Appendix C: First order semantics of OIL

In this section we will give a formal specification and semantics for OIL as well as for the common inference problems (class consistency and inferred subclass relations) performed with respect to an OIL ontology. We will only consider the **definitions** part of the ontology and we will ignore fields such as **documentation** that have no semantic significance.

The semantics of OIL rely on a translation into the $\mathcal{SHIQ}$ description logic. $\mathcal{SHIQ}$ has a highly expressive concept language that is able to fully capture the OIL core language, and we will define a satisfiability preserving translation $\sigma(\cdot)$ that maps OIL ontologies into $\mathcal{SHIQ}$ terminologies. This has the added benefit that an existing $\mathcal{SHIQ}$ reasoner implemented in the FaCT system can be used to reason with OIL ontologies.

The translation is quite straightforward and follows directly from the informal specification given in Section 3.1. An OIL ontology $\mathcal{O}$ consists of a list $d_1, \ldots, d_n$, where each d_i is either a class definition or a slot definition. This list of definitions is translated into a $\mathcal{SHIQ}$ terminology $\mathcal{T}$ (a set of axioms) as follows:

$$\sigma(d_1, \ldots, d_n) = \bigcup_{i=1,\ldots,n} \sigma(d_i)$$

A class definition is either a pair $\langle \mathsf{CN}, D\rangle$ or a triple $\langle \mathsf{CN}, P, D\rangle$, where CN is a class name, D is a class description and P is either **primitive** or **defined**; $\langle \mathsf{CN}, D\rangle$ is equivalent to $\langle \mathsf{CN}, \textbf{primitive}, D\rangle$. A class definition $\langle \mathsf{CN}, \textbf{primitive}, D\rangle$ is written $\mathsf{CN} \sqsubseteq D$ (it states that CN is a subclass of the class described by D) and a class definition $\langle \mathsf{CN}, \textbf{defined}, D\rangle$ is written $\mathsf{CN} \doteq D$ (it states that CN is equivalent to the class described by D).

A class description D consists of an optional **subclass-of** component, itself a list of one or more **class-expression**s $C_1, \ldots, C_n$, followed by a list of zero or more **slot-constraints** $A_1, \ldots, A_m$. We will write such a class description as

$$[C_1, \ldots, C_n, A_1, \ldots, A_m].$$

A **class-expression** is either a class name CN, a **slot-constraint**, a conjunction of class expressions, written $C_1 \sqcap \ldots \sqcap C_n$, a disjunction of class expressions, written $C_1 \sqcup \ldots \sqcup C_n$ or a negated class expression, written $\neg C$. A **slot-constraint** consists of a slot name SN followed by one or more constraints that apply to the slot, written $\mathsf{SN}[a_1, \ldots, a_n]$. Each constraint can be either:

- A **value** constraint with a list of one or more class-expressions, written $\exists C_1, \ldots, C_n$.
- A **value-type** constraint with a list of one or more class-expressions, written $\forall C_1, \ldots, C_n$.
- A **max-cardinality** constraint with a non-negative integer n followed (optionally) by a class expression C, written $\leqslant n, C$ ($\leqslant n, \top$ if the class expression is omitted).
- A **min-cardinality** constraint with a non-negative integer n followed (optionally) by a class expression C, written $\geqslant n, C$ ($\geqslant n, \top$ if the class expression is omitted).
- A **cardinality** constraint with a non-negative integer n followed (optionally) by a class expression C, written $=n, C$ ($=n, \top$ if the class expression is omitted).

