ORM2: Formalisation and Encoding in OWL2

Enrico Franconi[1], Alessandro Mosca[1], and Dmitry Solomakhin[2]

Free University of Bozen-Bolzano, Italy
lastname@inf.unibz.it,
Dmitry.Solomakhin@stud-inf.unibz.it

Abstract. This paper introduces $\mathsf{ORM2}^{\mathsf{plus}}$ – a new linear syntax and complete semantics expressed in first order logic of ORM2 – which can be shown correctly embedding the original proposal. A provably correct encoding of the core fragment $\mathsf{ORM2}^{\mathsf{zero}}$ in the $\mathcal{ALCQI}$ description logic (a fragment of OWL2 with qualified cardinality restrictions and inverse roles) is presented. Complexity of reasoning on ORM2 conceptual schemas, and the ExpTime-membership of reasoning on $\mathsf{ORM2}^{\mathsf{zero}}$, are also shown. On the basis of these results, a systematic critique of alternative approaches to the formalisation of ORM2 in (description) logics published so far is provided. A prototype has been implemented providing a backend for the automated support of implicit constraints deduction, schema consistency checks, and user-defined constraints entailment, for $\mathsf{ORM2}^{\mathsf{zero}}$ conceptual schemas along with its translation into $\mathcal{ALCQI}$ knowledge bases.

1 Introduction

The NIAM language, ancestor ORM, has been equipped with a first order logic (FOL) semantics in 1989 [1]. Since then, despite the remarkable evolution in terms of expressivity that ORM2 has experienced, much less attention has been paid to its formal foundations. This paper can be considered as an attempt to fill this gap. In particular, the issue of providing a logic formalism, equipped with sound and complete reasoning services, that captures the expressiveness of ORM2 is addressed, and a 'practical' fragment of ORM2 is introduced. The first contribution of the paper is thus the introduction of a completely new linear syntax and FOL semantics for a generalisation of ORM2, called $\mathsf{ORM2}^{\mathsf{plus}}$, allowing for the specification of join paths over an arbitrary number of relations. The syntax can be used to express the full set of ORM2 graphical symbols introduced in [2]. The new semantics has been proved to be equivalent with the original FOL semantics of NIAM, up to the differences in the expressivity of the two languages. The second contribution of the paper is driven by a practical objective: We identified a 'core' fragment of ORM2, called $\mathsf{ORM2}^{\mathsf{zero}}$, that can be translated in a sound a complete way into the ExpTime-complete description logic (DL) $\mathcal{ALCQI}$ [3]. Finally, we implemented a first prototype, built on top of available DL reasoners which provides an automated support for **schema consistency** checks, **implicit constraints deduction**, and user-defined **constraints entailment**. The rest of the paper is organised as follows: Section 2

P. Herrero et al. (Eds.): OTM 2012 Workshops, LNCS 7567, pp. 368–378, 2012.

introduces syntax and semantics of $\mathsf{ORM2}^{\mathsf{plus}}$. $\mathsf{ORM2}^{\mathsf{zero}}$ is presented in Section 3 together with its encoding in $\mathcal{ALCQI}$. A critique of alternative approaches to the formalisation of ORM2 in (description) logics published so far is provided in Section 4. Section 5 gives an overview of the implemented prototype. See [4] for a complete account of the work behind this paper, including full proofs.

