The Distributive, Graded Lattice of $\mathcal{EL}$ Concept Descriptions and its Neighborhood Relation

Francesco Kriegel[0000−0003−0219−0330]

Institute of Theoretical Computer Science,
Technische Universität Dresden, Dresden, Germany
francesco.kriegel@tu-dresden.de

Abstract. For the description logic $\mathcal{EL}$, we consider the neighborhood relation which is induced by the subsumption order, and we show that the corresponding lattice of $\mathcal{EL}$ concept descriptions is distributive, modular, graded, and metric. In particular, this implies the existence of a rank function as well as the existence of a distance function.

Keywords: Description logic · Distributive lattice · Modular lattice · Graded lattice · Metric lattice · Rank function · Distance function · Neighborhood relation · Upper neighbor · Lower neighbor

1 Introduction

Description Logics [3] are a family of well-founded languages for knowledge representation with a strong logical foundation as well as a widely explored hierarchy of decidability and complexity of common reasoning problems. The several reasoning tasks allow for an automatic deduction of implicit knowledge from given explicitly represented facts and axioms, and many reasoning algorithms have been developed. Description Logics are utilized in many different application domains, and in particular provide the logical underpinning of *Web Ontology Language (OWL)* [7] and its profiles.

$\mathcal{EL}$ is an example of a description logic with tractable reasoning problems, i.e., the usual inference problems can be decided in polynomial time, cf. Baader, Brandt, and Lutz in [2]. From a perspective of *lattice theory*, $\mathcal{EL}$ has not been deeply explored yet. Of course, it is apparent that the subsumption $\sqsubseteq$ with respect to some TBox $\mathcal{T}$ constitutes a quasi-order. Furthermore, in description logics supremums in the corresponding ordered set are usually called *least common subsumers*, and these exist in all cases if either no TBox is present, or if greatest fixed-point semantics are applied. Apart from that not much is known about the lattice of $\mathcal{EL}$ concept descriptions. In this document, we shall consider the neighborhood relation which is induced by the subsumption order, and we shall show that the lattice of $\mathcal{EL}$ concept descriptions is distributive, modular, graded, and metric. In particular, this implies the existence of a rank function as well as the existence of a distance function.

2 The Description Logic $\mathcal{EL}$

In this section we shall introduce the syntax and semantics of the light-weight description logic $\mathcal{EL}$ [3,2]. Throughout the whole document assume that Σ is a signature, i.e.,

ISBN 978–80–244–5328–6, Department of Computer Science, Palacký University Olomouc, 2018.

$\Sigma = \Sigma_{\mathsf{C}} \uplus \Sigma_{\mathsf{R}}$ is a disjoint union of a set Σ_{C} of *concept names* and a set Σ_{R} of *role names*. An $\mathcal{EL}$ *concept description*over Σ is a term that is constructed by means of the following inductive rule where $A \in \Sigma_{\mathsf{C}}$ and $r \in \Sigma_{\mathsf{R}}$.

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

The set of all $\mathcal{EL}$ concept descriptions over Σ is denoted by $\mathcal{EL}(\Sigma)$. The *size* $||C||$ of an $\mathcal{EL}$ concept description C is the number of nodes in its syntax tree, and we can recursively define it as follows: $||\top|| \coloneqq 1$, $||A|| \coloneqq 1$, $||C \sqcap D|| \coloneqq ||C|| + 1 + ||D||$, and $||\exists\, r.\, C|| \coloneqq 1 + ||C||$. A *concept inclusion* is an expression $C \sqsubseteq D$ where both the *premise* C as well as the *conclusion* D are concept descriptions. A *terminological box* (abbrv. *TBox*) is a finite set of concept inclusions.

An *interpretation* $\mathcal{I} \coloneqq (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ over Σ consists of a non-empty set $\Delta^{\mathcal{I}}$, called the *domain*, and an *extension function* $\cdot^{\mathcal{I}}$ that maps concept names $A \in \Sigma_{\mathsf{C}}$ to subsets $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and maps role names $r \in \Sigma_{\mathsf{R}}$ to binary relations $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. Then, the extension function is canonically extended to all $\mathcal{EL}$ concept descriptions by the following definitions.

$$\bot^{\mathcal{I}} \coloneqq \emptyset \qquad \top^{\mathcal{I}} \coloneqq \Delta^{\mathcal{I}} \qquad (C \sqcap D)^{\mathcal{I}} \coloneqq C^{\mathcal{I}} \cap D^{\mathcal{I}}$$
$$(\exists\, r.\, C)^{\mathcal{I}} \coloneqq \{\, d \in \Delta^{\mathcal{I}} \mid \exists\, e \in \Delta^{\mathcal{I}} \colon (d, e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}} \,\}$$

A concept inclusion $C \sqsubseteq D$ is *valid* in $\mathcal{I}$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. We then also refer to $\mathcal{I}$ as a *model* of $C \sqsubseteq D$, and denote this by $\mathcal{I} \models C \sqsubseteq D$. Furthermore, $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$, symbolized as $\mathcal{I} \models \mathcal{T}$, if each concept inclusion in $\mathcal{T}$ is valid in $\mathcal{I}$. The relation $\models$ is lifted to TBoxes as follows. A concept inclusion $C \sqsubseteq D$ is *entailed* by a TBox $\mathcal{T}$, denoted as $\mathcal{T} \models C \sqsubseteq D$, if each model of $\mathcal{T}$ is a model of $C \sqsubseteq D$ too. We then also say that C is *subsumed* by D with respect to $\mathcal{T}$. A TBox $\mathcal{T}$ *entails* a TBox $\mathcal{U}$, symbolized as $\mathcal{T} \models \mathcal{U}$, if $\mathcal{T}$ entails each concept inclusion in $\mathcal{U}$, or equivalently if each model of $\mathcal{T}$ is also a model of $\mathcal{U}$. In case $\mathcal{T} = \emptyset$ we may ommit the prefix "$\emptyset \models$". However, then we have to carefully interpret an expression $C \sqsubseteq D$—it either just denotes a concept inclusion, i.e., an axiom, without stating where it is valid; or it expresses that C is subsumed by D (w.r.t. $\emptyset$), i.e., $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ is satisfied in all interpretations $\mathcal{I}$.

