Computing Concept Referring Expressions for Queries on Horn $\mathcal{ALC}$ Ontologies - Appendix -

Moritz Illich , Birte Glimm
Institute of Artificial Intelligence, Ulm University, Germany
moritz.illich@uni-ulm.de, birte.glimm@uni-ulm.de

A Algorithm Proofs

Lemma 1. *Given a Horn $\mathcal{ALC}$ ontology $\mathcal{K}$, a generalized instance query Q, a CRE E, and its associated (Horn $\mathcal{ALC}$) current concept C, then $\mathcal{K} \models E \sqsubseteq Q \Leftrightarrow \mathcal{K} \models C \sqsubseteq Q$.*

Proof. ($\Leftarrow$) Per definition, a current concept C of some CRE E always satisfies $\mathcal{K} \models E \sqsubseteq C$, since we either have $\mathcal{K} \models C(a)$ for $E = \{a\}$ or C being a conjunct in the more general $E = C \sqcap \exists r^-.(E')$.
($\Rightarrow$) $\mathcal{K} \models C \sqsubseteq Q$ requires that the current concept C is at least as specific as Q. Therefore, we show that C is actually the most specific (Horn $\mathcal{ALC}$) concept with $\mathcal{K} \models E \sqsubseteq C$ and that for any other most specific candidate D with $\mathcal{K} \models E \sqsubseteq D$, we get $\mathcal{K} \models C \equiv D$. For that, we have to distinguish between the initial and the general form of the CRE E:
1) $E = \{a\}$ – Per definition, the initial current concept C is already the most specific concept with $\mathcal{K} \models \{a\} \sqsubseteq C$. Assuming that there is another most specific concept D, we could just define a new concept $C' = C \sqcap D$, for which $\mathcal{K} \models C' \sqsubseteq C$ and $\mathcal{K} \models C' \sqsubseteq D$. Since C and D are already most specific, the strict relations $\mathcal{K} \models C' \sqsubset C$ and $\mathcal{K} \models C' \sqsubset D$ must not hold. Hence, we obtain $\mathcal{K} \models C' \equiv C$ and $\mathcal{K} \models C' \equiv D$, and thus $\mathcal{K} \models C \equiv D$.
2) $E = C \sqcap \exists r^-.(E')$ with current concept C – Let D be a most specific concept with $\mathcal{K} \models C \sqcap \exists r^-.(E') \sqsubseteq D$. By the construction of E based on E', we then have $\mathcal{K} \models E' \sqsubseteq \exists r.D$, which requires that the existential restriction $\exists r.D$ is either already present in the ontology $\mathcal{K}$ or can be derived by means of other concepts and axioms. In the first case, $\exists r.D$ is also considered for the construction of E. However, since the concept actually occurring in E is C, which has been defined by selecting a reduced existential restriction from the minimal set $\mathcal{K}_{\exists}^{\mathsf{min}}(C)$, and combining it with related universal restrictions, we get $\mathcal{K} \models C \sqsubseteq D$. Since D is most specific (s.t. $\mathcal{K} \not\models C \sqsubset D$) and $\mathcal{K} \models E \sqsubseteq C$, we have $\mathcal{K} \models C \equiv D$.

In order to derive $\exists r.D$, we need another existential restriction $\exists r.A$ for which two cases may apply:
a) $\mathcal{K} \models A \sqsubseteq D$ is given, but since C has been selected for E, we have $\mathcal{K} \models C \sqsubseteq A$ and thus $\mathcal{K} \models C \sqsubseteq D$, which again results in $\mathcal{K} \models C \equiv D$.
b) There are some universal restrictions of the form $\forall r.B_i$, such that $D = A \sqcap B_1 \sqcap B_2 \sqcap \ldots$. Because C is also constructed by combining a reduced existential restriction $\exists r.A'$, that must satisfy $\mathcal{K} \models A' \sqsubseteq A$, with universal ones using the same role r, including the ones applied for D, we get $\mathcal{K} \models C \sqsubseteq D$ and furthermore $\mathcal{K} \models C \equiv D$ as before. □

A.1 Properties

In the following, we discuss why the described algorithm works properly in the way that for a generalized instance query, only certain, singular, and unique answers are produced with reference to Definition 2. Additionally, it is shown that the algorithm returns every wanted answer and, furthermore, always terminates.