2 ORM2 from a Formal Perspective

The modelling activity in ORM2 is supported by several tools that provide user friendly graphical interfaces to build complex conceptual schemas in real world application domains. Among these, the Natural Object-Role Modelling Architect (NORMA) tool is a plug-in to MS Visual Studio providing the most complete support for the ORM2 notation. This tool performs syntactic check on the graphical notation, returns warnings for not-admitted combinations of basic elements and constraints, and drives the modelling activity according to the ORM2 *Conceptual Schema Design Procedure* [2]. Nonetheless, the ability to avoid the definition of syntactically correct schemas that resolve to be semantically inconsistent is currently left to expertise and skill of the modeller itself. It is known that, due to design mistakes, a syntactically correct schema may (i) not admit any instantiation without the violation of some constraints, or (ii) admit only a partial instantiation. *Schema consistency*, *consistency of an object type*, and the fact that some constraints is already present in a schema as *implicit consequences*, are typical properties of a schema that, once checked, significantly improve its quality. The automated verification of these properties depends on the possibility to perform reasoning on the schema by means of a logic representation of it. With this goal in mind, this section presents the $\mathsf{ORM2}^{\mathsf{plus}}$ language. For each construct φ in the syntax, its corresponding semantics expressed in FOL is introduced in tables 1 and 2. The signature $\mathcal{S}$ of the linear syntax is made of the following symbols: (i) A set $\mathcal{E}$ of *entity types*; (ii) a set $\mathcal{V}$ of *value types* (a set of *object types* $\mathcal{O} = \mathcal{E} \cup \mathcal{V}$); (iii) a set $\mathcal{R}$ of *relations*; (iv) a set $\mathcal{A}$ of *roles*; (v) a set $\mathcal{D}$ of *domains*, and a set Λ of pairwise disjoint sets of values. Then, a binary relation $\varrho \subseteq \mathcal{R} \times \mathcal{A}$ linking role to relation symbols is also in $\mathcal{S}$: $R.a$ is the atomic elements of the syntax (given a relation symbol R, $\varrho_R = \{R.a | R.a \in \varrho\}$ is the set of *localised roles* w.r.t R, and $|\varrho_R| = arity(R)$). Finally, for each relation symbol R, a bijection $\tau_R \colon \varrho_R \to [1..|\varrho_R|]$, mapping role components and argument positions in a relation, is defined.

Given the signature $\mathcal{S}$, an $\mathsf{ORM2}^{\mathsf{plus}}$ ***conceptual schema*** Σ over $\mathcal{S}$ includes a finite combination of the constructs in tables 1 and 2. The list of constraints graphically introduced in [2] can now be linearised using a specialisation of the new syntax, where: (1) TYPE is for role **typing**; (2) FREQ for **frequency occurrence**; (3) MAND for **mandatory participation**; (4) $\mathsf{R\text{-}SET_H}$ for the family of **set-comparison** constraints; (5) $\mathsf{O\text{-}SET_H}$ for the family of **subtyping** constraints; (6) O-CARD and R-CARD for **object** and **role cardinality**; and (7) OBJ for **objectification**; (8) RING for **ring** constraints; and (9) V-VAL for **object value** constraints. Below, an example on how the introduced syntax can be used to encode a fragment of the schema in fig. 1.

Table 1. Linear Syntax (■) and FOL Semantics (□) table

■ $\mathsf{TYPE} \subseteq \varrho \times (\mathcal{E} \cup \mathcal{V})$

□ If $\mathsf{TYPE}(R.a, O) \in \Sigma$ then $\forall x_1 \ldots x_{\tau(R.a)} \ldots x_n . R(x_1, \ldots, x_{\tau(R.a)}, \ldots, x_n) \rightarrow O(x_{\tau(R.a)})$

■ $\mathsf{FREQ} \subseteq \wp(\varrho) \times (\wp(\varrho) \times \wp(\varrho)) \times (\mathbb{N} \times (\mathbb{N} \cup \{\infty\}))$

□ If $\mathsf{FREQ}(\{R^1.a_{11}, \ldots, R^1.a_{1n}, \ldots, R^k.a_{k1}, \ldots, R^k.a_{km}\}, \bowtie_{\mathbf{R}}, (\mathsf{min}, \mathsf{max})) \in \Sigma$ then