Two $\mathcal{EL}$ concept descriptions C and D are *equivalent* with respect to $\mathcal{T}$, and we shall write $\mathcal{T} \models C \equiv D$, if $\mathcal{T} \models \{C \sqsubseteq D,\ D \sqsubseteq C\}$. As a further abbreviation, let $\mathcal{T} \models C \sqsubsetneq D$ if both $\mathcal{T} \models C \sqsubseteq D$ and $\mathcal{T} \not\models C \sqsupseteq D$, and we then say that C is *strictly subsumed* by D with respect to $\mathcal{T}$. In the sequel of this document we may also write $C \leq_{\mathcal{T}} D$ instead of $\mathcal{T} \models C \leq D$ where $\leq$ is some suitable relation symbol, e.g., $\sqsubseteq$.

It is not hard to find $\mathcal{EL}$ concept descriptions that are equivalent, i.e., have the same extension in *all* interpretations, but are not equal. It is therefore helpful for technical details to have a unique *normal form* for $\mathcal{EL}$ concept descriptions. According to [4,9] an $\mathcal{EL}$ concept description C can be transformed into a *reduced form* that is equivalent to C by exhaustive application of the *reduction rule* $D \sqcap E \mapsto D$ whenever $\emptyset \models D \sqsubseteq E$ to the subconcepts of C (modulo commutativity and associativity of $\sqcap$). It is immediately clear that each $\mathcal{EL}$ concept description C is essentially a conjunction of other $\mathcal{EL}$ concept descriptions that are no conjunctions. In particular, if we define $\mathsf{Conj}(C)$ as the set of all top-level conjuncts in C, then C has the form $\bigsqcap \mathsf{Conj}(C)$ (modulo commutativity and associativity of $\sqcap$).

It is readily verified that the *subsumption* $\sqsubseteq_{\emptyset}$ constitutes a quasi-order on $\mathcal{EL}(\Sigma)$. Hence, the quotient of $\mathcal{EL}(\Sigma)$ with respect to the induced *equivalence* $\equiv_{\emptyset}$ is an ordered set. In what follows we will not distinguish between the equivalence classes and their representatives. Furthermore, $\top$ is the greatest element, and the quotient set $\mathcal{EL}(\Sigma)/{\equiv_\emptyset}$ is a lattice that we shall symbolize by $\boldsymbol{\mathcal{EL}}(\Sigma)$. It is easy to verify that the conjunction $\sqcap$ corresponds to the finitary *infimum* operation. In a description logic allowing for disjunctions $\sqcup$, it dually holds true that the disjunction $\sqcup$ corresponds to the finitary *supremum* operation. Unfortunately, this does not apply to our considered description logic $\mathcal{EL}$. As an obvious solution, we can simply define the notion of a *supremum* specifically tailored to the case of $\mathcal{EL}$ concept descriptions as follows. The *supremum* or *least common subsumer* (abbrv. *LCS*) of two $\mathcal{EL}$ concept descriptions C and D is an $\mathcal{EL}$ concept description E with the following properties.

1. $\emptyset \models \{C \sqsubseteq E,\ D \sqsubseteq E\}$
2. For each $\mathcal{EL}$ concept description F, if $\emptyset \models \{C \sqsubseteq F,\ D \sqsubseteq F\}$, then $\emptyset \models E \sqsubseteq F$.

Since all least common subsumers of C and D are unique up to equivalence, we may denote a representative of the corresponding equivalence class by $C \vee D$. It is well known that LCS-s always exist in $\mathcal{EL}$; in particular, the least common subsumer $C \vee D$ can be computed, modulo equivalence, by means of the following recursive formula.

$$\begin{aligned} C \vee D = {} & \bigsqcap(\Sigma_{\mathsf{C}} \cap \mathsf{Conj}(C) \cap \mathsf{Conj}(D)) \\ & \sqcap \bigsqcap\{\, \exists\, r.\,(E \vee F) \mid r \in \Sigma_{\mathsf{R}},\ \exists\, r.\, E \in \mathsf{Conj}(C),\ \text{and}\ \exists\, r.\, F \in \mathsf{Conj}(D)\,\} \end{aligned}$$

It is easy to see that the equivalence $\equiv_\emptyset$ is compatible with both $\sqcap$ and $\vee$. Of course, the definition of a LCS can be extended to an arbitrary number of arguments in the obvious way, and we shall then denote the LCS of the concept descriptions C_t, $t \in T$, by $\bigvee\{\, C_t \mid t \in T\,\}$.

3 The Neighborhood Problem

In this section we consider the *neighborhood problem* for $\mathcal{EL}$. We have already seen that the set of $\mathcal{EL}$ concept descriptions constitutes a lattice. It is only natural to consider the question whether there exists a neighborhood relation which corresponds to the subsumption order. Remark that for an order relation $\leq$ on some set P its *neighborhood relation* or *transitive reduction* is defined as

$$\prec \coloneqq \lneq \setminus (\lneq \circ \lneq) = \{\, (p, q) \mid p \lneq q \text{ and there exists no } x \text{ such that } p \lneq x \lneq q \,\}.$$

Clearly, if P is finite, then the transitive closure $\prec^+$ equals the irreflexive part $\lneq$. However, there are infinite ordered sets where this does not hold true; even worse, there are cases where $\prec^+$ is empty. Consider, for instance, the set $\mathbb{R}$ of real numbers with their usual ordering $\leq$. It is well-known that $\mathbb{R}$ is dense in itself, that is, for each pair $x \lneq y$, there is another real number z such that $x \lneq z \lneq y$—thus, there are no neighboring real numbers. In general, we say that $\leq$ is *neighborhood generated* if $\prec^+ = \lneq$ is satisfied. Clearly, $\leq$ is a neighborhood generated order relation if, and only if, there is a finite path $p = x_0 \prec x_1 \prec \ldots \prec x_n = q$ for each pair $p \leq q$. An alternative formulation is the following. $\leq$ is not neighborhood generated if, and only if, there exists some pair $p \lneq q$ such that every finite path $p = x_0 \lneq x_1 \lneq \ldots \lneq x_n = q$ can be refined, that is, there is some index i and an element y such that $x_i \lneq y \lneq x_{i+1}$. Of course, if the order relation $\leq$ is *bounded*, i.e., for each element $p \in P$, there exists a finite upper bound on the lengths of $\lneq$-paths issuing from p, then $\leq$ is neighborhood generated.