Certainty
Condition: $\mathcal{K} \models E \sqsubseteq Q$ and $|E^{\mathcal{I}}| > 0$ for all models $\mathcal{I}$ of $\mathcal{K}$
Proof: In the algorithm, a CRE E is selected as answer if $\mathcal{K} \models C \sqsubseteq Q$ for the current concept C, which implies $\mathcal{K} \models E \sqsubseteq Q$ according to Lemma 1. The second condition $|E^{\mathcal{I}}| > 0$ for every model $\mathcal{I}$ is also given, since E is based on an actual, present individual from the ontology $\mathcal{K}$, which means that the interpretation of the associated initial current concept C cannot be empty, hence $|C^{\mathcal{I}}| > 0$. This is also the case for the existential restrictions $\exists r.D$[1] being applied for the construction of E as they are only selected for the current concept C if $\mathcal{K} \models C \sqsubseteq \exists r.D$, thus requiring $|(\exists r.D)^{\mathcal{I}}| > 0$ for $|C^{\mathcal{I}}| > 0$. Moreover, the next current concept D also satisfies $|D^{\mathcal{I}}| > 0$ for every model $\mathcal{I}$ of $\mathcal{K}$, because the property $|(\exists r.D)^{\mathcal{I}}| > 0$ implies that there must exist an element $a \in (\exists r.D)^{\mathcal{I}}$ that has an r-relation to some element in $D^{\mathcal{I}}$, such that the latter cannot be empty. In this respect, the existence of the base individual assures the existence of the anonymous individual being referred to by the initially selected existential restriction, based on which the next anonymous individual is determined. This procedure can be repeated until the one individual is reached that is realized by the whole expression E, thus resulting in $|E^{\mathcal{I}}| > 0$.

Singularity
Condition: $|E^{\mathcal{U}}| = 1$ for the universal model $\mathcal{U}$ of $\mathcal{K}$
Proof: We can say that a CRE E is *not* singular if the property $|E^{\mathcal{U}}| > 1$ holds for the ontology's tree-like universal model $\mathcal{U}$. Therefore, we prove that this is never the case for any CRE E returned by the algorithm:

[1] $\exists r.D$ may result from the combination of some $\exists r.A$ and $\forall r.B_i$ with $\mathcal{K} \models C \sqsubseteq \exists r.A$ and $\mathcal{K} \models C \sqsubseteq \forall r.B_i$, thus implying $\mathcal{K} \models C \sqsubseteq \exists r.D$ with $D = A \sqcap B_1 \sqcap \ldots$.

For the initial $E = \{a\}$ we directly get $|E^{\mathcal{U}}| = 1$, because the interpretation of an individual uniquely refers to one domain element. In the more general case, a CRE E is defined as $E = D \sqcap \exists r^-.(E')$ where E' denotes the (recursive) part leading to the base individual. Here, $|E^{\mathcal{U}}| > 1$ would mean that there are at least two elements $a, b \in D^{\mathcal{U}}$ with $a \neq b$, each possessing an r-relation starting from either the same element $c \in E'^{\mathcal{U}}$, realized by $r(c, a)$ and $r(c, b)$, or two different elements $c_1, c_2 \in E'^{\mathcal{U}}$, given as $r(c_1, a)$ and $r(c_2, b)$, respectively. Since the second case would lead to a recursion of the same situation considering c_1 and c_2 instead of a and b, there must be at a minimum one point in the CRE where the first case applies which is at least realized when the initial nominal is considered with a concrete individual as single starting point. The necessary conditions that a single individual c really relates to two different elements a, b are described in the following lemma.

