First-Order Logic over Active Domain

We consider First-Order Logic (FOL) exactly as used in relational database queries. This requires to drop functions except for constants.

In particuar we assume to have a countably infinite set of constants Δ.

Moreover we assume that the interpretation of constants is the identity function, that is constants are interpreted as themselves.

This allows us to drop also the interpretation of constants from our interpretations, which now have the form:

$$\mathcal{I} = (\Delta, P_1^{\mathcal{I}}, P_2^{\mathcal{I}}, \ldots, P_n^{\mathcal{I}}).$$

First-Order Logic over Active Domain

We introduce special interpretations based on the *active domain*, denoted $\textsc{adom}$, which is the subset of Δ whose constants actually appear in the relations interpreting the predicate symbols. In particular given $\mathcal{I}$ we get the active domain interpretation:

$$\overline{\mathcal{I}} = (\textsc{adom}, P_1^{\mathcal{I}}, P_2^{\mathcal{I}}, \ldots, P_n^{\mathcal{I}}),$$

We call $\text{FOL}_{\textsc{adom}}$ the variant of FOL where formulas are to be interpreted over the active domain only: in particular we denote quantifications in this case as: $\exists x \in \textsc{adom}.\varphi$ and $\forall x \in \textsc{adom}.\varphi$

Equivalence of FOL and FOL$_{\text{ADOM}}$

Theorem

Every closed FOL formula ψ with no function symbols except constants, all of which must be present in the active domain, there is a FOL$_{\text{ADOM}}$ formula $\overline{\psi}$, such that

$$\mathcal{I} \models \psi \quad \textit{iff} \quad \overline{\mathcal{I}} \models \overline{\psi}.$$

We show the theorem constructively by define the corresponding formula $\overline{\psi}$ of ψ.

We proceed by induction on the subformulas of ψ:

- $\overline{P(t_1, \ldots, t_n)} \ = \ P(t_1, \ldots, t_n)$
- $\overline{(y = z)} \ = \ (y = z)$
- $\overline{\neg\varphi} \ = \ \neg\overline{\varphi}$
- $\overline{\varphi_1 \wedge \varphi_2} \ = \ \overline{\varphi_1} \wedge \overline{\varphi_2}$
- $\overline{\exists x.\varphi} \ = \ \exists x \in \text{ADOM}.\overline{\varphi} \vee \varphi_{fv} \vee \varphi_\infty$.
  where φ_{fv} and φ_∞ are defined below and are needed to suitably treat those values which are not in the active domain.

Equivalence of FOL and FOL$_{\text{ADOM}}$

The formula φ_{fv} captures the cases in which $x = z$ for some variable z occurring free in $\overline{\varphi}$. Namely:

$$\varphi_{fv} = \bigvee_{z \in FV} (\overline{\varphi})^x_z$$

where

- $FV = \{z_1, \ldots z_m\}$ are the variables occurring free in $\overline{\psi}$ and
- $(\overline{\varphi})^x_z$ stands for the formula obtained from $\overline{\psi}$ by syntactically replacing variable x with variable z.

Equivalence of FOL and FOL$_{\text{ADOM}}$

The formula φ_∞ captures the cases in which $x = d$ for some value $d \notin \text{ADOM}$. Namely:

$$\varphi_\infty = rem_x(\overline{\varphi})$$

where $rem_x(\overline{\varphi})$ is inductively defined over the subformulas of $\overline{\varphi}$ as follows (we assume wlog quantified variables are all distinct and different from x):

- $rem_x((x = x)) = true$
- $rem_x((x = v)) = false$
- $rem_x(P(\ldots, x, \ldots)) = false$
- $rem_x(P(\ldots, v, \ldots)) = P(\ldots, v, \ldots)$, if x does not occur
- $rem_x(\neg\phi) = \neg rem_x(\phi)$
- $rem_x(\phi_1 \wedge \phi_2) = rem_x(\phi_1) \wedge rem_x(\phi_2)$
- $rem_x(\exists y.\phi) = \exists y.rem_x(\phi)$

Equivalence of FOL and FOL$_{\text{ADOM}}$

To prove the theorem, we make use of the following the key Lemma.