In the sequel of this section, we shall address the neighborhood problem from different perspectives. We first consider the general problem of existence of neighbors, and then provide means for the computation of all upper neighbors and of all lower neighbors, respectively, in the cases where these exist. As it will turn out, neighbors only exist for all concept descriptions in the description logic $\mathcal{EL}$ without any TBox. The presence of either a TBox or of the bottom concept description $\bot$ prevents the existence of neighbors for some concept descriptions. Furthermore, the extensions of $\mathcal{EL}$ with greatest fixed-point semantics also allow for the construction of concept descriptions that do not possess neighbors. Eventually, a complexity analysis shows that deciding neighborhood in $\mathcal{EL}$ is in P, and that all upper neighbors of an $\mathcal{EL}$ concept description can be computed in deterministic polynomial time.

Definition 1. *Consider a signature Σ, let $\mathcal{T}$ be a TBox over Σ, and further assume that C and D are concept descriptions over Σ. Then, C is a* lower neighbor *or a* most general strict subsumee *of D with respect to $\mathcal{T}$, denoted as $\mathcal{T} \models C \prec D$, if the following statements hold true.*

1. $\mathcal{T} \models C \sqsubsetneq D$
2. *For each concept description E over Σ, it holds true that $\mathcal{T} \models C \sqsubseteq E \sqsubseteq D$ implies $\mathcal{T} \models E \equiv C$ or $\mathcal{T} \models E \equiv D$.*

Additionally, we then also say that D is an upper neighbor *or a* most specific strict subsumer *of C with respect to $\mathcal{T}$, and we may also write $\mathcal{T} \models D \succ C$.*

We first observe that neighborhood of concept descriptions is not preserved by the concept constructors. It is easy to see that $\emptyset \models A \sqcap B \prec A$. However, it holds true that $\emptyset \models \exists\, r.\, (A \sqcap B) \sqsubsetneq \exists\, r.\, A \sqcap \exists\, r.\, B \sqsubsetneq \exists\, r.\, A$, which shows $\emptyset \not\models \exists\, r.\, (A \sqcap B) \prec \exists\, r.\, A$. Furthermore, we have that $\emptyset \models A \sqcap B \sqcap (A \sqcap B) \equiv A \sqcap (A \sqcap B)$, and consequently $\emptyset \not\models A \sqcap B \sqcap (A \sqcap B) \prec A \sqcap (A \sqcap B)$. There are according counterexamples when neighborhood with respect to a non-empty TBox is considered.

It is easily verified that neighborhood with respect to the empty TBox $\emptyset$ does not coincide with neighborhood w.r.t. a non-empty TBox $\mathcal{T}$. For instance, $\emptyset \models A \prec \top$ holds true, but $\{\top \sqsubseteq A\} \models A \equiv \top$. For the converse direction, consider the counterexample where $\{A \sqsubseteq B,\ B \sqsubseteq A\} \models A \sqcap B \prec \top$ and $\emptyset \models A \sqcap B \sqsubsetneq A \sqsubsetneq \top$.

3.1 The Empty TBox

Since Baader and Morawska showed in [4, Proof of Proposition 3.5] that $\sqsubseteq_\emptyset$ is bounded, we can immediately draw the following conclusion.

Proposition 2. *The subsumption relation $\sqsubseteq_\emptyset$ is neighborhood generated.*

After this first promising result, we continue with describing the neighborhood relation $\prec_\emptyset$. For this purpose, we define $\mathsf{Upper}(C) := \{\, D \mid C \prec_\emptyset D \,\}$ as the set of all upper neighbors of a concept description C, and accordingly $\mathsf{Lower}(C)$ contains exactly all lower neighbors of C.

There is a well-known recursive characterization of $\sqsubseteq_\emptyset$ as follows: $C \sqsubseteq_\emptyset D$ if, and only if, $A \in \mathsf{Conj}(D)$ implies $A \in \mathsf{Conj}(C)$ for each concept name A, and for each $\exists\, r.\, F \in \mathsf{Conj}(D)$, there is some $\exists\, r.\, E \in \mathsf{Conj}(C)$ such that $E \sqsubseteq_\emptyset F$. With the help of that we can prove that there is the following necessary condition for neighboring concept descriptions.

Lemma 3. *Let C and D be some reduced $\mathcal{EL}$ concept descriptions over a signature Σ. If $\emptyset \models C \prec D$, then exactly one of the following statements holds true.*

1. *There is a concept name $A \in \Sigma_{\mathsf{C}}$ such that $\emptyset \models C \equiv D \sqcap A$.*
2. *It holds true that $\mathsf{Conj}(C) \cap \Sigma_{\mathsf{C}} = \mathsf{Conj}(D) \cap \Sigma_{\mathsf{C}}$, and there is exactly one existential restriction $\exists\, r.\, E \in \mathsf{Conj}(C)$ such that for all existential restrictions $\exists\, s.\, F \in \mathsf{Conj}(D)$, it holds true that $r = s$ and $\emptyset \models E \sqsubseteq F$ imply $\emptyset \not\models E \equiv F$.*

By means of the previous lemma we can deduce the following two propositions that describe the sets $\mathsf{Upper}(C)$ and $\mathsf{Lower}(C)$.

