Addendum to:
Efficient Axiomatization of OWL 2 EL Ontologies from Data by means of Formal Concept Analysis

Francesco Kriegel

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

In the following, (C) refers to the conference version (Kriegel, 2024) and (E) refers to the extended version (Kriegel, 2023).

1. On Page 14 in (E), the phrase "all stated results and proofs but Lemma XV are revised versions" must be extended to "all stated results and proofs but Lemmas XV and XVII are revised versions".
2. Theorem 13 in (C) and (E) states that, for each finite interpretation $\mathcal{I}$, a complete TBox of $\mathcal{EL}$ CIs, RRs, and RIs satisfied in $\mathcal{I}$ can be computed in exponential time. First of all, there is no issue with this result, but a discussion with Carsten Lutz and Franz Baader revealed that some details on the representation of the CIs in this TBox are not properly explained. Please find them below in Section 1.

1 Further Details on Theorem 13

Recall that the TBox mentioned in Theorem 13 consists of a rewriting of the canonical CI base in Theorem 10 and of the RRs and RIs above Theorem 13. Regarding the CIs, we first compute the CI base $\mathsf{Can}(\mathcal{I},\mathcal{T})$ in Theorem 10, which consists of CIs of the form $\bigsqcap \mathbf{C} \sqsubseteq \bigsqcap \mathbf{D}$ with $\mathbf{C},\mathbf{D} \subseteq \mathbf{M}$. Next, we replace in every premise $\bigsqcap \mathbf{C}$ each existential restriction $\exists r.\, X^{\mathcal{I}}$ by $\exists r.\, (X^{\mathcal{I}}\restriction_n)$, where possible values for the unfolding depth n are explained on Page 5 in (C) and on Page 14 in (E). Last, the conclusions $\bigsqcap \mathbf{D}$, which are possibly cyclic $\mathcal{EL}^{\bot}_{\mathsf{si}}$ CDs, are rewritten into $\mathcal{EL}$ by means of concept variables as per Proposition XXII in (E), see also Page 5 in (C) for a summary.

Now there is a lack of clarity with respect to the modified premises, which are obtained by replacing existential restrictions $\exists r.\, X^{\mathcal{I}}$ with unfoldings $\exists r.\, (X^{\mathcal{I}}\restriction_n)$. If we would naïvely transform these into $\mathcal{EL}$ CDs by means of the recursion on Page 5 in (E), then we could in the worst case obtain $\mathcal{EL}$ CDs of double exponential size since $X^{\mathcal{I}}$ is computed from the powering, which has single exponential size, and the unfolding depth n could be at most single exponential as well.

Instead, it would be more efficient to represent the unfolded filler $X^{\mathcal{I}}\restriction_n$ as the pair of the MMSCD $X^{\mathcal{I}}$ and the unfolding depth n, i.e. without actually unfolding into a potentially huge $\mathcal{EL}$ CD. This is a single exponential encoding, but it is not immediately clear how it could be treated by a reasoner. More suitable is the representation by layered copies of the powering as explained on Page 24 in (E). With that, each unfolded filler $X^{\mathcal{I}}\restriction_n$ is equivalent to an acylic $\mathcal{EL}_{\mathsf{si}}$ CD of single exponential size, and reasoning can be done by means of the rule-based calculus in the preliminaries. However, this is still not in $\mathcal{EL}$. To end up with an $\mathcal{EL}$ TBox, we would finally need to rewrite these acylic $\mathcal{EL}_{\mathsf{si}}$ CDs by means of concept variables, similar as for the conclusions.

To this end, we use concept variables that stand for objects on a layer between 0 and n, and write up the layered copies as CIs from the leaves to the root. For example, the CD $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_3$ with $\cdot^{\mathcal{C}} \coloneqq \{(c,c)\colon r,\ (c,c)\colon s,\ c\colon A\}$ is represented by the concept variable $X_{c,3}$ and the CIs $A \sqsubseteq X_{c,0}$, $A \sqcap \exists r.\, X_{c,0} \sqcap \exists s.\, X_{c,0} \sqsubseteq X_{c,1}$, $A \sqcap \exists r.\, X_{c,1} \sqcap \exists s.\, X_{c,1} \sqsubseteq X_{c,2}$, $A \sqcap \exists r.\, X_{c,2} \sqcap \exists s.\, X_{c,2} \sqsubseteq X_{c,3}$.

More formally, we provide the following extension of Proposition XXII in (E) which additionally deals with rewriting of the premises as sketched above. Applying it to the canonical CI base in Theorem 10 yields an equivalent $\mathcal{EL}$ TBox (with variables) that has single exponential size. Theorem 13 should now be sufficiently underpinned.

