Certain Answers in a Rough World[*]

Rafael Peñaloza[1,2], Veronika Thost[1], and Anni-Yasmin Turhan[1]

[1] Theoretical Computer Science, TU Dresden, Germany
[2] Center for Advancing Electronics Dresden
`{penaloza,thost,turhan}@tcs.inf.tu-dresden.de`

One of the main challenges in knowledge representation and reasoning is still to cope with vague and imprecise information in an adequate manner. This imprecision is found in many knowledge domains, particularly medicine and life sciences. A typical source of imprecision in these domains arises from the level of detail in which the knowledge is described. For example, a disease is usually diagnosed through a series of symptoms that a patient presents, but two individuals, say ANA and BOB, showing the same symptoms might in fact suffer from different maladies. Thus, while these individuals might be *equivalent* from a symptomatic point of view, they might be classified into different illness classes.

One of the many approaches suggested for handling imprecise knowledge is based on *rough* approximations. Generally speaking, the individuals in a domain are partitioned into equivalence classes, based on their indiscernibility according to the current level of detail. An individual belongs to the *upper approximation* of a class C (denoted $\overline{C}$), if it is indiscernible from some element of C. For instance, ANA and BOB are in the same *symptomatic equivalence* class. If BOB is diagnosed with, say the Cooties, then ANA potentially has the Cooties, too. In rough terminology, ANA is in the upper approximation of Cooties ($\overline{\text{Cooties}}$). An analogous lower approximation of a class can be defined, too. Intuitively, $\underline{C}$ contains the prototypical elements of the class C: if an element x belongs to $\underline{C}$, then every element indiscernible from x is guaranteed to belong to C.

Rough extensions of Description Logics (DLs) [1] have been proposed as a formalism for handling these upper and lower approximations [5]. An important example is the rough DL $\mathcal{EL}_\rho$, which extends $\mathcal{EL}$ with two new rough constructors. Formally, $\mathcal{EL}_\rho$ concepts are built from concept names A and role names r through the grammar rule $C ::= A \mid \top \mid C \sqcap C \mid \exists r.C \mid \overline{C} \mid \underline{C}$. The semantics of this logic is based on interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \rho_{\mathcal{I}})$ that extend standard interpretations by an equivalence relation $\rho_{\mathcal{I}}$ over the elements of $\Delta^{\mathcal{I}}$. The interpretation function is extended to the classical constructors in the usual way, and to the rough constructors by setting

$$(\overline{C})^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y \in C^{\mathcal{I}}.x\rho_{\mathcal{I}}y\},$$
$$(\underline{C})^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \forall y \in \Delta^{\mathcal{I}}.x\rho_{\mathcal{I}}y \Rightarrow y \in C^{\mathcal{I}}\}.$$

While the indiscernibility relation ρ is not used in the TBox directly, it is admitted in role assertions in $\mathcal{EL}_\rho$-ABoxes and role atoms in conjunctive queries.
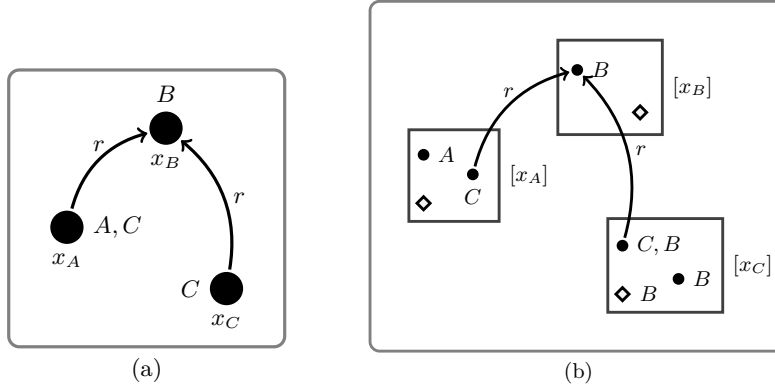

Fig. 1. (Partial) canonical interpretations for an $\mathcal{EL}$ (a) and an $\mathcal{EL}_\rho$ (b) KB.

It has been shown that standard reasoning, such as subsumption or instance checking is decidable in this logic in polynomial time [3]. Intuitively, the idea is to construct a minimal model, called the *canonical model*, that describes all the standard relations between named individuals and concept names.

Interestingly, there is a very tight connection between canonical models for $\mathcal{EL}$ knowledge bases, and those for $\mathcal{EL}_\rho$ knowledge bases. In $\mathcal{EL}$, the canonical model has a domain element x_C for every concept appearing in the KB. This element x_C acts as a representative for the concept C, and every concept containing this element is guaranteed to be a subsumer of C. In the case of $\mathcal{EL}_\rho$, the canonical model can be understood as a more detailed view into the classical canonical model for $\mathcal{EL}$. While each concept C appearing in the KB still produces a representative x_C, this representative defines a whole equivalence class, rather than a single domain element. This equivalence class, in turn, provides all the information regarding the upper and lower approximations of the concept C. This intuition is depicted in Figure 1. For example, the interpretation in part (b) of the figure expresses that $A \sqsubseteq \overline{C}$, through the auxiliary node in the class $[x_A]$, and that $C \sqsubseteq \underline{B}$, through the distinguished auxiliary (diamond shaped) element in the class $[x_C]$. Notice that this figure does not depict the whole canonical interpretation; some conclusions are missing from it.

