Conjunctive Query Entailment: Decidable in Spite of $\mathcal{O}$, $\mathcal{I}$, and $\mathcal{Q}$

Birte Glimm[1] and Sebastian Rudolph[2]

[1] University of Oxford, UK
[2] Institut AIFB, Universität Karlsruhe, DE

1 Introduction

We present a decidability result for entailment of conjunctive queries (CQs) in the very expressive Description Logic (DL) $\mathcal{ALCHOIQ}b$ [1] by establishing finite representability of countermodels in the case of non-entailment. Our result also generalizes to unions of conjunctive queries and $\mathcal{SHOIQ}$ provided the query contains only simple roles, and we are confident that the technique extends to $\mathcal{SROIQ}$ under the simple roles restriction as well. Full proofs and additional material can be found in the accompanying technical report [2].

Throughout the paper, we use the DL $\mathcal{ALCOIFb}$, where b stands for safe Boolean role expressions. Any $\mathcal{ALCHOIQ}b$ knowledge base (KB) can be polynomially reduced to an $\mathcal{ALCOIFb}$ KB with extended signature, while preserving query (non-)entailment [3]. W.l.o.g., we assume that KBs contain an empty ABox (with nominals the ABox can be internalized) and that all GCIs are *simplified* to one of the following forms:

$$\bigsqcap A_i \sqsubseteq \bigsqcup B_j \ \mid \ A \equiv \{o\} \ \mid \ A \sqsubseteq \forall U.B \ \mid \ A \sqsubseteq \exists U.B \ \mid \ \mathsf{func}(f), \qquad (1)$$

where $A_{(i)}$ and $B_{(j)}$ are atomic concepts, o is an individual name, U is a safe Boolean role expression, and f is a role that is declared functional. If $i = 0$, we interpret $\bigsqcap A_i$ as $\top$ and if $j = 0$, we interpret $\bigsqcup B_j$ as $\bot$. We use $\mathsf{con}(\mathcal{K})$, $\mathsf{rol}(\mathcal{K})$, and $\mathsf{nom}(\mathcal{K})$ to denote, respectively, the set of concept, role, and individual names occurring in $\mathcal{K}$, and $\mathsf{cl}(\mathcal{K})$ to denote the closure of $\mathcal{K}$. A role f is *(inverse) functional in* $\mathcal{K}$ if $\mathcal{K}$ contains an axiom $\mathsf{func}(f)$ ($\mathsf{func}(f^-)$).

Let N_V be a countably infinite set of variables, A a concept name, r a role name, and $x, y \in N_V$ variables. An *atom* is an expression $A(x)$ or $r(x, y)$. A *Boolean conjunctive query* q is a non-empty set of atoms. We use $\mathsf{var}(q)$ to denote the set of (existentially quantified) variables occurring in q and $\sharp(q)$ for the number of atoms in q. For $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ an interpretation, $A(x), r(x, y)$ atoms, and $\pi \colon \mathsf{var}(q) \to \Delta^{\mathcal{I}}$ a total function, we write (i) $\mathcal{I} \models^{\pi} A(x)$ if $\pi(x) \in A^{\mathcal{I}}$ and (ii) $\mathcal{I} \models^{\pi} r(x, y)$ if $\langle \pi(x), \pi(y) \rangle \in r^{\mathcal{I}}$. If $\mathcal{I} \models^{\pi} \mathsf{At}$ for all atoms $\mathsf{At} \in q$, we write $\mathcal{I} \models^{\pi} q$ and say that $\mathcal{I}$ *satisfies* q. We write $\mathcal{I} \models q$ if there exists a function π such that $\mathcal{I} \models^{\pi} q$ and call π a *match* for q in $\mathcal{I}$. If $\mathcal{I} \models \mathcal{K}$ implies $\mathcal{I} \models q$, we say that $\mathcal{K}$ *entails* q and write $\mathcal{K} \models q$. W.l.o.g., we assume that queries are connected. Given a KB $\mathcal{K}$ and a CQ q, the *query entailment problem* is to decide whether $\mathcal{K} \models q$.

As a running example, we use a KB $\mathcal{K}$ containing the axioms $\{o\} \sqsubseteq \exists r.A, A \sqsubseteq \exists r.A, A \sqsubseteq \exists s.B, B \sqsubseteq C \sqcup D, C \sqsubseteq \exists f.E, D \sqsubseteq \exists g.E, E \sqsubseteq B \sqcup \{o\}, \mathsf{func}(f^-), \mathsf{func}(g^-)$. The left hand side of Fig. 1 (p. 6) displays a model for $\mathcal{K}$.

Unless stated otherwise, we use A for a concept name, r for a role, f for a functional or inverse functional role, U for a safe Boolean role expression, o for a nominal, q for a connected Boolean conjunctive query, $\mathcal{K}$ for a simplified $\mathcal{ALCOIFb}$ knowledge base, and $\mathcal{I}$ for an interpretation $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$.

We show decidability by proving that non-entailment is always witnessed by a regular model which is finitely representable. Our procedure enumerates these finite representations and terminates if the entailment does not hold. If it holds, termination can be ensured because first-order logic is recursively enumerable [4]. To construct finitely representable models, we first "unravel" models into interpretations, called forest quasi-models, that have a real forest shape. We then provide a way of "collapsing" such forest quasi-models back into real models that still have a kind of forest shape, called forest model, but which have a possibly infinite set of roots. Finally, we introduce transformations that work on the unraveled forest quasi-models, and the resulting interpretations can be collapsed into forest models with only a finite set of roots. We can then adapt standard cycle detection techniques such as (tree) blocking to obtain the desired finite representations.

2 Model Construction

We first introduce interpretations and models that have a kind of forest shape. However, the notion of a forest is very weak since we do also allow for arbitrary relations between tree elements and roots.