Lemma. *Every TBox consisting of CIs of the form* $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \exists^{\mathsf{sim}}(\mathcal{D},d)$ *or* $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \bot$ *can be rewritten into an equivalent vTBox in polynomial time.*

Proof. Given a TBox $\mathcal{T}$ as above, we construct a vTBox $\exists \mathbb{X}.\, \mathcal{T}'$ as follows. Without loss of generality, we assume that all CDs in the premises share the same interpretation $\mathcal{C}$ and that all CDs in the conclusions share the same interpretation $\mathcal{D}$. To obtain a small vTBox these two interpretations $\mathcal{C}$ and $\mathcal{D}$ could be weakly reduced, but this is not necessary for our goal.

Initially, we do the following.

(a) For each CI $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \exists^{\mathsf{sim}}(\mathcal{D},d)$ in $\mathcal{T}$, we add the CI $X_{c,n} \sqsubseteq Y_d$ to $\mathcal{T}'$ and the concept variables $X_{c,n}$, Y_d to $\mathbb{X}$.

(b) For each CI $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \bot$ in $\mathcal{T}$, we add the CI $X_{c,n} \sqsubseteq \bot$ to $\mathcal{T}'$ and the concept variable $X_{c,n}$ to $\mathbb{X}$.

To encode the meaning of the premise variables $X_{c,n}$, we additionally need to do the following.

(c) Whenever a new concept variable $X_{c,n}$ with $n > 0$ has been added to $\mathbb{X}$, then we add the CI $\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqcap \bigsqcap\{ \exists r.X_{c',n-1} \mid (c,c') \in r^{\mathcal{C}} \} \sqsubseteq X_{c,n}$ to $\mathcal{T}'$ and all occurring concept variables $X_{c',n-1}$ to $\mathbb{X}$.
(d) Whenever a new concept variable $X_{c,0}$ has been added to $\mathbb{X}$, then we add the CI $\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqsubseteq X_{c,0}$ to $\mathcal{T}'$.

In a similar way, we encode the meaning of the conclusion variables Y_d.

(e) Whenever a new concept variable Y_d has been added to $\mathbb{X}$, then we add the CI $Y_d \sqsubseteq \bigsqcap\{ A \mid d \in A^{\mathcal{D}} \} \sqcap \bigsqcap\{ \exists r.Y_{d'} \mid (d,d') \in r^{\mathcal{D}} \}$ to $\mathcal{T}'$ and all occurring concept variables $Y_{d'}$ to $\mathbb{X}$.

Compared to Proposition XXII in (E), Rules (c) and (d) are new and take care of rewriting the unfolded premises.

We show that $\mathcal{T}$ and $\exists\mathbb{X}.\mathcal{T}'$ have the same models and are thus equivalent. Let $\mathcal{I}$ be a model of $\mathcal{T}$. We define the variable assignment $\mathcal{Z}$ by $\mathcal{Z}(X_{c,n}) \coloneqq (\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}}$ and $\mathcal{Z}(Y_d) \coloneqq (\exists^{\mathsf{sim}}(\mathcal{D},d))^{\mathcal{I}}$, and show that $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$.