Canonical interpretations are helpful for answering conjunctive queries w.r.t. $\mathcal{EL}$ knowledge bases [2]. Essentially, the canonical interpretation is extended with representatives for each individual name in the ABox; one can then use the information encoded in this interpretation, and answer the queries w.r.t. this interpretation only. Unfortunately, a naïve application of this idea would provide erroneous answers to some queries; for example, an interpretation like the one in Figure 1 (a) could return (x_A, x_B) to the query $\phi(x, y) = \exists z.r(x, z) \wedge r(y, z)$, although this is not true in all models of the KB. To avoid this problem, one can first *rewrite* the query into a first-order query, which can then be answered over the canonical interpretation. This is known as the *combined approach* [2].
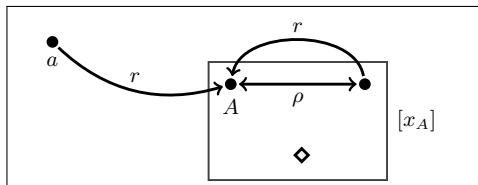
Fig. 2. Canonical interpretation of $\mathcal{K}_0$.

We exploit the connection between canonical models, and provide a combined approach for conjunctive query answering in $\mathcal{EL}_\rho$.

Answering Rough Queries in $\mathcal{EL}_\rho$

As mentioned already, the canonical interpretation $\mathcal{I}$ of an $\mathcal{EL}_\rho$ KB $\mathcal{K}$ provides a compact representation of all models of $\mathcal{K}$ from which all the subsumption and instance relationships among the individual and concept names that appear in $\mathcal{K}$ can be easily read. While all instance queries can be easily answered from this interpretation, for conjunctive queries the situation is more complex. The issue with answering conjunctive queries over the canonical interpretation arises from the succinctness of this interpretation. Mainly, in this interpretation there is one domain element x_C that represents the whole concept C. Thus, two individual names a, b that belong to the concept $\exists r.C$ will have the same r-successor x_C in $\mathcal{I}$, although in general these successors need not be the same.

Clearly, since $\mathcal{EL}_\rho$ is an extension of $\mathcal{EL}$, all the rewriting rules for query answering in $\mathcal{EL}$ are also required in the rough setting. However, the structure of the canonical model of an $\mathcal{EL}_\rho$ KB is more complex: each symbol gets a representative equivalence class, which is needed to convey the rough approximations of the concepts. Thus, some elements are connected by an equivalence relation, that can be understood as a symmetric, transitive and reflexive role ρ. This special kind of role needs to be treated carefully to avoid erroneous answers to a query. As a simple example, consider the KB $\mathcal{K}_0$ containing the assertion $\exists r.A(a)$ and the GCI $A \sqsubseteq \overline{\exists r.A}$. The canonical interpretation of this KB is depicted in Figure 2. The query $\phi(x) = \exists y, z.r(x,y) \wedge \rho(y,z) \wedge r(z,y)$ over this interpretation would return a as an answer, although it is not true that a satisfies this property in all models of $\mathcal{K}_0$. It is thus important to adapt the rewriting technique to handle also the equivalence relation that the rough constructors and the role assertions for ρ yield.

We give a computation algorithm following the combined approach for answering conjunctive queries in the rough DL $\mathcal{EL}_\rho$. As in case of $\mathcal{EL}$, the approach consists in computing the canonical interpretation $\mathcal{I}$ that simulates all models of the input KB $\mathcal{K}$, which can be done in polynomial time. This interpretation is first used as a guideline for rewriting the conjunctive query ϕ into a first-order query $\widehat{\phi}$, and then as the finite domain over which $\widehat{\phi}$ is answered. As a result, we obtain an effective method for answering conjunctive queries that can handle imprecision described as rough approximations of a concept or as indiscernibility information using ρ in role assertions. The full details can be found in [4].

References

1. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2nd edn. (2007)
2. Lutz, C., Toman, D., Wolter, F.: Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In: Proc. 21st Int. Joint Conf. on Artif. Intel. (IJCAI 2009). pp. 2070–2075 (2009)
3. Peñaloza, R., Zou, T.: Roughening the $\mathcal{EL}$ envelope. In: Fontaine, P., Ringeissen, C., Schmidt, R.A. (eds.) Proceedings of the 2013 International Symposium on Frontiers of Combining Systems (FroCoS 2013). Lecture Notes in Computer Science, vol. 8152, pp. 71–86. Springer-Verlag, Nancy, France (2013)
4. Peñaloza, R., Thost, V., Turhan, A.Y.: Answering conjunctive queries in the rough description logic $\mathcal{ELH}_{\perp\rho}$. LTCS-Report 14-04, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2014), see `http://lat.inf.tu-dresden.de/research/reports.html`.
5. Schlobach, S., Klein, M.C.A., Peelen, L.: Description logics with approximate definitions - precise modeling of vague concepts. In: Veloso, M.M. (ed.) Proc. 20th Int. Joint Conf. on Artif. Intel. (IJCAI 2007). pp. 557–562 (2007)