Proposition 4. *For each reduced $\mathcal{EL}$ concept description C over some signature Σ, the following recursive equation is satisfied (modulo equivalence).*

$$\begin{aligned}\mathsf{Upper}(C) = \{ \textstyle\bigsqcap \mathsf{Conj}(C) \setminus \{A\} \mid A \in \mathsf{Conj}(C) \} \\ \cup \{ \textstyle\bigsqcap \mathsf{Conj}(C) \setminus \{\exists\, r.\, E\} \cup \{ \exists\, r.\, F \mid F \in \mathsf{Upper}(E) \} \mid \exists\, r.\, E \in \mathsf{Conj}(C) \}\end{aligned}$$

For instance, consider the concept description $A \sqcap \exists\, r.\, B \sqcap \exists\, s.\, (A \sqcap B)$. It is in reduced form and has three upper neighbors, namely $\exists\, r.\, B \sqcap \exists\, s.\, (A \sqcap B)$, $A \sqcap \exists\, r.\, \top \sqcap \exists\, s.\, (A \sqcap B)$, and $A \sqcap \exists\, r.\, B \sqcap \exists\, s.\, A \sqcap \exists\, s.\, B$.

Proposition 5. *For every $\mathcal{EL}$ concept description C over some signature Σ, the following equation is satisfied (modulo equivalence).*

$$\begin{aligned}\mathsf{Lower}(C) = \{ C \sqcap A \mid A \in \Sigma_{\mathsf{C}} \text{ and } \emptyset \not\models C \sqsubseteq A \} \\ \cup \left\{ C \sqcap \exists\, r.\, D \;\middle|\; \begin{array}{l} r \in \Sigma_{\mathsf{R}},\ D \text{ is most general such that } \emptyset \not\models C \sqsubseteq \exists\, r.\, D, \\ \text{and } \emptyset \models C \sqsubseteq \exists\, r.\, E \text{ for all } E \text{ with } \emptyset \models D \prec E \end{array} \right\}\end{aligned}$$

While the recursive characterization of Upper in Proposition 4 immediately yields a procedure for enumerating all upper neighbors of a given concept description, the situation is not that apparent for lower neighbors. We can, however, constitute a procedure for computing lower neighbors by means of Proposition 5. Let C be an $\mathcal{EL}$ concept description over some signature Σ. Proceed as follows.

1. For each concept name $A \in \Sigma_{\mathsf{C}}$ with $\emptyset \not\models C \sqsubseteq A$, output $C \sqcap A$ as a lower neighbor of C.
2. For each role name $r \in \Sigma_{\mathsf{R}}$, recursively proceed as follows.
   (a) Let $D \coloneqq \top$.
   (b) While $\emptyset \models C \sqsubseteq \exists\, r.\, D$, replace D with a lower neighbor of D.
   (c) If $\emptyset \models C \sqsubseteq \exists\, r.\, E$ for all E with $\emptyset \models D \prec E$, then output $C \sqcap \exists\, r.\, D$ as a lower neighbor of C.

Eventually, we finish our investigations of $\prec_{\emptyset}$ with a complexity analysis. Using induction on the role depth of C, we can prove that, for each reduced $\mathcal{EL}$ concept description C, the set $\mathsf{Upper}(C)$ can be computed in deterministic time $\mathcal{O}(||C||^2)$. It is then apparent that the following proposition holds true.

Proposition 6. *The neighborhood relation $\prec_{\emptyset}$ can be decided in polynomial time, that is, $\prec_{\emptyset} \in \mathbf{P}$, and the mapping Upper is computable in deterministic polynomial time.*

3.2 The Bottom Concept Description

Now consider the extension of $\mathcal{EL}$ with the *bottom concept description* $\bot$ the semantics of which is defined as $\bot^{\mathcal{I}} \coloneqq \emptyset$ for any interpretation $\mathcal{I}$. Then $\sqsubseteq_\emptyset$ is not bounded, since the following infinite chain exists.

$$\emptyset \models \bot \sqsubsetneq \ldots \sqsubsetneq (\exists\, r.)^{n+1}\, \top \sqsubsetneq (\exists\, r.)^{n}\, \top \sqsubsetneq \ldots \sqsubsetneq \exists\, r.\, \exists\, r.\, \top \sqsubsetneq \exists\, r.\, \top \sqsubsetneq \top$$

However, $\sqsupseteq_\emptyset$ is still well-founded, since whenever a chain starts with $\bot$, then the second element must be a satisfiable concept description, that is, some C with $C \not\equiv_\emptyset \bot$, after which the chain can only have a bounded number of elements. Furthermore, $\sqsubseteq_\emptyset$ is not neighborhood generated, as $\bot$ does not have any upper neighbors. To see this, consider a concept description C such that $\bot \sqsubsetneq_\emptyset C$; it then follows that $\bot \sqsubsetneq_\emptyset C \sqcap \exists\, r.\, C \sqsubsetneq_\emptyset C$.

3.3 A Non-Empty TBox

A similar situation arises when considering subsumption with respect to a non-empty TBox $\mathcal{T}$. Firstly, consider the simple signature Σ where $\Sigma_{\mathsf{C}} \coloneqq \{A\}$ and $\Sigma_{\mathsf{R}} \coloneqq \{r\}$ and define the TBox $\mathcal{T} \coloneqq \{\top \sqsubseteq \exists\, r.\, \top,\ A \sqsubseteq \exists\, r.\, A\}$. We obtain that the quotient $\mathcal{EL}(\Sigma)/{\equiv_\mathcal{T}}$ consists of the classes $[(\exists\, r.)^n\, A]$ for $n \in \mathbb{N}$, and of the class $[\top]$. Furthermore, these concept descriptions are linearly ordered as follows.

$$\mathcal{T} \models A \sqsubsetneq \exists\, r.\, A \sqsubsetneq \exists\, r.\, \exists\, r.\, A \sqsubsetneq \exists\, r.\, \exists\, r.\, \exists\, r.\, A \sqsubsetneq \ldots \sqsubsetneq \top$$

Consequently, $\top$ does not have lower neighbors, and we also conclude that $\sqsubseteq_\mathcal{T}$ is not bounded and $\sqsupseteq_\mathcal{T}$ is not well-founded.