$$\forall \overline{y}[\exists \overline{x}^1 \ldots \overline{x}^k (\bigwedge_{j=1}^{k} R^j(\overline{x}^j) \wedge \bigwedge_{\mathbf{i1}=1}^{n} (x^1_{\tau(R^1.a_{1\mathbf{i1}})} = y_{1\mathbf{i1}}) \wedge \cdots \wedge \bigwedge_{\mathbf{ik}=1}^{m} (x^1_{\tau(R^1.a_{1\mathbf{ik}})} = y_{1\mathbf{ik}}) \wedge \bigwedge_{\bowtie_{\mathbf{R}}} (x^{r^+}_{\tau(R^{r^+}.a_{r^+\mathbf{v}_r})} = x^{r^-}_{\tau(R^{r^-}.a_{r^-\mathbf{w}_r})}))] \rightarrow$$

$$\exists^{\geq \mathsf{min}; \leq \mathsf{max}} \overline{y}[\exists \overline{x}^1 \ldots \overline{x}^k (\bigwedge_{j=1}^{k} R^j(\overline{x}^j) \wedge \bigwedge_{\mathbf{i1}=1}^{n} (x^1_{\tau(R^1.a_{1\mathbf{i1}})} = y_{1\mathbf{i1}}) \wedge \cdots \wedge \bigwedge_{\mathbf{ik}=1}^{m} (x^1_{\tau(R^1.a_{1\mathbf{ik}})} = y_{1\mathbf{ik}}) \wedge (\bigwedge_{\bowtie_{\mathbf{R}}} x^{r^+}_{\tau(R^{r^+}.a_{r^+\mathbf{v}_r})} = x^{r^-}_{\tau(R^{r^-}.a_{r^-\mathbf{w}_r})}))]$$

where:

(1) $\bowtie_{\mathbf{R}} = \{\ldots, \langle R^i.a_{i\mathbf{v}} = R^j.a_{j\mathbf{w}} \rangle, \ldots\}$, with $i \neq j$ and $1 \leq i, j \leq k$, is the finite set of join (role) pairs (given k relations $\mathbf{R}$, $| \bowtie_{\mathbf{R}} | = k - 1$)

(2) $R^i.a_{i\mathbf{x}} \in \varrho_{R^i}$ for any $R^i \in \mathcal{R}$, and

(3) the equalities in $\bigwedge_{\bowtie_{\mathbf{R}}}$ are specified according to $\bowtie_{\mathbf{R}}$

(e.g. given $\overline{x}^1, \overline{x}^2, \overline{x}^3$ s.t. $R^1(\overline{x}^1), R^2(\overline{x}^2), R^3(\overline{x}^3)$, if $\bowtie_{\mathbf{R}} = \{(R^1.a, R^2.b), (R^2.c, R^3.d)\}$ then $\bigwedge_{\bowtie_{\mathbf{R}}} =_{\mathsf{def}} (x^1_{\tau(R^1.a)} = x^2_{\tau(R^2.b)}) \wedge (x^2_{\tau(R^2.c)} = x^3_{\tau(R^3.d)})$)

■ $\mathsf{MAND} \subseteq \wp(\varrho) \times (\mathcal{E} \cup \mathcal{V})$

□ If $\mathsf{MAND}(\{R^1.a_{11}, \ldots, R^1.a_{1n}, \ldots, R^k.a_{k1}, \ldots, R^k.a_{km}\}, O) \in \Sigma$ then

$$\forall y[A(y) \rightarrow \bigvee_{i=1}^{n} \exists \overline{z}^i (R^1(\overline{z}^i) \wedge (z^i_{\tau(R^1.a_{1i})} = y) \vee \cdots \vee \bigvee_{j=1}^{m} \exists \overline{z}^j . R^k(\overline{z}^j) \wedge (z^j_{\tau(R^k.a_{kj})} = y))]$$

Table 1. (*continued*)

■ $\mathsf{R\text{-}SET_H} \subseteq (\wp(\varrho) \times (\wp(\varrho) \times \wp(\varrho))) \times (\wp(\varrho) \times (\wp(\varrho) \times \wp(\varrho))) \times (\mu\colon \varrho \to \varrho)$ where $\mathsf{H} = \{\mathsf{Sub}, \mathsf{Exc}\}$