In order to maintain the decidability of the language, cardinality constraints can only be applied to *simple* slots. A simple slot is one that is neither transitive nor has any transitive subslots. However, as the transitivity of a slot can be inferred (e.g., from the fact that the inverse of the slot is a transitive

$$
\begin{aligned}
\sigma(\mathsf{CN} \sqsubseteq D) &= \{\sigma(\mathsf{CN}) \sqsubseteq \sigma(D)\}\\
\sigma(\mathsf{CN} \doteq D) &= \{\sigma(\mathsf{CN}) \sqsubseteq \sigma(D), \sigma(D) \sqsubseteq \sigma(\mathsf{CN})\}\\
\sigma([C_1, \ldots, C_n, A_1, \ldots, A_m]) &= \top \sqcap \sigma(C_1) \sqcap \ldots \sqcap \sigma(C_n) \sqcap \sigma(A_1) \sqcap \ldots \sqcap \sigma(A_m)\\
\sigma(\mathsf{CN}) &= \mathsf{CN}\\
\sigma(\top) &= \top\\
\sigma(C_1 \sqcap \ldots \sqcap C_n) &= \sigma(C_1) \sqcap \ldots \sqcap \sigma(C_n)\\
\sigma(C_1 \sqcup \ldots \sqcup C_n) &= \sigma(C_1) \sqcup \ldots \sqcup \sigma(C_n)\\
\sigma(\neg C) &= \neg\sigma(C)\\
\sigma(\mathsf{SN}[a_1, \ldots, a_n]) &= \sigma(\mathsf{SN}(a_1)) \sqcap \ldots \sqcap \sigma(\mathsf{SN}(a_n))\\
\sigma(\mathsf{SN}(\exists C_1, \ldots, C_n)) &= \exists \mathsf{SN}.\sigma(C_1) \sqcap \ldots \sqcap \exists \mathsf{SN}.\sigma(C_n)\\
\sigma(\mathsf{SN}(\forall C_1, \ldots, C_n)) &= \forall \mathsf{SN}.\sigma(C_1) \sqcap \ldots \sqcap \forall \mathsf{SN}.\sigma(C_n)\\
\sigma(\mathsf{SN}(\leqslant n, C)) &= \leqslant n\mathsf{SN}.\sigma(C)\\
\sigma(\mathsf{SN}(\geqslant n, C)) &= \geqslant n\mathsf{SN}.\sigma(C)\\
\sigma(\mathsf{SN}(= n, C)) &= \leqslant n\mathsf{SN}.\sigma(C) \sqcap \geqslant n\mathsf{SN}.\sigma(C)
\end{aligned}
$$

Figure C-1: Translation of OIL class definitions into $\mathcal{SHIQ}$

slot), simple slot is defined in terms of the translation into $\mathcal{SHIQ}$: a slot SN in an ontology $\mathcal{O}$ is a simple slot iff $\sigma(\mathsf{SN})$ is a simple role in the $\mathcal{SHIQ}$ terminology $\sigma(\mathcal{O})$.

We can now define how the function $\sigma(\cdot)$ maps an OIL class definition into a set of $\mathcal{SHIQ}$ axioms. The definition is given in Figure C-1, where CN is a class name (or a $\mathcal{SHIQ}$ concept name), SN is a slot name (or $\mathcal{SHIQ}$ role name), D is a class description, C (possibly subscripted) is a class expression, A (possibly subscripted) is a slot constraint, a_i is a constraint (on a slot) and n is a non-negative integer.

A slot definition is a pair $\langle \mathsf{SN}, D\rangle$, where SN is a slot name and D is a slot description. A slot description D consists of an optional **subslot-of** component, itself a list of one or more slot names $\mathsf{RN}_1, \ldots, \mathsf{RN}_n$, followed by a list of zero or more global slot constraints (e.g., **inverse**) $S_1, \ldots, S_m$. We will write such a slot definition as:

$$\mathsf{SN}[\mathsf{RN}_1, \ldots, \mathsf{RN}_n, S_1, \ldots, S_m]$$

Each global constraint S_i on SN can be either:

- A **domain** constraint with a list of one or more class-expressions, written $\downarrow [C_1, \ldots, C_n]$.
- A **range** constraint with a list of one or more class-expressions, written $\uparrow [C_1, \ldots, C_n]$.
- An **inverse** constraint with a slot name RN, written ${}^{-}\mathsf{RN}$.
- A **properties** constraint with a list of one or more properties, written $[P_1, \ldots, P_n]$. Valid properties are **transitive**, written $+$ and **symmetrical**, written $\leftrightarrow$.