Secondly, we show that there is a TBox $\mathcal{T}$ and a concept description without any upper neighbors with respect to $\sqsubseteq_\mathcal{T}$. We try to keep things simple, and consider a rather small signature, namely Σ defined by $\Sigma_{\mathsf{C}} \coloneqq \{A, B\}$ and $\Sigma_{\mathsf{R}} \coloneqq \{r\}$, and we define a TBox by $\mathcal{T} \coloneqq \{\exists\, r.\, A \sqsubseteq A,\ B \sqsubseteq A,\ B \equiv \exists\, r.\, B\}$. It can be shown that, for each $\mathcal{EL}(\Sigma)$ concept description C, either C is equivalent to B w.r.t. $\mathcal{T}$ or there exists an $n \in \mathbb{N}$ such that $\mathcal{T} \models B \sqsubsetneq (\exists\, r.)^n\, A \sqsubsetneq C$, i.e., B does not have upper neighbors with respect to $\mathcal{T}$.

Proposition 7. *There is some TBox $\mathcal{T}$ for which the subsumption relation $\sqsubseteq_\mathcal{T}$ is not neighborhood generated.*

3.4 Greatest Fixed-Point Semantics

Unfortunately, the situation is also not rosy for extensions of $\mathcal{EL}$ with *greatest fixed-point semantics* [1,11]. It then also holds true that $\sqsubseteq_\emptyset$ is neither bounded nor neighborhood generated, and $\sqsupseteq_\emptyset$ is not well-founded. One culprit is a concept description which represents a cycle, for instance $\nu\, X.\, \exists\, r.\, X$, the extension of which is maximal w.r.t. the property of containing elements that have some other element in that extension as an r-successor.

4 The Distributive, Graded Lattice of $\mathcal{EL}$ Concept Descriptions

The goal of this section is to explore the properties of the lattice of $\mathcal{EL}$ concept descriptions ordered by subsumption with respect to the empty TBox. In particular, Blyth [5, Chapters 4 and 5] shows that it suffices to investigate whether this lattice is distributive and of locally finite length, such that as an immediate corollary we then obtain that also the Jordan-Dedekind chain condition is satisfied, which states that for each pair $C \sqsubseteq_\emptyset D$, all maximal chains in the intervall $[C, D]$ have the same length. Furthermore, this length can then be utilized to define a distance between C and D, and in particular to measure a distance from each concept description C to the top concept description $\top$, which we call the rank of C.

Lemma 8. *For each signature Σ, the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$ is distributive, i.e., for all concept descriptions $C, D, E \in \mathcal{EL}(\Sigma)$, it holds true that*

$$\begin{aligned} &\emptyset \models C \sqcap (D \vee E) \equiv (C \sqcap D) \vee (C \sqcap E), \\ \textit{and} \quad &\emptyset \models C \vee (D \sqcap E) \equiv (C \vee D) \sqcap (C \vee E). \end{aligned}$$

Lemma 9. *For each signature Σ, the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$ is of locally finite length, that is, for all concept descriptions $C, D \in \mathcal{EL}(\Sigma)$ with $\emptyset \models C \sqsubseteq D$, every chain in the interval $[C, D]$ has a finite length.*

According to Blyth [5, Chapters 4 and 5], the following statements are obtained as immediate consequences of Lemmas 8 and 9.

Corollary 10. *1. For each signature Σ, the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$ is modular, i.e., for all concept descriptions $C, D, E \in \mathcal{EL}(\Sigma)$, it holds true that*

$$\begin{aligned} &\emptyset \models (C \sqcap D) \vee (C \sqcap E) \equiv C \sqcap (D \vee (C \sqcap E)), \\ &\emptyset \models (C \vee D) \sqcap (C \vee E) \equiv C \vee (D \sqcap (C \vee E)), \\ &\emptyset \models C \sqsubseteq D \quad \textit{implies} \quad \emptyset \models C \vee (E \sqcap D) \equiv (C \vee E) \sqcap D, \\ \textit{and} \quad &\emptyset \models C \sqsupseteq D \quad \textit{implies} \quad \emptyset \models C \sqcap (E \vee D) \equiv (C \sqcap E) \vee D. \end{aligned}$$

2. For each signature Σ, the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$ is both upper and lower semi-modular, i.e., for all concept descriptions $C, D \in \mathcal{EL}(\Sigma)$, it holds true that

$$\emptyset \models C \sqcap D \prec C \quad \textit{if, and only if,} \quad \emptyset \models D \prec C \vee D.$$

3. For each signature Σ, the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$ satisfies the Jordan-Dedekind chain condition*, i.e., for all concept descriptions $C, D \in \boldsymbol{\mathcal{EL}}(\Sigma)$ with $\emptyset \models C \sqsubsetneq D$, it holds true that all maximal chains in the interval $[C, D]$ have the same length.*

The notion of a rank function can be defined for ordered sets. The following definition specifically tailors this notion for the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$.

Definition 11. *An $\mathcal{EL}$* rank function *is a mapping $|\cdot| \colon \mathcal{EL}(\Sigma) \to \mathbb{N}$ with the following properties.*

1. $|\top| = 0$
2. $\emptyset \models C \equiv D$ *implies* $|C| = |D|$ *(equivalence closed)*
3. $\emptyset \models C \sqsubsetneq D$ *implies* $|C| \gneq |D|$ *(strictly order preserving)*
4. $\emptyset \models C \prec D$ *implies* $|C| + 1 = |D|$ *(neighborhood preserving)*

For an $\mathcal{EL}$ concept description C, we say that $|C|$ is the rank *of C.*

Proposition 12. *For each $C \in \mathcal{EL}(\Sigma)$, let $|C| \coloneqq 0$ if $\emptyset \models C \equiv \top$, and otherwise set*

$$|C| \coloneqq \mathsf{max}\{\, n+1 \mid \exists\, D_1, \ldots, D_n \in \mathcal{EL}(\Sigma) \colon \emptyset \models C \prec D_1 \prec \ldots \prec D_n \prec \top \,\}.$$

Then, $|\cdot|$ is an $\mathcal{EL}$ rank function.

Since $\boldsymbol{\mathcal{EL}}(\Sigma)$ satisfies the Jordan-Dedekind chain condition, we infer that in order to compute the rank $|C|$ of an $\mathcal{EL}$ concept description C over Σ with $\emptyset \not\models C \equiv \top$, we simply need to find *one* chain $\emptyset \models C \prec D_1 \prec D_2 \prec \ldots \prec D_n \prec \top$, and then it follows that $|C| = n + 1$. Furthermore, $|C| = 0$ if $\emptyset \models C \equiv \top$.