1. Assume that the CI $X_{c,n} \sqsubseteq \bot$ is in $\mathcal{T}'$, i.e. $\mathcal{T}$ contains the CI $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \bot$. We obtain that $(X_{c,n})^{\mathcal{I}[\mathcal{Z}]} = \mathcal{Z}(X_{c,n}) = (\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}} = \emptyset$, where the second equality holds by definition of $\mathcal{Z}$ and the third equality holds since $\mathcal{I}$ is a model of $\mathcal{T}$. Thus $\mathcal{I}[\mathcal{Z}]$ satisfies $X_{c,n} \sqsubseteq \bot$.
2. Consider a CI $X_{c,n} \sqsubseteq Y_d$ in $\mathcal{T}'$, i.e. the CI $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \exists^{\mathsf{sim}}(\mathcal{D},d)$ is in $\mathcal{T}$. We obtain that $(X_{c,n})^{\mathcal{I}[\mathcal{Z}]} = \mathcal{Z}(X_{c,n}) = (\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}} \subseteq (\exists^{\mathsf{sim}}(\mathcal{D},d))^{\mathcal{I}} = \mathcal{Z}(Y_d) = (Y_d)^{\mathcal{I}[\mathcal{Z}]}$, where the second and fourth equality holds by definition of $\mathcal{Z}$ and the third inclusion holds since $\mathcal{I}$ is a model of $\mathcal{T}$. Thus $\mathcal{I}[\mathcal{Z}]$ satisfies $X_{c,n} \sqsubseteq Y_d$.
3. Let $\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqsubseteq X_{c,0}$ be a CI in $\mathcal{T}'$. Recall that $(X_{c,0})^{\mathcal{I}[\mathcal{Z}]} = \mathcal{Z}(X_{c,0}) = (\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_0)^{\mathcal{I}}$ and that $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_0 = \bigsqcap\{ A \mid c \in A^{\mathcal{C}} \}$ (see Page 5 in (E)). With that, we obtain $(\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \})^{\mathcal{I}[\mathcal{Z}]} = (\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \})^{\mathcal{I}} = (X_{c,0})^{\mathcal{I}[\mathcal{Z}]}$, i.e. $\mathcal{I}[\mathcal{Z}]$ satisfies the considered CI.
4. Next, consider a CI $\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqcap \bigsqcap\{ \exists r.X_{c',n-1} \mid (c,c') \in r^{\mathcal{C}} \} \sqsubseteq X_{c,n}$ in $\mathcal{T}'$. Recall that $(X_{c,n})^{\mathcal{I}[\mathcal{Z}]} = \mathcal{Z}(X_{c,n}) = (\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}}$ and that $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n = \bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqcap \bigsqcap\{ \exists r.(\exists^{\mathsf{sim}}(\mathcal{C},c')\restriction_{n-1}) \mid (c,c') \in r^{\mathcal{C}} \}$ (see Page 5 in (E)). Since we also have $(X_{c',n-1})^{\mathcal{I}[\mathcal{Z}]} = \mathcal{Z}(X_{c',n-1}) = (\exists^{\mathsf{sim}}(\mathcal{C},c')\restriction_{n-1})^{\mathcal{I}}$ for each c', it follows that $\mathcal{I}[\mathcal{Z}]$ satisfies the considered CI.
5. Let $Y_d \sqsubseteq \bigsqcap\{ A \mid d \in A^{\mathcal{D}} \} \sqcap \bigsqcap\{ \exists r.Y_{d'} \mid (d,d') \in r^{\mathcal{D}} \}$ be in $\mathcal{T}'$, and consider an element $x \in (Y_d)^{\mathcal{I}[\mathcal{Z}]}$. Recall that $(Y_d)^{\mathcal{I}[\mathcal{Z}]} = \mathcal{Z}(Y_d) = (\exists^{\mathsf{sim}}(\mathcal{D},d))^{\mathcal{I}}$, and so there is a simulation from $\mathcal{D}$ to $\mathcal{I}$ containing (d,x). Thus, $d \in A^{\mathcal{D}}$ implies $x \in A^{\mathcal{I}} = A^{\mathcal{I}[\mathcal{Z}]}$. Likewise, $(d,d') \in r^{\mathcal{D}}$ implies that there is x' with $(x,x') \in r^{\mathcal{I}}$ and such that there is a simulation from $\mathcal{D}$ to $\mathcal{I}$ containing (d',x'), i.e. $x' \in (Y_{d'})^{\mathcal{I}[\mathcal{Z}]}$, and therefore $x \in (\exists r.Y_{d'})^{\mathcal{I}[\mathcal{Z}]}$. We conclude that $\mathcal{I}[\mathcal{Z}]$ satisfies the considered CI.

Conversely, assume that $\mathcal{I}$ is a model of $\exists\mathbb{X}.\mathcal{T}'$, i.e. there is a variable assignment $\mathcal{Z}$ such that $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$.

First, we show by induction on n that $(\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}} \subseteq \mathcal{Z}(X_{c,n})$.

- Recall from Page 5 in (E) that $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_0 = \bigsqcap\{ A \mid c \in A^{\mathcal{C}} \}$. Since $\mathcal{T}'$ contains the CI $\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqsubseteq X_{c,0}$ and $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$, we obtain that $(\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_0)^{\mathcal{I}} \subseteq \mathcal{Z}(X_{c,0})$.
- Regarding the induction step, assume that $n > 0$. Recall from Page 5 in (E) that $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n = \bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqcap \bigsqcap\{ \exists r.\exists^{\mathsf{sim}}(\mathcal{C},c')\restriction_{n-1} \mid (c,c') \in r^{\mathcal{C}} \}$. Furthermore, $\mathcal{T}'$ contains the CI $\bigsqcap\{ A \mid c \in A^{\mathcal{C}} \} \sqcap \bigsqcap\{ \exists r.X_{c',n-1} \mid (c,c') \in r^{\mathcal{C}} \} \sqsubseteq X_{c,n}$. Since $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$ and the induction hypothesis yields that $(\exists^{\mathsf{sim}}(\mathcal{C},c')\restriction_{n-1})^{\mathcal{I}} \subseteq \mathcal{Z}(X_{c',n-1})$ for each involved c', we obtain that $(\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}} \subseteq \mathcal{Z}(X_{c,n})$.