Definition 1. *A* tree *T is a non-empty, prefix-closed subset of $\mathbb{N}^*$. A* forest *F is a subset of $R \times \mathbb{N}^*$, where R is a countable, possibly infinite set of elements $\{\rho_1, \ldots, \rho_n\}$ such that, for each $\rho \in R$, the set $\{w \mid (\rho, w) \in F\}$ is a tree. Each pair $(\rho, \varepsilon) \in F$ is called a* root *of F. For $(\rho, w), (\rho', w') \in F$, we call (ρ', w') a* successor *of (ρ, w) if $\rho' = \rho$ and $w' = w \cdot c$ for some $c \in \mathbb{N}$, where "·" denotes concatenation; (ρ', w') is a* predecessor *of (ρ, w) if $\rho' = \rho$ and $w = w' \cdot c$ for some $c \in \mathbb{N}$; (ρ', w') is a* neighbor *of (ρ, w) if (ρ', w') is a successor of (ρ, w) or vice versa. A node (ρ, w) is an* ancestor *of a node (ρ', w') if $\rho = \rho'$ and w is a prefix of w' and it is a* descendant *if $\rho = \rho'$ and w' is a prefix of w. We use $|w|$ to denote the* length *of w. The* branching degree *$d(w)$ of a node w in a tree T is the number of successors of w.*

A forest interpretation *of $\mathcal{K}$ is an interpretation $\mathcal{I}$ that satisfies:*

FI1 *$\Delta^{\mathcal{I}}$ is a forest with roots R;*

FI2 *there is a total and surjective function $\lambda \colon \mathsf{nom}(\mathcal{K}) \to R$ s.t. $\lambda(o) = (\rho, \varepsilon)$ iff $o^{\mathcal{I}} = (\rho, \varepsilon)$;*

FI3 *for each role $r \in \mathsf{rol}(\mathcal{K})$, if $\langle (\rho, w), (\rho', w') \rangle \in r^{\mathcal{I}}$, then either (a) $w = \varepsilon$ or $w' = \varepsilon$, or (b) (ρ, w) is a neighbor of (ρ', w').*

If $\mathcal{I} \models \mathcal{K}$ we say that $\mathcal{I}$ is a forest model for $\mathcal{K}$. With $\mathsf{nomFree}(\mathcal{K})$, we denote a knowledge base obtained from $\mathcal{K}$ by replacing each nominal concept $\{o\}$ with $o \in \mathsf{nom}(\mathcal{K})$ with a fresh concept name N_o. A forest quasi-interpretation for $\mathcal{K}$ is an interpretation $\mathcal{J}$ that satisfies FI1 and FI3, and the adapted version FI2′ of FI2 that there is a total and surjective function $\lambda \colon \mathsf{nom}(\mathcal{K}) \to R$ s.t. $\lambda(o) = (\rho, \varepsilon)$ iff $(\rho, \varepsilon) \in N_o^{\mathcal{J}}$ (there might be other $(\rho, w) \in \Delta^{\mathcal{J}}$ with $w \neq \varepsilon$ s.t. $(\rho, w) \in N_o^{\mathcal{J}}$). If $\mathcal{J} \models \mathsf{nomFree}(\mathcal{K})$ we say that $\mathcal{J}$ is a forest quasi-model for $\mathcal{K}$. A forest (quasi) interpretation $\mathcal{I}$ is a strict forest (quasi) interpretation if, in condition FI3, only (b) is allowed; it is a tree interpretation, if it has a single root. If there is a k such that $d(w) \leq k$ for each $(\rho, w) \in \Delta^{\mathcal{I}}$, then we say that $\mathcal{I}$ has branching degree k.

Let $\mathcal{I}, \mathcal{I}'$ be two forest interpretations of $\mathcal{K}$ with $\delta_1, \delta_2 \in \Delta^{\mathcal{I}}, \delta_1', \delta_2' \in \Delta^{\mathcal{I}'}$. The pairs $\langle \delta_1, \delta_2 \rangle, \langle \delta_1', \delta_2' \rangle$ are isomorphic w.r.t. $\mathcal{K}$, written $\langle \delta_1, \delta_2 \rangle \cong_{\mathcal{K}} \langle \delta_1', \delta_2' \rangle$ iff

- $\langle \delta_1, \delta_2 \rangle \in r^{\mathcal{I}}$ iff $\langle \delta_1', \delta_2' \rangle \in r^{\mathcal{I}'}$ for each $r \in \mathsf{rol}(\mathcal{K})$,
- $\delta_i \in A^{\mathcal{I}}$ iff $\delta_i' \in A^{\mathcal{I}'}$ for $i \in \{1, 2\}$ and each $A \in \mathsf{con}(\mathcal{K})$,
- $\delta_i = o^{\mathcal{I}}$ iff $\delta_i' = o^{\mathcal{I}'}$ for $i \in \{1, 2\}$ and each $o \in \mathsf{nom}(\mathcal{K})$.

We say that $\mathcal{I}$ and $\mathcal{I}'$ are isomorphic w.r.t. $\mathcal{K}$, written: $\mathcal{I} \cong_{\mathcal{K}} \mathcal{I}'$, if there is a bijection $\varphi : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{I}'}$ such that, for each $\delta_1, \delta_2 \in \Delta^{\mathcal{I}}$, $\langle \delta_1, \delta_2 \rangle \cong_{\mathcal{K}} \langle \varphi(\delta_1), \varphi(\delta_2) \rangle$ and δ_1 is a successor of δ_2 iff $\varphi(\delta_1)$ is a successor of $\varphi(\delta_2)$.

If clear from the context, we omit the subscript $\mathcal{K}$ of $\cong_{\mathcal{K}}$ and we extend the definition to forest quasi-interpretations in the obvious way. Forest quasi-models represent, intuitively, an intermediate step between arbitrary models of $\mathcal{K}$ and forest models of $\mathcal{K}$.

Since KBs are assumed to be simplified, it can mostly be checked locally (by looking at an element of the domain and its direct neighbors) whether an interpretation $\mathcal{I}$ is a model of $\mathcal{K}$. Only nominals impose a global restriction on the cardinality of concepts. We call an element $\delta \in \Delta^{\mathcal{I}}$ locally $\mathcal{K}$-consistent if it satisfies each GCI in $\mathcal{K}$ (a functionality restriction $\mathsf{func}(f)$ is satisfied if δ has at most one f-neighbor); $\mathcal{I}$ is a model of $\mathcal{K}$ if each $\delta \in \Delta^{\mathcal{I}}$ is locally $\mathcal{K}$-consistent and, for each $o \in \mathsf{nom}(\mathcal{K})$, there is exactly one element $\delta \in \Delta^{\mathcal{I}}$ such that $o^{\mathcal{I}} = \delta$. We now show how we can obtain a forest quasi-model from a model of $\mathcal{K}$ by using an adapted version of unraveling.