Corollary 13. *For each signature Σ, the lattice $\boldsymbol{\mathcal{EL}}(\Sigma)$ is graded.*

The next lemma provides an equation for the rank of a conjunction of n concept descriptions. By induction over n, it follows from Lemma 9, Corollary 10, and [5, Theorem 4.6].

Lemma 14. *Let $\mathcal{C}$ be a set of n $\mathcal{EL}$ concept descriptions over Σ. Then, the following equation holds true.*

$$|\bigsqcap \mathcal{C}| = \sum_{i=1}^{n} (-1)^{i+1} \cdot \sum_{\mathcal{D} \in \binom{\mathcal{C}}{i}} |\bigvee \mathcal{D}|$$

Let $C = A_1 \sqcap \ldots \sqcap A_m \sqcap \exists\, r_1.\, C_1 \sqcap \ldots \sqcap \exists\, r_n.\, C_n$ be a reduced $\mathcal{EL}$ concept description. Then its rank can be computed as follows, cf. Lemma 14.

$$\begin{aligned} |C| &= |A_1 \sqcap \ldots \sqcap A_m \sqcap \exists\, r_1.\, C_1 \sqcap \ldots \sqcap \exists\, r_n.\, C_n| \\ &= |A_1 \sqcap \ldots \sqcap A_m| + |\exists\, r_1.\, C_1 \sqcap \ldots \sqcap \exists\, r_n.\, C_n| - |\top| \\ &= m + |\exists\, r_1.\, C_1 \sqcap \ldots \sqcap \exists\, r_n.\, C_n| \end{aligned}$$

Furthermore, it holds true that $\emptyset \models \exists\, r.\, C \vee \exists\, s.\, D \equiv \top$ if $r \neq s$. It follows that we can further simplify the rank computation as follows.

$$\begin{aligned} |\exists\, r_1.\, C_1 \sqcap \ldots \sqcap \exists\, r_n.\, C_n| &= |\bigsqcap\{\bigsqcap\{\, \exists\, r_i.\, C_i \mid i \in \{1, \ldots, n\} \text{ and } r_i = r \,\} \mid r \in \Sigma_{\mathsf{R}} \,\}| \\ &= \sum_{r \in \Sigma_{\mathsf{R}}} |\bigsqcap\{\, \exists\, r_i.\, C_i \mid i \in \{1, \ldots, n\} \text{ and } r_i = r \,\}| \end{aligned}$$

The rank of the conjunction of existential restrictions can be computed by means of Lemma 14, and finally it is readily verified that the rank of one existential restriction $\exists\, r.\, C$ satisfies the following equation.

$$|\exists\, r.\, C| = 1 + |\bigsqcap\{\, \exists\, r.\, D \mid \emptyset \models C \prec D \,\}|$$

Definition 15. *An $\mathcal{EL}$* metric *or $\mathcal{EL}$* distance function *is a mapping $\mathsf{d} \colon \mathcal{EL}(\Sigma) \times \mathcal{EL}(\Sigma) \to \mathbb{N}$ with the following properties.*

1. $\mathsf{d}(C, D) \geq 0$ *(non-negative)*
2. $\mathsf{d}(C, D) = 0$ *if, and only if, $\emptyset \models C \equiv D$* *(equivalence closed)*

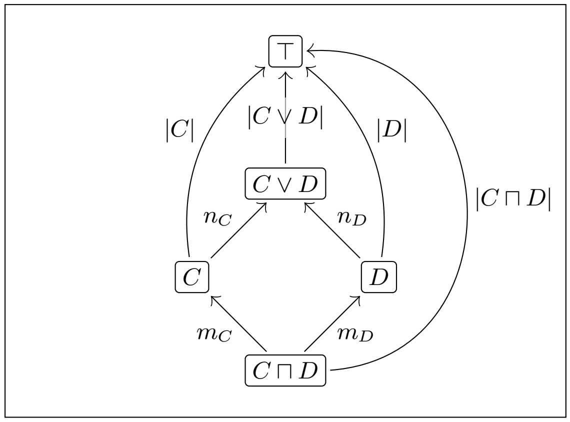

Fig. 1. Obtaining a distance function from the rank function

3. $\mathsf{d}(C, D) = \mathsf{d}(D, C)$ *(symmetric)*
4. $\mathsf{d}(C, E) \leq \mathsf{d}(C, D) + \mathsf{d}(D, E)$ *(triangle inequality)*

We then also say that $\mathsf{d}(C, D)$ *is the* distance *between* C *and* D*.*

Lemma 14 for the case $n = 2$ yields that in the rectangle shown in Figure 1 opposite edges have the same length, where length means length of a maximal chain between the endpoints. It is easy to see that $|C \sqcap D| = |C| + m_C = |D| + m_D$ and $|C \vee D| = |C| - n_C = |D| - n_D$. Thus, we infer that $m_C = |C \sqcap D| - |C| = |D| - |C \vee D| = n_D$, and similarily that $m_D = n_C$. Consequently, we can define an $\mathcal{EL}$ distance function in the following way.

Proposition 16. *For all* $C, D \in \mathcal{EL}(\Sigma)$*, define*

$$\mathsf{d}(C, D) := |C \sqcap D| - |C \vee D|.$$

Then, d *is an* $\mathcal{EL}$ *metric.*

We can justify the name of a distance function as follows. If we consider the graph of $\mathcal{EL}$ concept descriptions such that edges exist exactly between neighboring concept descriptions, that is, if we consider the graph $(\mathcal{EL}(\Sigma), \prec_\emptyset \cup \succ_\emptyset)$, then the distance $\mathsf{d}(C, D)$ is the length of a shortest path between C and D.

Corollary 17. $\boldsymbol{\mathcal{EL}}(\Sigma)$ *is a metric lattice, i.e., a lattice which is also a metric space.*

It is easy to verify that $\boldsymbol{\mathcal{EL}}(\Sigma)$ is complete, not bounded, not precompact, not compact, locally compact, proper if Σ is finite, neither connected nor path connected, and separable. The induced topology τ_d is perfectly normal Hausdorff or T_6. Furthermore, all subsets are both open and closed, and all mappings $f\colon \boldsymbol{\mathcal{EL}}(\Sigma) \to (X, \mathsf{d}')$ are continuous.