We can now define how the function $\sigma(\cdot)$ maps an OIL slot definition into a set of $\mathcal{SHIQ}$ axioms. The definition is given in Figure C-2, where RN and SN are slot names (or $\mathcal{SHIQ}$ role names), C_i is a class expression, S_i is a global slot constraint and P_i is a property.

The meaning of a $\mathcal{SHIQ}$ terminology, and of the common inference problems, is given in terms of a Tarski style model theoretic semantics using *interpretations*. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

$$\begin{aligned}
\sigma(\mathsf{SN}[\mathsf{RN}_1,\ldots,\mathsf{RN}_n,S_1,\ldots,S_m]) &= \sigma(\mathsf{SN}[\mathsf{RN}_1,\ldots,\mathsf{RN}_n] \cup \sigma(\mathsf{SN}[S_1,\ldots,S_m]) \\
\sigma(\mathsf{SN}[\mathsf{RN}_1,\ldots,\mathsf{RN}_n] &= \textstyle\bigcup_{i=1,\ldots,n} \sigma(\mathsf{SN} \sqsubseteq \mathsf{RN}_i) \\
\sigma(\mathsf{SN}[S_1,\ldots,S_m]) &= \textstyle\bigcup_{i=1,\ldots,m} \sigma(\mathsf{SN}(S_i)) \\
\sigma(\mathsf{SN} \sqsubseteq \mathsf{RN}) &= \{\mathsf{SN} \sqsubseteq \mathsf{RN}\} \\
\sigma(\mathsf{SN}(\downarrow [C_1,\ldots,C_n])) &= \textstyle\bigcup_{i=1,\ldots,n} \{\exists \mathsf{SN}.\top \sqsubseteq \sigma(C_i)\} \\
\sigma(\mathsf{SN}(\uparrow [C_1,\ldots,C_n])) &= \textstyle\bigcup_{i=1,\ldots,n} \{\top \sqsubseteq \forall \mathsf{SN}.\sigma(C_i)\} \\
\sigma(\mathsf{SN}(^-\mathsf{RN})) &= \{\mathsf{SN}^- \sqsubseteq \mathsf{RN}, \mathsf{RN} \sqsubseteq \mathsf{SN}^-\} \\
\sigma(\mathsf{SN}([P_1,\ldots,P_n])) &= \textstyle\bigcup_{i=1,\ldots,n} \{\sigma(\mathsf{SN}(P_i))\} \\
\sigma(\mathsf{SN}(+)) &= \{\mathsf{SN} \in \mathbf{S}_+\} \\
\sigma(\mathsf{SN}(\leftrightarrow)) &= \{\mathsf{SN}^- \sqsubseteq \mathsf{SN}, \mathsf{SN} \sqsubseteq \mathsf{SN}^-\}
\end{aligned}$$

Figure C-2: Translation of OIL slot definitions into $\mathcal{SHIQ}$

consists of a set $\Delta^{\mathcal{I}}$, called the *domain* of $\mathcal{I}$, and a *valuation* $\cdot^{\mathcal{I}}$ which maps every concept to a subset of $\Delta^{\mathcal{I}}$ and every role to a subset of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ such that, for all concepts C, D, roles R, S, and non-negative integers n, the following equations are satisfied, where $\sharp M$ denotes the cardinality of a set M:

$$\begin{aligned}
(R^-)^{\mathcal{I}} &= \{\langle x,y\rangle \mid \langle y,x\rangle \in R^{\mathcal{I}}\} && \text{(inverse roles)} \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} && \text{(conjunction)} \\
(C \sqcup D)^{\mathcal{I}} &= C^{\mathcal{I}} \cup D^{\mathcal{I}} && \text{(disjunction)} \\
(\neg C)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} && \text{(negation)} \\
(\exists R.C)^{\mathcal{I}} &= \{x \mid \exists y.\langle x,y\rangle \in R^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\} && \text{(value constraint)} \\
(\forall R.C)^{\mathcal{I}} &= \{x \mid \forall y.\langle x,y\rangle \in R^{\mathcal{I}} \text{ implies } y \in C^{\mathcal{I}}\} && \text{(value-type constraint)} \\
(\geqslant nR.C)^{\mathcal{I}} &= \{x \mid \sharp\{y.\langle x,y\rangle \in R^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\} \geqslant n\} && \text{(minimum cardinality)} \\
(\leqslant nR.C)^{\mathcal{I}} &= \{x \mid \sharp\{y.\langle x,y\rangle \in R^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\} \leqslant n\} && \text{(maximum cardinality)}
\end{aligned}$$

In order to avoid considering roles such as R^{--} (i.e., the inverse of an inverse) we will define a function Inv such that $\mathit{Inv}(R)$ is R^- and $\mathit{Inv}(R^-)$ is R. A role R is *directly subsumed* by a role S w.r.t. a terminology $\mathcal{T}$ iff either $\{R \sqsubseteq S\} \subseteq \mathcal{T}$ or $\{\mathit{Inv}(R) \sqsubseteq \mathit{Inv}(S)\} \subseteq \mathcal{T}$. A role R is *subsumed* by a role S w.r.t. $\mathcal{T}$ (written $\mathcal{T} \models R \sqsubseteq S$) iff R is directly subsumed by a S or there is a role S' such that R is directly subsumed by a S' and $\mathcal{T} \models S' \sqsubseteq S$. A role R is *equivalent* to a role S w.r.t. $\mathcal{T}$ (written $\mathcal{T} \models R \doteq S$) iff $\mathcal{T} \models R \sqsubseteq S$ and $\mathcal{T} \models S \sqsubseteq R$. A role R is transitive in $\mathcal{T}$ iff $\{S \in \mathbf{S}_+\} \subseteq \mathcal{T}$ for some role S such that $R \doteq S$ or $\mathit{Inv}(R) \doteq S$ (this defines $\mathbf{S}_+$, the set of transitive role names). A role R is a *simple* role in $\mathcal{T}$ iff there is no role S such that S is transitive in $\mathcal{T}$ and $\mathcal{T} \models S \sqsubseteq R$.

An interpretation $\mathcal{I}$ *satisfies* a $\mathcal{SHIQ}$ terminology $\mathcal{T}$ iff for every axiom $R \sqsubseteq S$ in $\mathcal{T}$, $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$, for every axiom $C \sqsubseteq D$ in $\mathcal{T}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ and for every transitive role S in $\mathcal{T}$, $S^{\mathcal{I}} = (S^{\mathcal{I}})^+$. Such an interpretation is called a *model* of $\mathcal{T}$ (written $\mathcal{I} \models \mathcal{T}$).

A concept C is satisfiable with respect to a $\mathcal{SHIQ}$ terminology $\mathcal{T}$ (written $\mathcal{T} \models C \neq \bot$) iff there a model $\mathcal{I}$ of $\mathcal{T}$ with $C^{\mathcal{I}} \neq \emptyset$. A concept C is subsumed by a concept D w.r.t. $\mathcal{T}$ (written $\mathcal{T} \models C \sqsubseteq D$) $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds for each model $\mathcal{I}$ of $\mathcal{T}$.

An OIL ontology $\mathcal{O}$ is called *consistent* iff $\sigma(\mathcal{O}) \models \top \neq \bot$. A class CN in an ontology $\mathcal{O}$ is called *consistent* iff $\sigma(\mathcal{O}) \models \sigma(\mathsf{CN}) \neq \bot$. A class CN is a *subclass* of a class DN in an ontology $\mathcal{O}$ iff $\sigma(\mathcal{O}) \models \sigma(\mathsf{CN}) \sqsubseteq \sigma(\mathsf{DN})$.