Definition 2. *Let $\mathcal{I}$ be a model for $\mathcal{K}$ and choose a function that returns, for a concept $C = \exists U.B \in \mathsf{cl}(\mathcal{K})$ and $\delta \in C^{\mathcal{I}}$ some $\delta_{C,\delta} \in \Delta^{\mathcal{I}}$ s.t. $\langle \delta, \delta_{C,\delta} \rangle \in U^{\mathcal{I}}$ and $\delta_{C,\delta} \in B^{\mathcal{I}}$. W.l.o.g., we assume that if $\delta \in C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$ with $C_i = \exists U_i.B_i \in \mathsf{cl}(\mathcal{K}), \mathsf{choose}(C_i, \delta) = \delta_i$ for $i \in \{1, 2\}$ and $\langle \delta, \delta_1 \rangle \cong \langle \delta, \delta_2 \rangle$, then $\delta_1 = \delta_2$.*

An unraveling *for some $\delta \in \Delta^{\mathcal{I}}$, denoted $\downarrow(\mathcal{I}, \delta)$, is an interpretation obtained from $\mathcal{I}$ and δ as follows: we define the set $S \subseteq (\Delta^{\mathcal{I}})^*$ of sequences to be the smallest set such that δ is a sequence and $\delta_1 \cdots \delta_n \cdot \delta_{n+1}$ is a sequence, if*

- *$\delta_1 \cdots \delta_n$ is a sequence,*
- *if $n > 2$ and $\langle \delta_n, \delta_{n-1} \rangle \in f^{\mathcal{I}}$ for some functional role f, then $\delta_{n+1} \neq \delta_{n-1}$,*
- *$\delta_{n+1} = \mathsf{choose}(C, \delta_n)$ for some $C = \exists U.B \in \mathsf{cl}(\mathcal{K})$.*

Now fix a set $F \subseteq \{\delta\} \times \mathbb{N}^*$ and a bijection $\lambda \colon F \to S$ such that (i) F is a forest, (ii) $\lambda(\delta, \varepsilon) = \delta$, (iii) if $(\delta, w), (\delta, w \cdot c) \in F$ with $w \cdot c$ a successor of w, then $\lambda(\delta, w \cdot c) = \lambda(\delta, w) \cdot \delta_{n+1}$ for some $\delta_{n+1} \in \Delta^{\mathcal{I}}$. For each $(\delta, w) \in F$, set $\mathsf{Tail}(\delta, w) = \delta_n$ if $f(\delta, w) = \delta_1 \cdots \delta_n$. The unraveling for δ is the interpretation $\mathcal{J}$ with $\Delta^{\mathcal{J}} = F$ and, for each $(\delta, w) \in \Delta^{\mathcal{J}}$:

(a) for each $o \in \mathsf{nom}(\mathcal{K}), N_o^{\mathcal{J}} = \{(\delta, w) \in \Delta^{\mathcal{J}} \mid \mathsf{Tail}(\delta, w) \in o^{\mathcal{I}}\}$ for $N_o \in N_C$ a fresh concept name;
(b) for each concept name $A \in \mathsf{con}(\mathcal{K})$, $(\delta, w) \in A^{\mathcal{J}}$ iff $\mathsf{Tail}(\delta, w) \in A^{\mathcal{I}}$;
(c) for each role name $r \in \mathsf{rol}(\mathcal{K})$, $\langle (\delta, w), (\delta, w') \rangle \in r^{\mathcal{J}}$ iff (δ, w') is a neighbor of (δ, w), and $\langle \mathsf{Tail}(\delta, w), \mathsf{Tail}(\delta, w') \rangle \in r^{\mathcal{I}}$.

Let $R \subseteq \Delta^{\mathcal{I}}$ contain each δ s.t. $o^{\mathcal{I}} = \delta$ for some $o \in \mathsf{nom}(\mathcal{K})$. The union of all $\downarrow(\mathcal{I}, \delta)$ with $\delta \in R$ is called an unraveling for $\mathcal{I}$, denoted $\downarrow(\mathcal{I})$, where unions of interpretations are defined in the natural way.

Please note that the function Tail can also be seen as a homomorphism (up to signature extension) from the elements in the unraveling to elements in the original model. Fig. 1b shows the unraveling for our example KB and model. The dotted lines for the non-root elements labeled N_o indicate that a copy of the whole tree should be appended.

Unravelings are the first step in the process of transforming an arbitrary model of $\mathcal{K}$ into a forest model since the resulting interpretation is a strict forest quasi-model of $\mathcal{K}$.

Lemma 1. *Let $\mathcal{I}$ be a model of $\mathcal{K}$, then $\mathcal{J} = \downarrow(\mathcal{I})$ is a strict forest quasi-model for $\mathcal{K}$ with branching degree bounded in $|\mathsf{cl}(\mathcal{K})|$.*

In the following steps, we traverse a forest quasi-model in an order in which elements with smaller tree depth are always of smaller order than elements with greater tree depth. Elements with the same tree depth are ordered lexicographically. The bounded branching degree of unravelings then guarantees that, after a finite number of steps, we go on to the next level in the forest and process all nodes eventually. Further, we can merge nodes such that, finally, all nominal placeholders (in the extension of some N_o) can be interpreted as nominals without violating functionality restrictions. In fact, we do not only have to merge nominal placeholders, but also elements that are related to a nominal placeholder by an inverse functional role since, by definition of the semantics, these elements have to correspond to the same element in a model. In order to identify such elements, we define *backwards counting paths* as follows:

Definition 3. *Let $\mathcal{I}$ be a (quasi) forest model for $\mathcal{K}$. We call $p = \delta_1 \cdot \ldots \cdot \delta_n$ a path from δ_1 to δ_n if, for each i with $1 \leq i < n$, $\langle \delta_i, \delta_{i+1} \rangle \in r_i^{\mathcal{I}}$ for some role $r_i \in \mathsf{rol}(\mathcal{K})$. The length $|p|$ of a path p is $n - 1$. We write $\delta_1 \xrightarrow{U_1} \delta_2 \ldots \xrightarrow{U_{n-1}} \delta_n$ to denote that $\langle \delta_i, \delta_{i+1} \rangle \in U_i^{\mathcal{I}}$ for each $1 \leq i < n$. The path p is a descending path if there is some $(\rho, \varepsilon) \in \Delta^{\mathcal{I}}$ s.t., for each $1 \leq i \leq n, \delta_i = (\rho, w_i)$ and, for each $1 \leq i < n, |w_i| < |w_{i+1}|$; p is a backwards counting path (BCP) in $\mathcal{I}$ if $\delta_n \in o^{\mathcal{I}}$ ($\delta_n \in N_o^{\mathcal{I}}$) for some $o \in \mathsf{nom}(\mathcal{K})$ and, for each $1 \leq i < n, \langle \delta_i, \delta_{i+1} \rangle \in f_i^{\mathcal{I}}$ for*

some inverse functional role f_i; p is a descending BCP *if it is descending and a BCP. Given a BCP $p = \delta_1 \xrightarrow{f_1} \delta_2 \ldots \xrightarrow{f_n} \delta_{n+1}$ with $\delta_{n+1} \in o^{\mathcal{J}}$ ($\delta_{n+1} \in N_o^{\mathcal{J}}$), we call the sequence $f_1 \cdots f_n o$ a* path sketch *of p.*

Let $\leq$ be a total order over N_I, $\mathcal{K}$ a consistent $\mathcal{ALCOIFb}$ KB, and $\mathcal{J}$ a forest quasi-interpretation for $\mathcal{K}$. We extend the order to elements in $\Delta^{\mathcal{J}}$ as follows: let $w_1 = w_p \cdot c_1^1 \cdots c_1^n, w_2 = w_p \cdot c_2^1 \cdots c_2^m \in \mathbf{N}^*$ where $w_p \in \mathbf{N}^*$ is the longest common prefix of w_1 and w_2, then $w_1 < w_2$ if either $n < m$ or both $n = m$ and $c_1^1 < c_2^1$. For $i \in \{1, 2\}$ and $(\rho_i, \varepsilon) \in \Delta^{\mathcal{J}}$, let $o_i \in \mathsf{nom}(\mathcal{K})$ be the smallest nominal such that $(\rho_i, \varepsilon) \in N_{o_i}^{\mathcal{J}}$. Now $(\rho_1, w_1) < (\rho_2, w_2)$ if either (i) $|w_1| < |w_2|$ or (ii) $|w_1| = |w_2|$ and $o_1 < o_2$ or (ii) $|w_1| = |w_2|, o_1 = o_2$ and $w_1 < w_2$. When collapsing, we create new elements of the form $(\rho w, w')$ from (ρ, ww'). We extend, therefore, the order as follows: $(\rho_1 w_1, w_1') < (\rho_2 w_2, w_2')$ if $(\rho_1, w_1 w_1') < (\rho_2, w_2 w_2')$.

Please note that (ρ, w) is already a descending BCP if $(\rho, w) \in o^{\mathcal{I}}$ ($N_o^{\mathcal{I}}$). We now show how we can "collapse" a forest quasi-model into a forest model provided it satisfies some admissibility restrictions. During the traversal, we distinguish two situations: (i) we encounter an element (ρ, w) that starts a descending BCP and we have not seen another element before that starts a descending BCP with the same path sketch. In this case, we promote (ρ, w) to become a new root node of the form $(\rho w, \varepsilon)$ and we shift the subtree rooted in (ρ, w) with it; (ii) we encounter a node (ρ, w) that starts a descending BCP, but we have already seen a node (ρ', w') that starts a descending BCP with that path sketch and which is now a root of the form $(\rho' w', \varepsilon)$. In this case, we delete the subtree rooted in (ρ, w) and identify (ρ, w) with $(\rho' w', \varepsilon)$. If (ρ, w) is an f-successor of its predecessor for some inverse functional role f, we delete all f^--successors of $(\rho' w', \varepsilon)$ and their subtrees in order to satisfy the functionality restriction. We use a notion of collapsing admissibility to characterize models s.t. the the predecessor of (ρ, w) satisfies the same atomic concepts as the deleted successor of (ρ', w'), which ensures that local consistency is preserved.

Definition 4. *Let $\mathcal{K}' = \mathsf{nomFree}(\mathcal{K})$ and $\mathcal{J}$ a forest quasi-interpretation for $\mathcal{K}$. We define $\sim$ as the smallest equivalence relation on $\Delta^{\mathcal{J}}$ that satisfies $\delta_1 \sim \delta_2$ if δ_1, δ_2 start descending BCPs with identical path sketches.*

If $\mathcal{J}$ is a strict forest quasi-model for $\mathcal{K}$, we call $\mathcal{J}_0 = \mathcal{J}$ an initial collapsing *for $\mathcal{J}$ and the smallest element $(\rho_0, w_0) \in \Delta^{\mathcal{J}_0}$ with $w_0 \neq \varepsilon$ that starts a descending BCP the* focus *of $\mathcal{J}_0$. Let $\mathcal{J}_i$ be a collapsing for $\mathcal{J}$ and $(\rho_i, w_i) \in \Delta^{\mathcal{J}_i}$ the focus of $\mathcal{J}_i$. We obtain a collapsing $\mathcal{J}_{i+1}$ for $\mathcal{J}$ from $\mathcal{J}_i$ with focus (ρ_{i+1}, w_{i+1}) the smallest element starting a descending BCP and $(\rho_{i+1}, w_{i+1}) > (\rho_i, w_i)$ according to the following two cases:*

1. *There is no element $(\rho, \varepsilon) \in \Delta^{\mathcal{J}_i}$ s.t. $(\rho, \varepsilon) < (\rho_i, w_i)$ and $(\rho, \varepsilon) \sim (\rho_i, w_i)$. Then $\mathcal{J}_{i+i}$ is obtained from $\mathcal{J}_i$ by renaming each element $(\rho_i, w_i w_i') \in \Delta^{\mathcal{J}_i}$ to $(\rho_i w_i, w_i')$.*
2. *There is an element $(\rho, \varepsilon) \in \Delta^{\mathcal{J}_i}$ s.t. $(\rho, \varepsilon) < (\rho_i, w_i)$ and $(\rho, \varepsilon) \sim (\rho_i, w_i)$. Let (ρ, ε) be the smallest such element.*
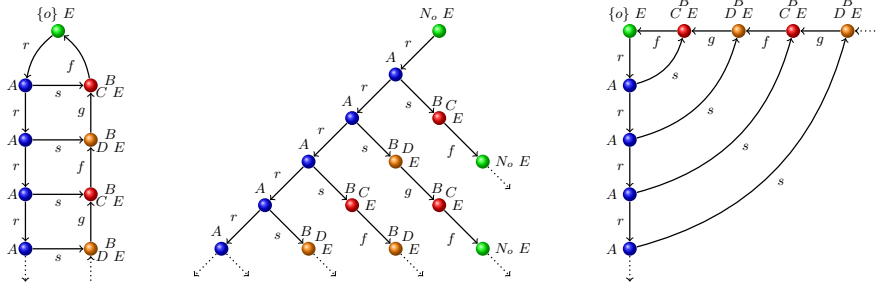
Fig. 1. a) A representation of a model for the running example. b) Result of unraveling this model. c) Result of collapsing this unraveling with infinitely many new root elements displayed in the top line.

