\mathcal{EL}$. LTCS-Report LTCS-15-09, TU Dresden, Germany (2015). See `http://lat.inf.tu-dresden.de/research/reports.html`
2. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: Proc. IJCAI 2003 (2003)
3. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook. Cambridge University Press (2003)
4. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. In: Proc. IJCAI 1999 (1999)
5. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. LTCS-Report LTCS-98-09. RWTH Aachen, Germany (1998). See `http://lat.inf.tu-dresden.de/research/reports.html`
6. Borgwardt, S., Distel, F., Peñaloza, R.: The limits of decidability in fuzzy description logics with general concept inclusions. Artificial Intelligence 218 (2015)
7. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In: Proc. ECAI 2004 (2004)
8. Donini, F.M., Lenzerini, M., Nardi, D., Schaerf, A.: Deduction in concept languages: From subsumption to instance checking. J. Log. Comput. 4(4) (1994)
9. Ecke, A., Peñaloza, R., Turhan, A.Y.: Similarity-based relaxed instance queries. Journal of Applied Logic (in press, 2015)
10. Ecke, A., Peñaloza, R., Turhan, A.: Answering instance queries relaxed by concept similarity. In: Proc. KR 2014 (2014)
11. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman (1979)
12. Küsters, R.: Non-Standard Inferences in Description Logics. LNCS (LNAI), vol. 2100. Springer, Heidelberg (2001)
13. Küsters, R., Molitor, R.: Approximating most specific concepts in description logics with existential restrictions. AI Commun. 15(1) (2002)
14. Lehmann, K., Turhan, A.-Y.: A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts. In: del Cerro, L.F., Herzig, A., Mengin, J. (eds.) JELIA 2012. LNCS, vol. 7519, pp. 307–319. Springer, Heidelberg (2012)
15. Suntisrivaraporn, B.: A similarity measure for the description logic EL with unfoldable terminologies. In: Proc. INCoS 2013 (2013)
16. Baader, F., Brewka, G., Fernández Gil, O.: Adding threshold concepts to the description logic $\mathcal{EL}$. In: Proc. DL 2015 (2015)