Lemma

Let ψ be a (possibly open) formula with no function symbols except constants, all of which must be present in the active domain, then for all assignment α of free variables in ψ:

$$\mathcal{I}, \alpha \models \psi \quad \textit{iff} \quad \mathcal{I}, \alpha \models \overline{\psi}.$$

Proof. By induction on the depth of nested quantification in ψ.
Base case: there are no quantification. Then the result is immediate considering that the extension of predicate is the same in both databases, and equality atoms mention only constant from the active domain.
Induction on the formula guarantees that the same holds for all formulas not introducing variables (not's and and's).

Continues

Equivalence of FOL and FOL$_{\text{ADOM}}$

Continuing

Inductive case: we assume that the thesis holds for formulas with k nested quantifiers, and we shot it holds for those with $k+1$.
Let us consider formulas of the form $\exists x.\varphi$. For them
$\overline{\exists x.\varphi} \;=\; \exists x \in \text{ADOM}.\overline{\varphi} \vee \varphi_{fv} \vee \varphi_\infty$

- Assume that $\mathcal{I}, \alpha \models \exists x.\varphi$ by assigning to x a value $d \in \text{ADOM}$. Then with the same assignment for x the formula $\exists x \in \text{ADOM}.\overline{\varphi} \vee \varphi_{fv} \vee \varphi_\infty$ is true.

Continues

Equivalence of FOL and FOL$_{\text{ADOM}}$

Continuing

- Assume that $\mathcal{I}, \alpha \models \exists x.\varphi$ by assigning to x a value $d \notin \text{ADOM}$, then we can concentrate on $\varphi_{fv} \vee \varphi_\infty$. Suppose that for some other free variable $z \in FV$ we have that $\alpha(z) = d$. Then $I, \alpha[x \leftarrow d] \models \varphi$ iff $I, \alpha \models \varphi_{fv}$. Suppose that for all free variables $z \in FV$ we have that $\alpha(z) \neq d$. Then we show by induction of φ that $I, \alpha[x \leftarrow d] \models \varphi$ iff $I, \alpha \models \varphi_\infty$.
  - $I, \alpha[x \leftarrow d] \models (x = x)$ iff $I, \alpha \models rem_x((x = x))$, since $rem_x((x = x)) = true$.
  - $I, \alpha[x \leftarrow d] \models (x = v)$ iff $I, \alpha \models rem_x((x = v))$, since $rem_x((x = v)) = true$, notice that v is either a constant or a (free) variable different from x.
  - $I, \alpha[x \leftarrow d] \models P(\ldots, x, \ldots)$ iff $I, \alpha \models rem_x(P(\ldots, x, \ldots))$, $rem_x(P(\ldots, x, \ldots)) = false$, notice $d \notin \text{ADOM}$.
  - $I, \alpha[x \leftarrow d] \models P(\ldots, v, \ldots)$ iff $I, \alpha \models rem_x(P(\ldots, v, \ldots))$, since $rem_x(P(\ldots, v, \ldots)) = P(\ldots, v, \ldots)$ (recall x does not occur in $P(\ldots, v, \ldots)$).
  - The inductive cases are straightforward by definition of $rem_x(\cdot)$.

QED

Equivalence of FOL and FOL$_{\text{ADOM}}$

Proof of the Main Theorem.
The above lemma shows: $\mathcal{I}, \alpha \models \psi$ iff $\mathcal{I}, \alpha \models \overline{\psi}$. If ψ is closed then we can drop the initial assignment, thus getting

$$\mathcal{I} \models \psi \quad \textit{iff} \quad \mathcal{I} \models \overline{\psi}.$$

Moreover, since the only quantification appearing in $\overline{\psi}$ is bounded to elements in the active domain we get :

$$\mathcal{I} \models \overline{\psi} \quad \textit{iff} \quad \overline{\mathcal{I}} \models \overline{\psi}.$$

Hence, the thesis holds.

QED