(a) $\Delta^{\mathcal{J}_{i+1}} = \Delta^{\mathcal{J}_i} \setminus (\{(\rho_i, w_i w_i') \mid w_i' \in \mathbb{N}^*\} \cup \{(\rho, w) \mid w = c \cdot w', c \in \mathbb{N}, w' \in \mathbb{N}^*, (\rho_i, w_i) \text{ has a predecessor } (\rho_i, w_i') \text{ such that } \langle(\rho_i, w_i'), (\rho_i, w_i)\rangle \in f^{\mathcal{J}_i}$ for an inverse functional role f in $\text{rol}(\mathcal{K})$ and $\langle(\rho, c), (\rho, \varepsilon)\rangle \in f^{\mathcal{J}_i}\})$;

(b) for each $A \in \text{con}(\mathcal{K})$ and $\delta \in \Delta^{\mathcal{J}_{i+1}}$, $\delta \in A^{\mathcal{J}_{i+1}}$ iff $\delta \in A^{\mathcal{J}_i}$;

(c) for each $r \in \text{rol}(\mathcal{K})$ and $\delta_1, \delta_2 \in \Delta^{\mathcal{J}_{i+1}}$, $\langle\delta_1, \delta_2\rangle \in r^{\mathcal{J}_{i+1}}$ iff (a) $\langle\delta_1, \delta_2\rangle \in r^{\mathcal{J}_i}$ or (b) δ_1 is the predecessor of (ρ_i, w_i) in $\mathcal{J}_i$, $\delta_2 = (\rho, \varepsilon)$, and $\langle\delta_1, (\rho_i, w_i)\rangle \in r^{\mathcal{J}_i}$.

For a collapsing $\mathcal{J}_i$, $\text{safe}(\mathcal{J}_i)$ is the restriction of $\mathcal{J}_i$ to elements (ρ, w) s.t. $(\rho, w) \in \mathcal{J}_j$ for all $j \geq i$. With $\mathcal{J}_\omega$ we denote the non-disjoint union of all interpretations $\text{safe}(\mathcal{J}_i)$ obtained from subsequent collapsings $\mathcal{J}_i$ for $\mathcal{J}$. The interpretation obtained from $\mathcal{J}_\omega$ by interpreting each $o \in \text{nom}(\mathcal{K})$ as $(\rho, \varepsilon) \in N_o^{\mathcal{J}_\omega}$ is denoted by $\text{collapse}(\mathcal{J})$ and called a purified interpretation w.r.t. $\mathcal{J}$. If $\text{collapse}(\mathcal{J}) \models \mathcal{K}$, we call $\text{collapse}(\mathcal{J})$ a purified model of $\mathcal{K}$.

Definition 5. *Let $\mathcal{J}$ be a forest quasi-interpretation for $\mathcal{K}$, then $\mathcal{J}$ is* collapsing-admissible *if there exists a function* $\text{ch} \colon (\text{cl}(\mathcal{K}) \times \Delta^{\mathcal{J}}) \to \Delta^{\mathcal{J}}$ *s.t., for each $C = \exists U.B \in \text{cl}(\mathcal{K})$ and $\delta \in C^{\mathcal{J}}$,*

1. $\langle\delta, \text{ch}(C, \delta)\rangle \in U^{\mathcal{J}}, \text{ch}(C, \delta) \in B^{\mathcal{J}}$ *and, if there is no functional role f s.t.* $\langle\delta, \text{ch}(C, \delta)\rangle \in f^{\mathcal{J}}$, *then $\text{ch}(C, \delta)$ is a successor of δ,*
2. *if there is some $\delta' \in C^{\mathcal{J}}$ s.t. δ and δ' start descending BCPs with identical path sketches, then $\langle\delta, \text{ch}(C, \delta)\rangle \cong \langle\delta', \text{ch}(C, \delta')\rangle$.*

Since, in unravelings, elements that start (desccending) BCPs with identical path sketches have been generated from the same element in the unravelled model, collapsing-admissibility is immediate and the function ch can be defined using the function choose from the unraveling. We can use collapsing-admissibility to show that, whenever we delete an element and the subtree rooted in it during the collapsing, the predecessor of the focus is a suitable replacement.

Lemma 2. *Let $\mathcal{J}$ be a strict forest quasi-model for $\mathcal{K}$ with branching degree b that is collapsing-admissible. Then $\text{collapse}(\mathcal{J})$ is a forest model for $\mathcal{K}$ that still has branching degree b.*

Since unravelings are strict forest quasi-models with bounded branching degree by Lemma 1 and collapsing-admissible, it is an immediate consequence of Lemma 2 that collapsing an unraveling yields a forest model with bounded branching degree. At this point, the number of roots might still be infinite and we could have obtained the same result by unraveling an arbitrary model, where we take all elements on BCPs as roots instead of taking just the nominals and creating new roots in the collapsing process. In the next sections, however, we show how we can transform an unraveling of a counter-model for the query such that it remains collapsing-admissible and such that it can in the end be collapsed into a forest model with a finite number of roots that is still a counter model for the query. For this transformation it is much more convenient to work with real (strict) trees and forests, which is why use (strict) forest quasi-interpretations.

3 Quasi-Entailment in Quasi-Models

In this section, we provide a characterization for query entailment in forest quasi-models that mirrors query entailment for the corresponding "proper models". In our further argumentation, we talk about the initial part of a tree, i.e., the part that remains if one cuts branches down to a fixed length. For a forest interpretation $\mathcal{I}$ and some $n \in \mathbb{N}$, we denote, therefore, with $\mathsf{cut}_n(\mathcal{I})$ the interpretation obtained from $\mathcal{I}$ by restricting $\Delta^{\mathcal{I}}$ to those pairs (ρ, w) for which $|w| \leq n$. One can show that, in the case of purified models, we find only finitely many unraveling trees of depth n that "look different" (i.e., that are non-isomorphic).