Lemma 2. *Let $\mathcal{K}$ be a Horn $\mathcal{ALC}$ ontology, $\mathcal{U}$ the universal model of $\mathcal{K}$, and assume that $\langle c, a\rangle, \langle c, b\rangle \in r^{\mathcal{U}}$ for some role r, and (anonymous) $a, b \in D^{\mathcal{U}}$ for some concept D. We have $a \neq b$ iff there exists a concept C such that $c \in C^{\mathcal{U}}$, $\mathcal{K} \models C \sqsubseteq \exists r.A$, $\mathcal{K} \models C \sqsubseteq \exists r.B$, $\mathcal{K} \models A \sqsubseteq D$, $\mathcal{K} \models B \sqsubseteq D$, and $\mathcal{K} \not\models A \equiv B$.*

Proof. ($\Rightarrow$) Because we are considering the universal model $\mathcal{U}$ where all entailments of $\mathcal{K}$ are directly present, the element c having two r-relations explicitly demands for two existential restrictions $\exists r.A$ and $\exists r.B$ with $\mathcal{K} \models C \sqsubseteq \exists r.A$ and $\mathcal{K} \models C \sqsubseteq \exists r.B$ for some concept C with $c \in C^{\mathcal{U}}$, as well as $a \in A^{\mathcal{U}}$ and $b \in B^{\mathcal{U}}$. Combining the latter with the assumed condition $a, b \in D^{\mathcal{U}}$ furthermore requires $\mathcal{K} \models A \sqsubseteq D$ and $\mathcal{K} \models B \sqsubseteq D$. In addition, A and B must not be equivalent, otherwise $C \sqsubseteq \exists r.A$ and $C \sqsubseteq \exists r.B$ could be modeled by only one r-relation in I.
($\Leftarrow$) Since $c \in C^{\mathcal{U}}$, $\mathcal{K} \models C \sqsubseteq \exists r.A$, $\mathcal{K} \models C \sqsubseteq \exists r.B$, and $\mathcal{K} \not\models A \equiv B$, we have $a \neq b$ in the universal model $\mathcal{U}$. Because of $\mathcal{K} \models A \sqsubseteq D$ and $\mathcal{K} \models B \sqsubseteq D$, we also have $a, b \in D^{\mathcal{U}}$. □

The occurrence of this situation is prevented in our algorithm by choosing some $\exists r.D$ with $\mathcal{K} \models C \sqsubseteq \exists r.D$ for the current concept C only if there does not exist another candidate $\exists r.A$ with $\mathcal{K} \models A \sqsubseteq D$, or for the case that $\mathcal{K} \models A \equiv D$, by using only one of either $\exists r.D$ or $\exists r.A$ as candidate. This is due to the selection based on the minimal set $\mathcal{K}_{\exists}^{\mathsf{min}}(C)$ of reduced existential restrictions defined in Definition 3 and thus guarantees singularity of the constructed CREs. Besides, the applied cycle notation does not compromise singularity as it only serves as an abbreviation to compactly represent a possibly infinite number of CREs that may appear due to the repeated application of certain existential restrictions, where each construction still adheres to the same scheme being applied for non-cyclic CREs.

Uniqueness
Condition: $E^{\mathcal{U}} \neq E'^{\mathcal{U}}$ for the universal model $\mathcal{U}$ of $\mathcal{K}$
Proof: For the tree-like universal model $\mathcal{U}$, having two different CREs E, E' with the unwanted characteristic $E^{\mathcal{U}} = E'^{\mathcal{U}}$ would mean that i) they both describe the same path in the tree using different concepts or ii) one refers to a sub-path of the other leading to the same anonymous individual. Another way where neither E nor E' is a sub-path of the other is not possible, because two different individuals (nodes) a, b referring to the same anonymous element c would require relations (edges) of the form $r_1(a, c_1)$ and $r_2(b, c_2)$ with $c_1 = c_2$. However, equality between individuals cannot be entailed in Horn $\mathcal{ALC}$.
i) Because an edge in a tree path relates to some $\exists r.B$, E' may describe the same path as E if it uses some $\exists r.B'$ with $B \sqsubseteq B'$ (including $B \equiv B'$) for the associated edge instead. In Algorithm 2, this is prevented w.r.t. Line 2 and 11.
ii) Assuming that a and a' are the base individuals of E and E', respectively, and, thus, the starting points of the associated tree paths, there must be some $r(a, a')$ (describing the connection between two explicitly given individuals) such that E' refers to a sub-path of E. If $a'^{\mathcal{U}} \in B^{\mathcal{U}}$ for some concept B, some $\exists r.B$ would describe the same edge as $r(a, a')$, given that $a^{\mathcal{U}} \in (\exists r.B)^{\mathcal{U}}$, which is why this existential restriction must not be chosen[2]. In Algorithm 2, this situation is handled in Line 13.