Lemma 18. *Let* $C \in \mathcal{EL}(\Sigma)$*, then* $\mathsf{d}(C, \bigvee \mathsf{Upper}(C)) = |\mathsf{Upper}(C)|$*.*

According the the previous lemma, we can compute ranks as follows.

1. Let $D := C$ and $r := 0$.

2. While $\emptyset \not\models D \equiv \top$, compute the set $\mathsf{Upper}(D)$ of upper neighbors of D, set $r \coloneqq r + |\mathsf{Upper}(D)|$ and $D \coloneqq \bigvee \mathsf{Upper}(D)$.
3. Return r.

In [6] Ecke, Peñaloza, and Turhan defined the notion of a concept similarity measure as a function of type $\mathcal{EL}(\Sigma) \times \mathcal{EL}(\Sigma) \to [0,1]$, and then considered so-called *relaxed instances* of concept descriptions with respect to ontologies. Simply speaking, a is a relaxed instance of C if there is a concept that is similar enough to C and has a as an instance. It is straight-forward to consider these relaxed instances also with respect to the distance function we have just introduced. More formally, we define them as follows.

Definition 19. *Consider an interpretation $\mathcal{I}$ over some signature Σ and a concept description $C \in \mathcal{EL}(\Sigma)$, and let $n \in \mathbb{N}$. Then, the expression $\mathsf{C} \leq n. C$ is called a* relaxed concept description*, and its extension is defined by*

$$(\mathsf{C} \leq n. C)^{\mathcal{I}} \coloneqq \bigcup \{\, D^{\mathcal{I}} \mid D \in \mathcal{EL}(\Sigma) \text{ and } \mathsf{d}(C, D) \leq n \,\}.$$

Suppose that $\mathcal{O}$ is an ontology over some signature Σ, and further let $a \in \Sigma_{\mathsf{I}}$ be an individual name, $C \in \mathcal{EL}(\Sigma)$ a concept description, and $n \in \mathbb{N}$. We then say that a is a relaxed instance *of C with respect to $\mathcal{O}$ and distance threshold n, denoted as $\mathcal{O} \models a \sqsubseteq \mathsf{C} \leq n. C$, if it holds true that $a^{\mathcal{I}} \in (\mathsf{C} \leq n. C)^{\mathcal{I}}$ for each model $\mathcal{I}$ of $\mathcal{O}$.*

For transforming our distance function d into a similarity function $\mathsf{s}\colon \mathcal{EL}(\Sigma) \times \mathcal{EL}(\Sigma) \to [0,1]$ we can proceed as follows. We begin with transforming d into a metric with range $[0,1)$. For that purpose, we choose an order-preserving, sub-additive function $f\colon [0,\infty) \to [0,1)$ with $\ker(f) = \{0\}$. Note that a function $f\colon [0,\infty) \to \mathbb{R}$ is sub-additive if $f'' < 0$ and $f(0) = 0$. Then $f \circ \mathsf{d}$ is such a metric with range $[0,1)$. Suitable functions are the following.

- $f\colon x \mapsto \frac{x}{1+x}$ or more generally $f\colon x \mapsto (\frac{x}{1+x})^y$ for $y > 0$
- $f\colon x \mapsto 1 - \frac{1}{2^x}$ or more generally $f\colon x \mapsto 1 - y^x$ for $y \in (0,1)$

Then, $\mathsf{s} \coloneqq 1 - f \circ \mathsf{d}$ is a similarity function on $\mathcal{EL}(\Sigma)$. It is easy to verify that then s satifies the following properties which have been defined by Lehmann and Turhan in [10], for all $\mathcal{EL}$ concept descriptions C, D, E over Σ.

1. $\mathsf{s}(C,D) = \mathsf{s}(D,C)$ (symmetric)
2. $1 + \mathsf{s}(C,D) \geq \mathsf{s}(C,E) + \mathsf{s}(E,D)$ (triangle inequality)
3. $\emptyset \models C \equiv D$ implies $\mathsf{s}(C,E) = \mathsf{s}(D,E)$ (equivalence invariant)
4. $\emptyset \models C \equiv D$ if, and only if, $\mathsf{s}(C,D) = 1$ (equivalence closed)
5. $\emptyset \models C \sqsubseteq D \sqsubseteq E$ implies $\mathsf{s}(C,D) \geq \mathsf{s}(C,E)$ (subsumption preserving)
6. $\emptyset \models C \sqsubseteq D \sqsubseteq E$ implies $\mathsf{s}(C,E) \leq \mathsf{s}(D,E)$ (reverse subsumption preserving)

However, as it turns out such a similarity measure $1 - f \circ \mathsf{d}$ does not satisfy the property of *structural dependance*. For instance, consider a signature Σ without role names and such that $\Sigma_C \coloneqq \{A\} \cup \{\, B_n \mid n \in \mathbb{N} \,\}$. It is now readily verified that

$$(1 - f \circ \mathsf{d})(A \sqcap \bigsqcap \{\, B_\ell \mid \ell \leq n \,\}, \bigsqcap \{\, B_\ell \mid \ell \leq n \,\}) = 1 - f(1)$$

for all $n \in \mathbb{N}$, and since $f(1) > 0$ we conclude that the sequence does not converge to 1 for $n \to \infty$.

For extending our rank function $|\cdot|$ and our distance function d to $\mathcal{EL}^\bot$, we can simply define $|\bot| \coloneqq \infty$, $\mathsf{d}(\bot,\bot) \coloneqq 0$, and $\mathsf{d}(\bot, C) \coloneqq \mathsf{d}(C,\bot) \coloneqq \infty$ for $\emptyset \models C \not\equiv \bot$. When transforming the extended metric into a similarity measure then two concept descriptions have a similarity of 0 if, and only if, exactly one of them is unsatisfiable. In $\mathcal{EL}$ without the bottom concept description $\bot$, a similarity of 0 can never occur when utilizing the above construction.