For our further considerations, we introduce the notion of "anchored n-components". These are certain substructures of forest quasi-interpretations that we use to define a notion of quasi entailment.

Definition 6. *Let $\mathcal{J}$ be a forest quasi-interpretation and $\delta \in \Delta^{\mathcal{J}}$. An interpretation $\mathcal{C}$ is called* anchored n-component *of $\mathcal{J}$ with witness δ if $\mathcal{C}$ can be created by restricting $\mathcal{J}$ to a set $W \subseteq \Delta^{\mathcal{J}}$ obtained as follows:*

- *Let $\mathcal{J}_\delta$ be the subtree of $\mathcal{J}$ that is started by δ and let $\mathcal{J}_{\delta,n} := \mathsf{cut}_n(\mathcal{J}_\delta)$. Select a subset $W' \subseteq \Delta^{\mathcal{J}_{\delta,n}}$ that is closed under predecessors.*
- *For every $\delta' \in W'$, let P be a finite set (possibly empty) of descending BCPs $\mathbf{p}$ starting from δ' and let $W_{\delta'}$ contain all nodes from all $\mathbf{p} \in P$.*
- *Set $W = W' \cup \bigcup_{\delta' \in W'} W_{\delta'}$.*

The following definition and lemma employ the notion of anchored n-components to come up with the notion of *quentailment* (short for quasi-entailment), a criterion that reflects query-entailment in the world of forest quasi-models. Fig. 2 illustrates this correspondence.

Definition 7. *Let $\mathcal{J}$ be a forest quasi-model for $\mathcal{K}$ and q a CQ with $\sharp(q) = n$ and $V = \mathsf{var}(q)$. We say that $\mathcal{J}$* quentails *q, written $\mathcal{J} \mathrel{\approx\!\!\!\mid} q$, if $\mathcal{J}$ contains connected anchored n-components $\mathcal{C}_1, \ldots, \mathcal{C}_\ell$ and there are variable assignment functions $\mu_i \colon V \to 2^{\Delta^{\mathcal{C}_i}}$ such that:*
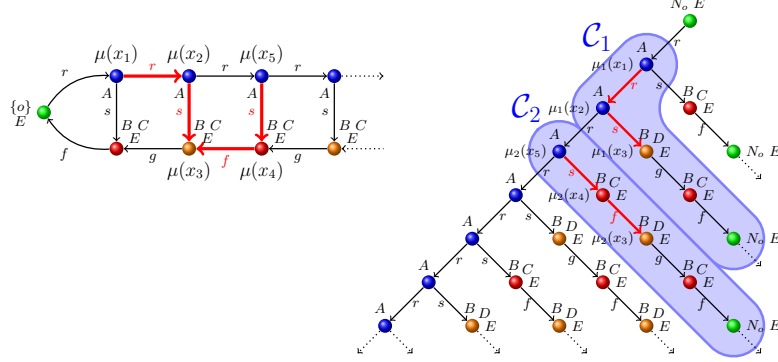
Fig. 2. Correspondence between entailment and quentailment. Left: match μ for $q = \{r(x_1, x_2), s(x_2, x_3), f(x_4, x_3), s(x_5, x_4)\}$ into the "proper model". Right: anchored components $\mathcal{C}_1$ and $\mathcal{C}_2$ and functions μ_1 and μ_2 establishing $\downarrow(\mathcal{I}) \approx\!\!\!| \, q$.

Q1 For every $x \in V$, there is at least one $\mathcal{C}_i$, such that $\mu_i(x) \neq \emptyset$

Q2 For all $A(x) \in q$, we have $\mu_i(x) \subseteq A^{\mathcal{J}}$ for some i.

Q3 For every $r(x, y) \in q$ there is a $\mathcal{C}_i$ such that there are $\delta_1 \in \mu_i(x)$ and $\delta_2 \in \mu_i(y)$ such that $\langle \delta_1, \delta_2 \rangle \in r^{\mathcal{J}}$.

Q4 If, for some $x \in V$, there are connected anchored n-components $\mathcal{C}_i$ and $\mathcal{C}_j$ with $\delta \in \mu_i(x)$ and $\delta' \in \mu_j(x)$, then there is
  - a sequence $\mathcal{C}_{n_1}, \ldots, \mathcal{C}_{n_k}$ with $n_1 = i$ and $n_k = j$ and
  - a sequence $\delta_1, \ldots, \delta_k$ with $\delta_1 = \delta$ and $\delta_k = \delta'$ as well as $\delta_m \in \mu_{n_m}(x)$ for all $1 \leq m < k$,

  such that, for every m with $1 \leq m < k$, we have that
  - $\mathcal{C}_{n_m}$ contains a descending BCP p_1 started by δ_m,
  - $\mathcal{C}_{n_{m+1}}$ contains a descending BCP p_2 started by δ_{m+1},
 - p_1 and p_2 have the same path sketch.

Note that an anchored component may contain none, one or several instantiations of a variable $x \in V$. Intuitively, the definition ensures, that we find matches of query parts which when fitted together by identifying BCP-equal elements yield a complete query match.

Lemma 3. *For any model $\mathcal{I}$ of $\mathcal{K}$, $\downarrow(\mathcal{I}) \approx\!\!\!| \, q$ implies $\mathcal{I} \models q$ and, for any collapsing-admissible strict forest quasi-model $\mathcal{J}$ of $\mathcal{K}$, $\mathsf{collapse}(\mathcal{J}) \models q$ implies $\mathcal{J} \approx\!\!\!| \, q$.*

4 Limits and Forest Transformations

One of the major obstacles for a decision procedure for CQ entailment is that for DLs including inverses, nominals, and cardinality restrictions (or alternatively functionality), there are potentially infinitely many new roots. If we want to eliminate new roots such that only finitely many remain, they have to be replaced by "uncritical" elements. We will construct such elements as "environment-limits"
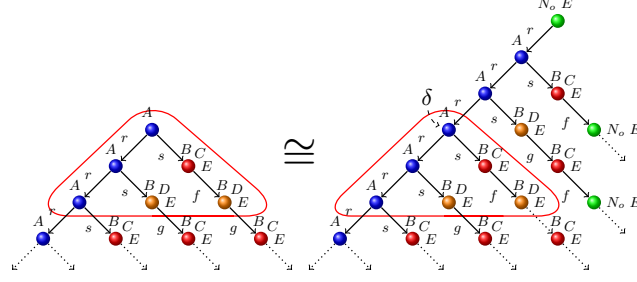
Fig. 3. Pseudo forest model (right) and according 2-secure replacement for δ (left).