Next, consider a CI $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \bot$ in $\mathcal{T}$. Then $\mathcal{T}'$ contains the CI $X_{c,n} \sqsubseteq \bot$ and thus $\mathcal{Z}(X_{c,n}) = \emptyset$. Since $(\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}} \subseteq \mathcal{Z}(X_{c,n})$, we conclude that $\mathcal{I}$ satisfies the considered CI.

Last, let $\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n \sqsubseteq \exists^{\mathsf{sim}}(\mathcal{D},d)$ be a CI in $\mathcal{T}$. Then $\mathcal{T}'$ contains the CI $X_{c,n} \sqsubseteq Y_d$ and, since $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$, we infer that $\mathcal{Z}(X_{c,n}) \subseteq \mathcal{Z}(Y_d)$. We already know that $(\exists^{\mathsf{sim}}(\mathcal{C},c)\restriction_n)^{\mathcal{I}} \subseteq \mathcal{Z}(X_{c,n})$, and it remains to show that $\mathcal{Z}(Y_d) \subseteq (\exists^{\mathsf{sim}}(\mathcal{D},d))^{\mathcal{I}}$. To this end, let $x \in \mathcal{Z}(Y_d)$ and consider the relation $\mathfrak{S} \coloneqq \{ (v,u) \mid u \in \mathcal{Z}(Y_v) \}$. Since $x \in \mathcal{Z}(Y_d)$, $\mathfrak{S}$ contains (d,x). We proceed with verifying that $\mathfrak{S}$ is a simulation from $\mathcal{D}$ to $\mathcal{I}$, which then yields that $x \in (\exists^{\mathsf{sim}}(\mathcal{D},d))^{\mathcal{I}}$.

(S1) Let $(v,u) \in \mathfrak{S}$, i.e. $u \in \mathcal{Z}(Y_v)$, and further let $v \in A^{\mathcal{D}}$. Then $\mathcal{T}'$ contains a CI $Y_v \sqsubseteq A \sqcap \cdots$ by Rule (e). Since $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$, we have that $\mathcal{Z}(Y_v) \subseteq A^{\mathcal{I}}$. We conclude that $u \in A^{\mathcal{I}}$.
(S2) Let $(v,u) \in \mathfrak{S}$, i.e. $u \in \mathcal{Z}(Y_v)$, and further let $(v,v') \in r^{\mathcal{D}}$. Then $\mathcal{T}'$ contains a CI $Y_v \sqsubseteq \exists r.Y_{v'} \sqcap \cdots$ by Rule (e). Since $\mathcal{I}[\mathcal{Z}]$ is a model of $\mathcal{T}'$, we have that $\mathcal{Z}(Y_v) \subseteq (\exists r.Y_{v'})^{\mathcal{I}}$, i.e. $(u,u') \in r^{\mathcal{I}}$ for some $u' \in \mathcal{Z}(Y_{v'})$. The latter yields $(v',u') \in \mathfrak{S}$.

Finally, it remains to show that the rewriting can be obtained in polynomial time. Rules (a) and (b) are applied once for each CI in $\mathcal{T}$ and yield CIs of constant size. Rules (c) and (d) yield the CIs that describe the layered copies of the premise structure $\mathcal{C}$. The size of each introduced CI is linear in $\mathcal{C}$, and the overall number of introduced CIs is polynomial in $\mathcal{C}$ and the respective unfolding depth n. Thus, the exhaustive application of these two rules finishes in polynomial time. The last Rule (e) merely rewrites the conclusion structure $\mathcal{D}$ into CIs and thus finishes in linear time. □

References

Francesco Kriegel (2023). *Efficient Axiomatization of OWL 2 EL Ontologies from Data by means of Formal Concept Analysis (Extended Version)*. LTCS-Report 23-01. Dresden, Germany: Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden. DOI: 10.25368/2023.214.

Francesco Kriegel (2024). Efficient Axiomatization of OWL 2 EL Ontologies from Data by means of Formal Concept Analysis. In: *Proceedings of the 38th Annual AAAI Conference on Artificial Intelligence (AAAI 2024), February 20–27, 2024, Vancouver, Canada*. DOI: 10.1609/aaai.v38i9.28930.