Completeness
For the completeness property of our algorithm, it is not required to construct every possible CRE that represents a correct, singular answer for the stated query, but rather to find a CRE for every, possibly anonymous, individual that serves as an answer. In this respect, the above uniqueness characteristic does not cause any conflict and we can show that a singular CRE is constructed by the algorithm for every wanted answer of the query. For that, we assume there may exist a separate CRE E which describes a correct, singular and unique answer but that cannot be constructed by the algorithm and prove that this is not possible:

Since the considered CRE E is not generated by the algorithm, it must possess a unique subpart that does not occur in any of the answers constructed by the algorithm. This subpart may be either the nominal exploited as starting point for E or, more generally, some part $D \sqcap \exists r^-.(C \sqcap ...)$ (or initially $D \sqcap \exists r^-.\{a\}$ with $\mathcal{K} \models C(a)$) where for the (current) concept C the existential restriction $\exists r.D$ is applied. While it is easy to see that the first case cannot occur, since the algorithm deals with every individual in the ontology during the initialization, the second one demands more consideration.

I. At first, we take a look at possible cases when the above C is not used as a current concept during the algorithm processing with some ontology $\mathcal{K}$:
a) *C was not chosen for initial algorithm call*
(i) $\mathcal{K} \models C(a)$ *does not hold for any individual a*: This means that C cannot be applied to search for existential restrictions that relate the individual a to some other anonymous one and, hence, it should not be used for the construction of CREs.
(ii) *C is not the most specific concept with $\mathcal{K} \models C(a)$*: Accordingly, there exists another most specific concept D with

[2] In other words, since a CRE generally describes a chain of relations that link the represented anonymous individual to an explicitly named one, the connection from a to a' means that we can refer to the same anonymous individuals by beginning with a as if we would directly start from a'.

$\mathcal{K} \models D(a)$ and $\mathcal{K} \models D \sqsubseteq C$ that is chosen instead. Due to $\mathcal{K} \models D \sqsubseteq C$ in combination with $\mathcal{K} \models C \sqsubseteq \exists r.A$ implying $\mathcal{K} \models D \sqsubseteq \exists r.A$ for some arbitrary existential restriction $\exists r.A$, we can create a CRE that is equivalent to E and leads to the same anonymous individual, which is sufficient for completeness as we are not interested in duplicate answers.

b) *C was not part of the previously selected existential restriction* – Some existential restriction $\exists r.C$ was not applied in the previous construction step, which refers to the cases discussed in the following part II.

II. Next, the situations are presented where C denotes the current concept, but the existential restriction $\exists r.D$ is not chosen by the algorithm:

a) *$\exists r.D$ does not occur in $\mathcal{K}$* – $\exists r.D$ can only be constructed by means of reasoning for which there are two possible cases in order to obtain $\mathcal{K} \models C \sqsubseteq \exists r.D$:

(i) *We have $\mathcal{K} \models C \sqsubseteq \exists r.A$ and $\mathcal{K} \models A \sqsubseteq D$*: With reference to Lemma 2, these conditions imply that $\exists r.D$ does not lead to a singular CRE and, hence, should be avoided.

(ii) *From $\mathcal{K} \models C \sqsubseteq \exists r.A$ and $\mathcal{K} \models C \sqsubseteq \forall r.B$, we derive $\mathcal{K} \models C \sqsubseteq \exists r.D$ with $D = A \sqcap B$*: In the algorithm, we also combine existential and universal restrictions, but with the condition that we consider every $\forall r.B$ with $\mathcal{K} \models C \sqsubseteq \forall r.B$. In this regard, $\exists r.D$ may be not constructed if it does not include every appropriate $\forall r.B$. Assuming that $\exists r.D'$ describes the existential restriction constructed by considering every suitable $\forall r.B$, i.e., D' is a conjunction containing every associated concept B, we get $\mathcal{K} \models D' \sqsubseteq D$, which again indicates non-singularity for $\exists r.D$.