We close this section with some first investigations on the complexities of decision problems and computation problems related to the introduced rank function. So far, it is unknown whether the rank function can be tractably computed, i.e., in deterministic polynomial time. However, if $|C|$ is computed in the naïve way by constructing an arbitrary chain of neighbors from C to $\top$ and then determinining its length, at least deterministic exponential time with respect to the size of C is necessary. To see this, consider the concept description $C_n \coloneqq \exists\, r. \bigsqcap\{A_1, \dots, A_n\}$ for each $n \in \mathbb{N}$. It is well-known that there are exponentially many subsets of $\{A_1, \dots, A_n\}$ with $\lfloor \frac{n}{2} \rfloor$ elements; let $X_1, \dots, X_\ell$ be an enumeration of these, and define $D_m \coloneqq \bigsqcap\{\, \exists\, r. \bigsqcap X_i \mid i \in \{m, \dots, \ell\}\,\}$. Clearly, then $C_n \sqsubsetneq_\emptyset D_1 \sqsubsetneq_\emptyset D_2 \sqsubsetneq_\emptyset \cdots \sqsubsetneq_\emptyset D_\ell \sqsubsetneq_\emptyset \top$ is an exponentially long chain of strict subsumptions. We conclude that $|C_n|$ is at least exponential in n.

Given a concept description C and a natural number n (in binary encoding), then we can decide in triple exponential time whether the rank of C is equal to n, at most n, or at least n. A procedure can construct a chain of n neighbors and then check whether $\top$ is reached. If n is fixed, then this requires only deterministic polynomial time.

5 Conclusion

We have investigated the *neighborhood problem* for the description logic $\mathcal{EL}$ and some of its variants. We found that existence of neighbors can in general only be guaranteed for the case of $\mathcal{EL}$ without a TBox, without the bottom concept description, and without greatest fixed-point semantics. The presence of a TBox, the bottom concept description, or greatest fixpoint semantics allow for the construction of concept descriptions that do not have neighbors in certain directions. For the case of $\mathcal{EL}$ we proposed sound and complete procedures for deciding neighborhood as well as for computing all upper neighbors and all lower neighbors, respectively. Furthermore, we have shown that deciding neighborhood and computing all upper neighbors requires only deterministic polynomial time. The complexity of computing all lower neighbors is currently an open question; possibly there exists a less expensive procedure than the one presented here.

As further results, we have proven that the lattice of $\mathcal{EL}$ concept descriptions is distributive, modular, graded, and metric. In particular, this means that there exists a rank function as well as a distance function on this lattice. Some first complexity results on problems related to these rank and distance functions were found. However, the exact complexities are currently unknown; we do not know whether the presented upper bounds are sharp, and lower bounds are also not available. This implies that there could possibly exist faster procedures for computing ranks and distances than those introduced in this document. In particular, better formulas for computing or approximating ranks of $\mathcal{EL}$ concept description should be sought. Some initial experiments could lead to the claim that the rank of an $\mathcal{EL}$ concept description with role depth n, that is, for which the nesting depth of existential quantifiers is n, could be n-exponential in the size of C.

As an important consequence we infer that the algorithm *NextClosures* [8] can be utilized for enumerating canonical bases of closure operators in $\mathcal{EL}$.

Other possible future research could consider extensions to more expressive description logics. Of course, these logics should be considered without TBoxes for deciding existence of neighbors in general. Eventually, a further direction for future research is a more fine-grained characterization of existence of neighbors. That is, given a description logic where neighbors need not exist in general, how can we decide whether a concept description has neighbors and how can we enumerate these?

Acknowledgements The author gratefully thanks both Franz Baader and Bernhard Ganter for fruitful discussions on neighborhood generated orders, and furthermore thanks the anonymous reviewers for their constructive hints and helpful remarks.

References

1. Baader, F.: Terminological Cycles in a Description Logic with Existential Restrictions. In: Gottlob, G., Walsh, T. (eds.) IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003. pp. 325–330. Morgan Kaufmann (2003)
2. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ Envelope. In: Kaelbling, L.P., Saffiotti, A. (eds.) IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005. pp. 364–369. Professional Book Center (2005)
3. Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017)
4. Baader, F., Morawska, B.: Unification in the Description Logic $\mathcal{EL}$. Logical Methods in Computer Science **6**(3) (2010)
5. Blyth, T.S.: Lattices and Ordered Algebraic Structures. Universitext, Springer London, 1st edn. (2005)
6. Ecke, A., Peñaloza, R., Turhan, A.: Similarity-based Relaxed Instance Queries. Journal of Applied Logic **13**(4), 480–508 (2015)
7. Hitzler, P., Krötzsch, M., Rudolph, S.: Foundations of Semantic Web Technologies. Chapman and Hall/CRC Press (2010)
8. Kriegel, F.: NextClosures with Constraints. In: Huchard, M., Kuznetsov, S. (eds.) Proceedings of the Thirteenth International Conference on Concept Lattices and Their Applications, Moscow, Russia, July 18-22, 2016. CEUR Workshop Proceedings, vol. 1624, pp. 231–243. CEUR-WS.org (2016)
9. Küsters, R.: Non-Standard Inferences in Description Logics, Lecture Notes in Computer Science, vol. 2100. Springer (2001)
10. Lehmann, K., Turhan, A.: A Framework for Semantic-Based Similarity Measures for $\mathcal{ELH}$-Concepts. In: del Cerro, L.F., Herzig, A., Mengin, J. (eds.) Logics in Artificial Intelligence - 13th European Conference, JELIA 2012, Toulouse, France, September 26-28, 2012. Proceedings. Lecture Notes in Computer Science, vol. 7519, pp. 307–319. Springer (2012)
11. Lutz, C., Piro, R., Wolter, F.: Enriching $\mathcal{EL}$-Concepts with Greatest Fixpoints. In: Coelho, H., Studer, R., Wooldridge, M. (eds.) ECAI 2010 - 19th European Conference on Artificial Intelligence, Lisbon, Portugal, August 16-20, 2010, Proceedings. Frontiers in Artificial Intelligence and Applications, vol. 215, pp. 41–46. IOS Press (2010)