Exercises: OBQA Basics

Exercise 1: Chasing

There are two homomorphisms from the body of the first rule to \(D\). By mapping \(x\) to \(a\) and \(y\) to \(b\), one get:

\[ D_1 = D \cup \{s(b, \bot_1)\}. \]

The second homomorphism maps \(x\) to \(b\) and \(y\) to \(c\), but can be extended in a homomorphism from the head of the rule to \(D\) (and hence \(D_1\)), by mapping \(z\) to \(d\). Hence the first rule is not applicable through this second homomorphism.

We consider the second rule. There are two homomorphisms from the body of the second rule to \(D_1\). The first one maps \(x\) to \(c\) and \(y\) to \(d\), while the second one maps \(x\) to \(b\) and \(y\) to \(\bot_1\). Applying the first one on \(D_1\) creates

\[ D_2 = D_1 \cup \{p(c, d, \bot_2)\}. \]

The second one is still applicable on \(D_2\), yielding \(D_3\):

\[ D_3 = D_2 \cup \{p(b, \bot_1, \bot_3)\}. \]

There are two homomorphisms from the body of the third rule to \(D_3\): the first one maps \(x\) to \(a\), \(y\) to \(b\), \(z\) to \(\bot_1\) and \(t\) to \(\bot_3\). The second one maps \(x\) to \(b\), \(y\) to \(c\), \(z\) to \(d\) and \(t\) to \(\bot_2\).

Applying the first one generates

\[ D_4 = D_3 \cup \{s(b, \bot_1), q(\bot_1)\}, \]

while applying the second one generates:

\[ D_5 = D_4 \cup \{s(c, d), q(d)\}. \]

No rule is newly applicable, hence \(D_5\) is the result of the considered chase sequence.

It holds that the chase w.r.t. \(R\) is finite whatever the initial instance is. Intuitively, this is because there are no possible cycles of rule applications. Indeed, the first rule can trigger an application of the second rule, but not of the two others. The second rule can trigger an application of the third rule, but not of the two others. Finally, the third rule cannot trigger an application of any of the rules: indeed, while there is a unification between the head of that rule and the body of the second rule, if an atom \(s(x, y)\) has been created by an application of the third rule, then there must be a \(z\) such that \(p(x, y, z)\) is in the current instance, otherwise the third rule would not have been applicable. Hence, the second rule is not applicable by mapping its body to \(s(x, y)\), as this homomorphism could necessarily be extended into a homomorphism from its head to the current instance.

Exercise 2: Atomic Head

Let \(\forall x \forall y B(x, y) \rightarrow \exists z H(y, z)\) be an existential rule. We create the following rules:

- \(\forall x \forall y B(x, y) \rightarrow \exists z p_R(y, z)\), where \(p_R\) is a fresh predicate of suitting arity;
- for each atom \(\alpha \in H(y, z)\), one rule \(\forall y \forall z p_R(y, z) \rightarrow \alpha\).
Then given a ruleset $\mathcal{R}$, we define $\mathcal{R}'$ as the union of the sets built for each rule of $\mathcal{R}$ as above.

We can show that for all instance $D$ on the original vocabulary, for all conjunctive query $q$ on the original vocabulary, $D, \mathcal{R} \models q$ if and only if $D, \mathcal{R}' \models q$. To do that, we show that for any chase sequence from $D$ w.r.t. $\mathcal{R}$, creating $\hat{D}$, we can build a chase sequence from $D$ w.r.t. $\mathcal{R}'$, creating $\hat{D}'$, such that there is a homomorphism from $\hat{D}$ to (the restriction on the original vocabulary of) $\hat{D}'$. We show the same thing conversely.

To build such chase sequences, we replace a rule of application in the original rule set by applying all the corresponding rules in the modified ruleset.

**Exercise 3: Weak Acyclicity**

The predicate graph contains four vertices $(r, 1), (r, 2), (s, 1), (s, 2)$ and the following edges:

- **normal edges:**
  - from $(r, 2)$ to $(r, 1)$ (first rule)
  - from $(s, 1)$ to $(r, 2)$ (second rule)
  - from $(s, 2)$ to $(r, 1)$ (second rule)
  - from $(r, 1)$ to $(r, 1)$ (third rule)
  - from $(r, 2)$ to $(r, 2)$ (third rule)

- **special edges:**
  - from $(r, 2)$ to $(r, 2)$ (first rule)

Note that there exists a critical cycle, being the only special edge.

If there are no critical cycle, then for any position $(p, i)$, there is an upper bound on the number of special edges on a path leading to $(p, i)$. We denote this number by the rank of $(p, i)$. We show by induction on the rank that there is then a bound the number of terms that can appear at position $(p, i)$ in an atom of the chase, which in turn upperbounds the size of the chase.

To get this bound, note the following properties of the edges:

- if there is a rule application that creates an atom having a term $t$ at position $(q, j)$, where $t$ was at position $(p, i)$ in an atom used in the body, then there is a normal edge from $(p, i)$ to $(q, j)$;
- if there is a rule application that creates an atom containing a fresh null at position $(q, j)$, then there is a special edge from the position $(p, i)$ of any frontier term to $(q, j)$ in the predicate graph.

Noticing this, let us consider the case of positions of rank $0$. As there is no path containing a special edge leading to $(p, i)$ of rank $0$ in the predicate graph, all the terms that may appear at position $(p, i)$ in the chase actually belong to the original terms: there are thus at most $|\text{terms}(D)|$ terms appearing in such a position.

Let us know that we have a bound $n_i$ for the number of terms appearing at a position of rank $i$ for any $i \leq k$, and let us derive a bound $n_{k+1}$ for positions of rank $k + 1$. A term appearing at such a position is either a term created in such a position (and then propagated to other positions of the same rank) or a term created in a position of strictly smaller rank and then propagated. The second set of terms can be upper bounded by $\Sigma_{i=0}^k n_i$. The term created must have been created by a rule application: each rule can be applied at most once by mapping of its frontier terms to terms of rank at most $k$: there are thus at most $|\mathcal{R}| \times (\Sigma_{i=0}^k n_i)^w$ possible rule applications, where $w$ is the maximal size of a frontier. Hence, one can set $n_{k+1} = \Sigma_{i=0}^k n_i + e \times |\mathcal{R}| \times (\Sigma_{i=0}^k n_i)^w$, where $e$ is the maximal number of existentially quantified variables appearing in a rule head.

Finally, let us consider the following rule:

$$\forall x \forall y \ r(x, y) \land r(y, x) \rightarrow \exists z \exists t \ r(x, z) \land r(z, t) \land r(t, x).$$

It is easy to check that its predicate graph contains critical cycle. However, an application of that rule cannot trigger a new application of itself, which implies that the chase stops for any instance with respect to that single rule.