b) $\mathcal{K} \not\models C \sqsubseteq \exists r.D$ – Obviously, there is no need to choose the existential restriction as it cannot be applied to relate the current concept C to another one.

c) $\mathcal{K} \models C \sqsubseteq \exists r.D$

(i) *$\exists r.D$ is not in $\mathcal{K}_{\exists}^{\mathsf{min}}(C)$*: With reference to Definition 3, there exists another $\exists r.A$ with $\mathcal{K} \models C \sqsubseteq \exists r.A$ for which $\mathcal{K} \models A \sqsubseteq D$, which means that $\exists r.D$ cannot guarantee singularity of a constructed CRE as stated before.

(ii) *Some $\exists r.D'$ with $\mathcal{K} \models D \equiv D'$ was already selected as candidate*: The CRE resulting from $\exists r.D$ only differs on a syntactical level from the CRE using $\exists r.D'$ while representing a semantically equivalent individual, hence, it can be ignored.

(iii) *It is the initial algorithm call and there exists a $r(a, b)$ for the base individual a with $\mathcal{K} \models r(a, b)$ and $\mathcal{K} \models D(b)$*: As explained in the uniqueness proof, $\exists r.D$ should not be applied, because it leads to a CRE being semantically equivalent to the one directly starting with b as base individual.

(iv) *$\exists r.D$ was used before and is blocked by the cycle detection*: Due to the cycle notation, all the CREs that can be created through the repeated usage of $\exists r.D$ are summarized in one, which means that the separate answer E is returned by the algorithm, too, although not explicitly.

In summary, the above arguments show that it is not possible to define such a separate CRE E without impeding its correctness and singularity or resulting in unwanted duplicate answers. Additionally, the cycle notation may be seen as a compromise to implicitly return a possibly infinite number of CREs in form of just one compact instance in order to ensure the termination of the algorithm.

Termination

The termination of our algorithm is closely linked to the fact that the considered Horn $\mathcal{ALC}$ ontology always consists of a finite number of axioms and concepts, such that the search for both existential and universal restrictions, as well as the extraction of suitable concepts for individuals during the initialization phase always come to a halt, while producing collections that also contain finitely many elements. Accordingly, the only potential for non-termination is given by iterations and recursive calls, where the first stop at all time as they only work with finite collections of elements, while the latter are performed for each element of the set (next) containing every existential restriction that has been chosen in order to further extend the current CRE. Since these elements are based on selected existential and universal restrictions from the ontology, there can only be a limited number of available instances. Therefore, non-termination may only happen due to the repeated usage of an element, which, however, does never occur as the reappearing of an existential restriction activates the cycle detection, thus blocking its repeated application, so that eventually no element can be used for a recursive call anymore.

B Implementation and Evaluation

Table 1 below provides an overview of the computed CREs obtained as answers for the query $\top$. The fact that the returned CREs for anonymous individuals are rather small (consisting mainly of only one existential restriction), while the percentage of concretely named individuals (depth = 0) is much higher, is actually not surprising, since ontologies are typically meant to represent knowledge about entities and their relations, for which it is desired that relevant individuals are also explicitly modeled. In this sense, anonymous individuals may be seen as "missing data" that obviously should be avoided when designing an ontology. This furthermore illustrates the importance of the here discussed topic, allowing us to find such individuals that are not yet explicitly represented and replace them by named ones in order to complete the ontology, as it is already the case for the tested HAO and DTO, for example. Besides, none of the computed answers contain any cycle, which may be explained by the small number of applied (inverse) existential restrictions in the CREs.

Ontology	#ans	depth (%)					
		0	1	2-4	5-7	8-10	11-12
codinteraction-A	30	70	30	-	-	-	-
ore_ont_2608	279	76	24	-	-	-	-
ore_ont_4516	284	76	24	-	-	-	-
ore_ont_3313	9,734	21	8	34	27	9	1
HAO	2,764	100	-	-	-	-	-
VO	176	95	5	-	-	-	-
DTO	3	100	-	-	-	-	-

Table 1: Overview of computed CREs returned as answers for the query $\top$, showing their total number (#ans) and the rounded percentage of instances with a certain depth, i.e., number of contained inverse existential restrictions