Rewriting Minimisations for Efficient Ontology-Based Query Answering

Tassos Venetis, Giorgos Stoilos, and Vasilis Vassalos

Athens University of Economics and Business
Department of Informatics, Athens, Greece.

Lemma 1. *Let $\mathcal{T}$ be a OWL 2 DL-TBox, let $\mathcal{Q}$ be a CQ, let $\mathcal{R}$ be a UCQ rewriting for $\mathcal{Q},\mathcal{T}$, let* ans *be a query answering system complete for some fragment $\mathcal{L}$ of OWL 2 DL, and let $\mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ be as defined in Definition* **??**. *Then, for every $\mathcal{A}$ we have $\mathsf{cert}(Q,\mathcal{T}\cup\mathcal{A}) = \mathsf{cert}(\mathcal{R}^{\mathcal{T}}_{\mathcal{L}},\mathcal{A}_s)$, where $\mathcal{A}_s$ is the saturation computed by* ans *for $\mathcal{T}\cup\mathcal{A}$.*

Proof. First, we show the following property for $\mathcal{L}$-derived queries:

> (♦): for every $\mathcal{Q}_1 \in \mathcal{R}$ either $\mathcal{Q}_1 \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ or some other $\mathcal{Q}_2 \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ exists such that $\mathcal{Q}_1$ is $\mathcal{L}$-derived by $\mathcal{Q}_1$.

Proof of Property (♦): The minimal subsets $\mathcal{T}'$ used in the derived relation can be used to construct a directed acyclic graph $G = \langle V, E\rangle$ of the queries in $\mathcal{R}$: more precisely, for $\mathcal{Q}_1$ and $\mathcal{Q}_2$ we have $\langle \mathcal{Q}_1, \mathcal{Q}_2\rangle \in E$ iff $\mathcal{Q}_1$ derives $\mathcal{Q}_2$ with minimal subset $\mathcal{T}_1$ and there is no $\mathcal{Q}'$ such that $\mathcal{Q}'$ derives $\mathcal{Q}_2$ with subset $\mathcal{T}' \subseteq \mathcal{T}_1$. It follows that G is a directed acyclic graph with one top element $\mathcal{Q}$ for which its minimal set is $\mathcal{T}' = \emptyset$. Let i denote the distance of a query from the root. We can show the property using induction over the distance i:

- By Definition **??**, $\mathcal{Q}$ is in $\mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ hence Property (♦) holds for $\mathcal{Q}$ (the root element).
- Assume Property (♦) holds for all queries up to a level i and consider some query $\mathcal{Q}_{i+1}$. Consider *all* parents of $\mathcal{Q}_{i+1}$ in G—that is, all queries $\mathcal{Q}^1,\ldots,\mathcal{Q}^\ell$ that (minimally) derive $\mathcal{Q}_{i+1}$. If at least one of them, call it $\mathcal{Q}^k$, $\mathcal{L}$-derives $\mathcal{Q}_{i+1}$ then we are done; by Property (♦) if $\mathcal{Q}^k \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ then $\mathcal{Q}_{i+1}$ is $\mathcal{L}$-derived by $\mathcal{Q}^k$ and $\mathcal{Q}^k \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ or, in a different case, some other $\mathcal{Q}'' \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$ exists that $\mathcal{L}$-derives $\mathcal{Q}^k$ and hence also $\mathcal{L}$-derives $\mathcal{Q}_{i+1}$. Assume in contrast that *no* $\mathcal{Q}^j$ $\mathcal{L}$-derives $\mathcal{Q}_{i+1}$. Then, since the TBoxes $\mathcal{T}^j$ used in the construction of G were minimal, any other $\mathcal{T}'$ for which some $\mathcal{Q}'$ derives $\mathcal{Q}_{i+1}$ must contain $\mathcal{T}^j$. Since $\mathcal{T}^j$ is not an $\mathcal{L}$-TBox $\mathcal{T}'$ cannot be an $\mathcal{L}$-TBox; hence, no query in $\mathcal{R}$ $\mathcal{L}$-derives $\mathcal{Q}_{i+1}$ and we must have $\mathcal{Q}_{i+1} \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$.

Now we show the claim.

Consider some arbitrary but fixed fragment $\mathcal{L}$ of OWL 2 DL and some ABox-saturation system ans complete for $\mathcal{L}$. Consider now some arbitrary tuple of individuals $\vec{a}$ s.t. $\vec{a} \in \mathsf{cert}(\mathcal{Q},\mathcal{T}\cup\mathcal{A})$. Since $\mathcal{R}$ is a UCQ rewriting for $\mathcal{Q},\mathcal{T}$ some $\mathcal{Q}' \in \mathcal{R}$ must exist such that $\vec{a} \in \mathsf{cert}(\mathcal{Q}',\mathcal{A})$. If $\mathcal{Q}' \in \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$, then for $\mathcal{A}_s$ the saturation computed by ans for $\mathcal{T}\cup\mathcal{A}$ we have $\mathcal{A}_s \supseteq \mathcal{A}$ hence by monotonicity of DLs $\vec{a} \in \mathsf{cert}(\mathcal{Q}',\mathcal{A}_s) \subseteq \mathsf{ans}(\mathcal{R}^{\mathcal{T}}_{\mathcal{L}},\mathcal{T}\cup\mathcal{A})$. Otherwise, assume that $\mathcal{Q}' \notin \mathcal{R}^{\mathcal{T}}_{\mathcal{L}}$.

By Property ($\blacklozenge$) some $\mathcal{Q}'' \in \mathcal{R}_{\mathcal{L}}^{\mathcal{T}}$ must exist that $\mathcal{L}$-derives $\mathcal{Q}'$. This means that some $\mathcal{T}'$ exists such that $\mathcal{Q}'$ is in the UCQ rewriting of $\mathcal{Q}'', \mathcal{T}'$. By definition of a UCQ rewriting (soundness property) we have that $\vec{a} \in \mathsf{cert}(\mathcal{Q}', \mathcal{A})$ implies $\vec{a} \in \mathsf{cert}(\mathcal{Q}'', \mathcal{T}' \cup \mathcal{A})$. Moreover, since $\mathcal{Q}''$ $\mathcal{L}$-derives $\mathcal{Q}'$, $\mathcal{T}'$ is an $\mathcal{L}$-TBox. Consequently, $\mathcal{T}' \subseteq \mathcal{T}|_{\mathcal{L}}$ hence $\mathsf{cert}(\mathcal{Q}'', \mathcal{T}' \cup \mathcal{A}) = \mathsf{ans}(\mathcal{Q}'', \mathcal{T}' \cup \mathcal{A}) \subseteq \mathsf{ans}(\mathcal{Q}'', \mathcal{T}|_{\mathcal{L}} \cup \mathcal{A}) = \mathsf{ans}(\mathcal{Q}'', \mathcal{T} \cup \mathcal{A})$ and $\mathcal{Q}'' \in \mathcal{R}_{\mathcal{L}}^{\mathcal{T}}$. $\square$