– new domain elements which can be approximated with arbitrary precision by already present domain elements – possibly without themselves being present in the domain.[3]

Definition 8. *Let $\mathcal{I}$ with $\delta \in \Delta^{\mathcal{I}}$ be a model of $\mathcal{K}$. A tree interpretation $\mathcal{J}$ is said to be* generated by δ, *written: $\mathcal{J} \lhd \delta$, if it is isomorphic to the restriction of $\downarrow(\mathcal{I}, \delta)$ to elements of $\{(\delta, cw) \mid (\delta, cw) \in \Delta^{\downarrow(\mathcal{I}, \delta)}, c \notin H\}$ for some $H \subseteq \mathbb{N}$. The set of* limits *of $\mathcal{I}$, written $\lim \mathcal{I}$, is the set of all tree interpretations $\mathcal{J}$ s.t., for every $k \in \mathbb{N}$, there are infinitely many $\delta \in \Delta^{\mathcal{I}}$ with $\mathsf{cut}_k(\mathcal{L}) \cong \mathsf{cut}_k(\mathcal{J})$ for some $\mathcal{L} \lhd \delta$.*

The right hand side of Fig. 3 displays one limit element of our example model. The following lemma gives some useful properties of limits.

Lemma 4. *Let $\mathcal{K}' = \mathsf{nomFree}(\mathcal{K})$, $\mathcal{I}$ a purified model of $\mathcal{K}$, and n some fixed natural number. Then the following hold:*

1. *Let $\mathcal{L}'$ be a tree interpretation such that there are infinitely many $\delta \in \Delta^{\mathcal{I}}$ with $\mathcal{L}' \cong \mathsf{cut}_n(\mathcal{L})$ for some $\mathcal{L} \lhd \delta$. Then, there is at least one limit $\mathcal{J} \in \lim \mathcal{I}$ such that $\mathsf{cut}_n(\mathcal{J}) \cong \mathcal{L}'$.*
2. *Every $\mathcal{J} \in \lim \mathcal{I}$ is locally $\mathcal{K}'$-consistent apart from its root (ρ, ε).*
3. *For every $\mathcal{J} \in \lim \mathcal{I}$, every root (ρ, ε) in $\mathcal{J}$ has no BCP to any $(\rho, w) \in \Delta^{\mathcal{J}}$.*
4. *Every $\mathcal{J} \in \lim \mathcal{I}$ is collapsing-admissible.*

Having defined and justified limit elements as convenient building blocks for restructuring forest quasi-interpretations, the following definition states how this restructuring is carried out.

Definition 9. *Let $\mathcal{I}$ be a model for $\mathcal{K}$ and $\mathcal{J}$ some forest quasi-model for $\mathcal{K}$ with $\delta \in \Delta^{\mathcal{J}}$. A strict tree quasi-interpretation $\mathcal{J}' \in \lim \mathcal{I}$ is called an n-secure replacement for δ if (i) $\mathsf{cut}_n(\downarrow(\mathcal{J}, \delta))$ is isomorphic to $\mathsf{cut}_n(\mathcal{J}')$ and (ii) for every anchored n-component of $\mathcal{J}'$ with witness δ', there is an isomorphic anchored n-component of $\mathcal{J}$ with witness δ. If $\delta \in \Delta^{\mathcal{J}}$ has an n-secure replacement in $\lim \mathcal{I}$, δ is n-replaceable w.r.t. $\mathcal{I}$ and it is n-irreplaceable w.r.t. $\mathcal{I}$ otherwise.*

---

[3] As an analogy, consider the fact that any real number can be approximated by a sequence of rational numbers, even if it is itself irrational.
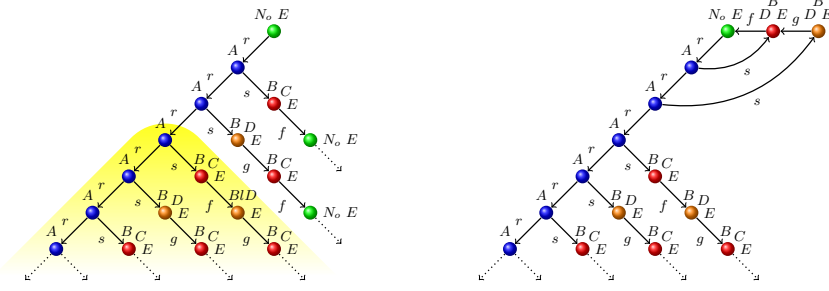
Fig. 4. a) Result of replacing the element δ by the 2-secure replacement depicted in Fig. 3. The inserted component is highlighted. b) Result of collapsing the forest quasi-model displayed on the left hand side.

Now, let $(\rho, w) \in \Delta^{\mathcal{J}}$ be n-replaceable w.r.t. $\mathcal{I}$ and $\mathcal{J}'$ an according n-replacement for (ρ, w) from $\lim \mathcal{I}$ with root (ς, ε). The result of replacing (ρ, w) by $\mathcal{J}'$ is an interpretation $\mathcal{R}$ with $\Delta^{\mathcal{R}} = \Delta^{\mathcal{J}}_{\mathsf{red}} \cup \{(\rho, ww'') \mid (\varsigma, w'') \in \Delta^{\mathcal{J}'}\}$ for $\Delta^{\mathcal{J}}_{\mathsf{red}} = (\Delta^{\mathcal{J}} \setminus \{(\rho, ww') \mid |w'| > 1\})$ s.t.

