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

Franz Baader[1], Gerhard Brewka[2], and Oliver Fernández Gil[2]*
baader@tcs.inf.tu-dresden.de
{brewka,fernandez}@informatik.uni-leipzig.de

[1] Theoretical Computer Science, TU Dresden, Germany
[2] Department of Computer Science, University of Leipzig, Germany

The description logic (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 [6]. On the other hand, though quite inexpressive, $\mathcal{EL}$ can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT.[3] In $\mathcal{EL}$ we can, for 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:

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

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.

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

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 [5] 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. 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.

We name our new logic $\tau\mathcal{EL}(m)$, where the membership degree function m is a parameter in defining the logic. In [2], we propose one specific such function *deg*, but we do not claim this is the only reasonable way to define such a function. Nevertheless, membership functions are not arbitrary. There are two properties we require such functions to satisfy:

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

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

Property *M2* expresses the intuition that the membership value should not depend on the syntactic form of a concept, but only on its semantics.

The set of $\tau\mathcal{EL}(m)$ concept descriptions is defined inductively, starting from finite sets of concept names $\mathsf{N_C}$ and role names $\mathsf{N_R}$, as follows:

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

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 $\widehat{C}, \widehat{D}$ are $\tau\mathcal{EL}(m)$ concept descriptions. For a given interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, the semantics of the new threshold concepts is defined as follows:

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

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

To make things more concrete, we introduce in [2] a specific membership function, denoted *deg*, which satisfies properties *M1* and *M2*. Given an interpretation $\mathcal{I}$, an element $d \in \Delta^{\mathcal{I}}$, and an $\mathcal{EL}$ concept description C, this function measures to which degree d satisfies the conditions for membership expressed by C. To come up with such a function, we use the homomorphism characterization of crisp membership in $\mathcal{EL}$. In $\mathcal{EL}$, concept descriptions and interpretations can be translated into $\mathcal{EL}$ description trees and $\mathcal{EL}$ description graphs, respectively (see [4,1]). Then, homomorphisms between $\mathcal{EL}$ description trees can be used to characterized subsumption in $\mathcal{EL}$ [4]. The proof of this result can be easily adapted to obtain the following characterization of *element-hood* in $\mathcal{EL}$.

Theorem 1. *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 φ from T_C to $G_{\mathcal{I}}$ such that $\varphi(v_0) = d$.*

Using Theorem 1 as a starting point, we consider all partial mappings h 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 $deg^{\mathcal{I}}(d, C)$. Intuitively, to compute the degree associated to a partial mapping h, we define the *weighted* homomorphism induced by h as a function $h_w : \mathsf{dom}(h) \to [0, 1]$. Basically, in the definition of this function, the individual d is punished (in the sense that its membership degree is lowered) for each missing property (i.e., required element-hood in a concept name or an existential restriction) in a uniform way (see [2] for the precise definition).

In [2], we describe an algorithm that, given a finite interpretation $\mathcal{I}$, computes $deg^{\mathcal{I}}(d, C)$ in polynomial time. This polynomial time algorithm is inspired by the polynomial time algorithm for checking the existence of a homomorphism between $\mathcal{EL}$ description trees [3,4], and similar to the algorithm for computing the similarity degree between $\mathcal{EL}$ concept descriptions introduced in [9].

The main technical contribution of this work is, however, the investigation of the complexity of terminological (subsumption, satisfiability) and assertional (consistency, instance) reasoning in $\tau\mathcal{EL}(deg)$. To provide lower bounds, we show NP-hardness of the satisfiability problem by a simple reduction from the well-known NP-complete problem ALL-POS ONE-IN-THREE 3SAT [8]. The corresponding NP upper bound for satisfiability is an immediate consequence of the following polynomial bounded model property.

Lemma 1. *Let $\widehat{C}$ be a $\tau\mathcal{EL}(deg)$ concept description of size m. If $\widehat{C}$ is satisfiable, then there exists an interpretation $\mathcal{J}$ such that $\widehat{C}^{\mathcal{J}} \neq \emptyset$ and $|\Delta^{\mathcal{J}}| \leq m$.*

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 1 to obtain a polynomial bounded model property for satisfiability of concepts of the form $\widehat{C} \sqcap \neg\widehat{D}$. The same is true for ABox consistency. Regarding instance checking, the bound on the size of counter models is exponential w.r.t. combined complexity, but fortunately still polynomial w.r.t. data complexity (in the sense of [7]).

Overall, we thus obtain the following complexity results for reasoning in $\tau\mathcal{EL}(deg)$.

Theorem 2. *In the DL $\tau\mathcal{EL}(deg)$, satisfiability is NP-complete, subsumption is coNP-complete, and ABox consistency is NP-complete. Moreover, instance checking is coNP-complete w.r.t. data complexity.*

Due to the space constraints, we could not provide technical details and proofs in this extended abstract. They can be found in the technical report [2].

References

1. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: Gottlob, G., Walsh, T. (eds.) IJCAI. pp. 325–330. Morgan Kaufmann (2003)
2. Baader, F., Brewka, G., Fernández Gil, O.: Adding threshold concepts to the description logic $\mathcal{EL}$. LTCS-Report 15-09, Chair for Automata Theory, Institute for Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2015), see http://lat.inf.tu-dresden.de/research/reports.html.
3. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. LTCS-Report LTCS-98-09, LuFG Theoretical Computer Science, RWTH Aachen, Germany (1998), see http://lat.inf.tu-dresden.de/research/reports.html.
4. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. In: Dean, T. (ed.) IJCAI. pp. 96–103. Morgan Kaufmann (1999)
5. Borgwardt, S., Distel, F., Peñaloza, R.: The limits of decidability in fuzzy description logics with general concept inclusions. Artificial Intelligence 218, 23–55 (2015)
6. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In: de Mántaras, R.L., Saitta, L. (eds.) Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004. pp. 298–302. IOS Press (2004)
7. Donini, F.M., Lenzerini, M., Nardi, D., Schaerf, A.: Deduction in concept languages: From subsumption to instance checking. J. Log. Comput. 4(4), 423–452 (1994), `http://dx.doi.org/10.1093/logcom/4.4.423`
8. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman (1979)
9. Suntisrivaraporn, B.: A similarity measure for the description logic EL with unfoldable terminologies. In: 2013 5th International Conference on Intelligent Networking and Collaborative Systems, Xi'an city, Shaanxi province, China, September 9-11, 2013. pp. 408–413. IEEE (2013), `http://dx.doi.org/10.1109/INCoS.2013.77`