- for each $A \in \mathsf{con}(\mathcal{K}')$, $A^{\mathcal{R}} = (A^{\mathcal{J}} \cap \Delta^{\mathcal{J}}_{\mathsf{red}}) \cup \{(\rho, ww') \mid (\varsigma, w') \in A^{\mathcal{J}'}\}$
- for each $r \in \mathsf{rol}(\mathcal{K}')$, $r^{\mathcal{R}} = (r^{\mathcal{J}} \cap \Delta^{\mathcal{J}}_{\mathsf{red}} \times \Delta^{\mathcal{J}}_{\mathsf{red}}) \cup \{\langle(\rho, ww'), (\rho, ww'')\rangle \mid \langle(\varsigma, w'), (\varsigma, w'')\rangle \in r^{\mathcal{J}'}\}$

For $\mathcal{J} = \downarrow(\mathcal{I})$, an interpretation $\mathcal{J}'$ is called an n-secure transformation of $\mathcal{J}$ if it is obtained by (possibly infinitely) repeating the following step:

Choose one unvisited w.r.t. tree-depth minimal node (ρ, w) that is n-replaceable w.r.t. $\mathcal{I}$. Replace (ρ, w) with one of its n-secure replacements from $\lim \mathcal{I}$ and mark (ρ, w) as visited.

Fig. 3 displays a 32-secure replacement in the considered unraveling of our example model and Fig. 4a displays the result of carrying out this replacement step on our example. The following lemma ensures that not too many elements (actually defined in terms of the original model) are exempt from being replaced.

Lemma 5. *Every purified model $\mathcal{I}$ of $\mathcal{K}$ contains only finitely many distinct elements that start a BCP and are the cause for n-irreplaceable nodes in the unraveling of $\mathcal{I}$.*

Next, we can show that the process of unraveling, n-secure transformation and collapsing preserves the property of being a model of a knowledge base and (with the right choice of n) also preserves the property of *not* entailing a conjunctive query. Moreover, this model conversion process ensures that the resulting model contains only finitely many new nominals (witnessed by a bound on the length of BCPs). Fig. 4b illustrates these properties for our example model. Note that only two new nominals are left whereas collapsing the original unraveling yields infinitely many.

Lemma 6. *Let $\mathcal{I}$ be a purified model of $\mathcal{K}$, $\mathcal{J} = \downarrow(\mathcal{I})$, and $\mathcal{J}'$ an n-secure transformation of $\mathcal{J}$. Then the following hold:*

1. collapse($\mathcal{J}'$) is a model of $\mathcal{K}$.
2. There is a natural number m such that $\mathcal{J}'$ does not contain any node whose shortest descending BCP has a length greater than m.
3. If, for some CQ q, we have $\mathcal{J} \not\models q$ and $n > \sharp(q)$, then $\mathcal{J}' \not\models q$.
4. If, for some CQ q, we have $\mathcal{I} \not\models q$ and $n > \sharp(q)$, then collapse($\mathcal{J}'$) $\not\models q$.

Now we are able to establish our first milestone on the way to showing finite representability of countermodels.

Theorem 1. *For every $\mathcal{ALCOIF}b$ KB $\mathcal{K}$ and CQ q s.t. $\mathcal{K} \not\models q$, there is a forest model $\mathcal{I}$ of $\mathcal{K}$ with finitely many roots and bounded branching degree s.t. $\mathcal{I} \not\models q$.*

5 Finite Representations of Models

We can now use standard techniques from tableau algorithms (adapted to work on models) to construct finite representations for a forest model of $\mathcal{K}$ with finitly many roots. In particular the tableau algorithm with n-tree-blocking, $n \geq \sharp(q)$, for deciding CQ entailment in $\mathcal{SHIQ}$, $\mathcal{SHOQ}$, and $\mathcal{SHOI}$ with only simple roles in the query [5, 6] works exactly like that. We call an interpretation on which we applied n-tree-blocking and discarded all blocked elements an n-representation. Such an n-representation corresponds to a complete and clash-free completion graph in tableau algorithms. Since we fixed a bound on the number of roots in Theorem 1 and otherwise only consider the closure $cl(\mathcal{K})$ of $\mathcal{K}$, one can show that there are only finitely many non-isomorphic n-blocking-trees even though we take links back to roots into account. Similarly, we can now use an adapted version of the technique for building a tableau from a complete and clash-free completion graph to show that we can obtain a model for $\mathcal{K}$ from an n-representation.

Lemma 7. *Let $n \geq \sharp(q)$. If $\mathcal{K} \not\models q$, then there is an n-representation $\mathcal{R}$ of $\mathcal{K}$ s.t. $\mathcal{R} \not\models q$, and from $\mathcal{R}$ one can build a model $\mathcal{I}$ of $\mathcal{K}$ such that $\mathcal{I} \not\models q$.*

Thus, we can enumerate all (finite) n-representations for $\mathcal{K}$ and check whether they entail q. Together with the semi-decidability result for FOL [4], we get the desired theorem.

Theorem 2. *It is decidable whether $\mathcal{K} \models q$ for $\mathcal{K}$ an $\mathcal{ALCOIF}b$ knowledge base and q a Boolean conjunctive query.*

This solves the long-standing open problem of deciding conjunctive query entailment in the presence of nominals, inverse roles, and qualified number restrictions. Since the approach is purely a decision procedure, the computational complexity of the problem remains open and will be part of our future work. Similarly, we will embark on extending our results to $\mathcal{SHOIQ}$ with non-simple roles as query predicates.

References

1. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F.: The Description Logic Handbook. Cambridge University Press (2003)
2. Glimm, B., Rudolph, S.: Nominals, inverses, counting, and conjunctive queries or Why infinity is your friend! Technical report, University of Oxford (2009) `http://www.comlab.ox.ac.uk/files/2175/paper.pdf`.
3. Rudolph, S., Krötzsch, M., Hitzler, P.: Terminological reasoning in $\mathcal{SHIQ}$ with ordered binary decision diagrams. In: Proc. 23rd National Conference on Artificial Intelligence (AAAI 2008), AAAI Press/The MIT Press (2008) 529–534
4. Gödel, K.: Über die Vollständigkeit des Logikkalküls. PhD thesis, Universität Wien (1929)
5. Ortiz, M.: Extending CARIN to the description logics of the $\mathcal{SH}$ family. In: Proceedings of Logics in Artificial Intelligence, European Workshop (JELIA 2008), Lecture Notes in Artificial Intelligence (2008) 324–337
6. Ortiz, M., Calvanese, D., Eiter, T.: Data complexity of query answering in expressive description logics via tableaux. Journal of Automated Reasoning **41**(1) (2008) 61–98