□ • If $\mathsf{R\text{-}SET_{Sub}}((\{R^1.a_{11}, \ldots, R^1.a_{1n}, \ldots, R^k.a_{k1}, \ldots, R^k.a_{km}\}, \bowtie_{\mathbf{R}}), (\{S^1.b_{11}, \ldots, S^1.b_{1v}, \ldots, S^q.b_{q1}, \ldots, S^q.b_{qw}\}, \bowtie_{\mathbf{S}}, \mu) \in \Sigma$ then

$$\forall \overline{y}[\exists \overline{x}^1 \ldots \overline{x}^k (\bigwedge_{j=1}^{k} R^j(\overline{x}^j) \wedge \bigwedge_{\mathbf{i1}=1}^{n} (x^1_{\tau(R^1.a_{1\mathbf{i1}})} = y_{1\mathbf{i1}}) \wedge \cdots \wedge \bigwedge_{\mathbf{ik}=1}^{m} (x^1_{\tau(R^1.a_{1\mathbf{ik}})} = y_{1\mathbf{ik}}) \wedge \bigwedge_{\bowtie_{\mathbf{R}}} (x^{r^+}_{\tau(R^{r^+}.a_{r^+\mathbf{v}_r})} = x^{r^-}_{\tau(R^{r^-}.a_{r^-\mathbf{w}_r})})) \rightarrow$$

$$\exists \overline{z}^1 \ldots \overline{z}^q (\bigwedge_{i=1}^{q} S^i(\overline{z}^i) \wedge \bigwedge_{\mathbf{i1}=1}^{n} (z^{f_{\mu(1\mathbf{i1})}}_{\tau(\mu(R^1.a_{1\mathbf{i1}}))} = y_{1\mathbf{i1}}) \wedge \cdots \wedge \bigwedge_{\mathbf{ik}=1}^{m} (z^{f_{\mu(1\mathbf{ik})}}_{\tau(\mu(R^1.a_{1\mathbf{ik}}))} = y_{1\mathbf{ik}}) \wedge \bigwedge_{\bowtie_{\mathbf{S}}} (z^{s^+}_{\tau(S^{s^+}.b_{s^+\mathbf{v}_s})} = z^{s^-}_{\tau(S^{s^-}.b_{s^-\mathbf{w}_s})}))]$$

• If $\mathsf{R\text{-}SET_{Exc}}((\{R^1.a_{11}, \ldots, R^1.a_{1n}, \ldots, R^k.a_{k1}, \ldots, R^k.a_{km}\}, \bowtie_{\mathbf{R}}), (\{S^1.b_{11}, \ldots, S^1.b_{1v}, \ldots, S^q.b_{q1}, \ldots, S^q.b_{qw}\}, \bowtie_{\mathbf{S}}, \mu) \in \Sigma$ then

$$\forall \overline{y}[\exists \overline{x}^1 \ldots \overline{x}^k (\bigwedge_{j=1}^{k} R^j(\overline{x}^j) \wedge \bigwedge_{\mathbf{i1}=1}^{n} (x^1_{\tau(R^1.a_{1\mathbf{i1}})} = y_{1\mathbf{i1}}) \wedge \cdots \wedge \bigwedge_{\mathbf{ik}=1}^{m} (x^1_{\tau(R^1.a_{1\mathbf{ik}})} = y_{1\mathbf{ik}}) \wedge \bigwedge_{\bowtie_{\mathbf{R}}} (x^{r^+}_{\tau(R^{r^+}.a_{r^+\mathbf{v}_r})} = x^{r^-}_{\tau(R^{r^-}.a_{r^-\mathbf{w}_r})})) \rightarrow$$

$$\neg(\exists \overline{z}^1 \ldots \overline{z}^q (\bigwedge_{i=1}^{q} S^i(\overline{z}^i) \wedge \bigwedge_{\mathbf{i1}=1}^{n} (z^{f_{\mu(1\mathbf{i1})}}_{\tau(\mu(R^1.a_{1\mathbf{i1}}))} = y_{1\mathbf{i1}}) \wedge \cdots \wedge \bigwedge_{\mathbf{ik}=1}^{m} (z^{f_{\mu(1\mathbf{ik})}}_{\tau(\mu(R^1.a_{1\mathbf{ik}}))} = y_{1\mathbf{ik}}) \wedge \bigwedge_{\bowtie_{\mathbf{S}}} (z^{s^+}_{\tau(S^{s^+}.b_{s^+\mathbf{v}_s})} = z^{s^-}_{\tau(S^{s^-}.b_{s^-\mathbf{w}_s})})))]$$

where:

(1) given $\varrho^{\mathsf{C_A}} = \{R^1.a_{11}, \ldots, R^1.a_{1n}, \ldots, R^k.a_{k1}, \ldots, R^k.a_{km}\}$, and $\varrho^{\mathsf{C_B}} = \{S^1.b_{11}, \ldots, S^1.b_{1v}, \ldots, S^q.b_{q1}, \ldots, S^q.b_{qw}\}$, μ is a partial bijection s.t. for any $\langle \varrho^{\mathsf{C_A}}, \varrho^{\mathsf{C_B}}, \mu \rangle \in \mathsf{R\text{-}SET_H}$, we have $\varrho^{C_A} = \{R.a | \mu(R.a) \in \varrho^{C_B}\}$, and

(2) $f_{\mu(x\mathbf{y})} = z$ iff $\mu(R^x.a_{x\mathbf{y}}) \in \varrho_{S^z}$

$$
\begin{array}{rl}
\mathsf{ENTITYTYPES} : & \{\mathsf{PhoneCall}, \mathsf{MobileCall}, \mathsf{PhonePoint}, \mathsf{Cell}, \mathsf{Landline}, \mathsf{HomePoint}\} \\
\mathsf{VALUETYPES} : & \{\mathsf{PhoneCall\_Id}, \mathsf{PhonePoint\_\sharp}\} \\
\mathsf{RELATIONS} : & \{\mathsf{HasOriginFrom}, \mathsf{HasDestinationTo}, \mathsf{HasMOriginFrom}, \mathsf{HasPhoneCall\_Id}, \\
& \mathsf{HasPhonePoint\_\sharp}\} \\
\mathsf{TYPE}(\mathsf{HasOriginFrom.1}, \mathsf{PhoneCall}) & \mathsf{MAND}(\{\mathsf{HasOriginFrom.1}\}, \mathsf{PhoneCall}) \\
& \mathsf{FREQ}(\{\mathsf{HasOriginFrom.1}\}, \{\}, (1,1)) \\
\mathsf{TYPE}(\mathsf{HasOriginFrom.2}, \mathsf{PhonePoint}) & \\
\mathsf{TYPE}(\mathsf{HasDestinationTo.1}, \mathsf{PhoneCall}) & \mathsf{MAND}(\{\mathsf{HasDestinationTo.1}\}, \mathsf{PhoneCall}) \\
\mathsf{TYPE}(\mathsf{HasDestinationTo.2}, \mathsf{PhonePoint}) & \\
\mathsf{TYPE}(\mathsf{HasMOriginFrom.1}, \mathsf{MobileCall}) & \\
\mathsf{TYPE}(\mathsf{HasMOriginFrom.2}, \mathsf{Cell}) & \\
\mathsf{TYPE}(\mathsf{HasPhoneCall\_Id.1}, \mathsf{PhoneCall}) & \\
: \mathsf{MAND}(\{\mathsf{HasPhoneCall\_Id.1}\}, \mathsf{PhoneCall}) & \\
& \mathsf{FREQ}(\{\mathsf{HasPhoneCall\_Id.1}\}, \{\}, (1,1)) \\
\mathsf{TYPE}(\mathsf{HasPhoneCall\_Id.2}, \mathsf{PhoneCall_{Id}}) & \mathsf{FREQ}(\{\mathsf{HasPhoneCall\_Id.2}\}, \{\}, (1,1)) \\
\mathsf{TYPE}(\mathsf{HasPhonePoint\_\sharp.1}, \mathsf{PhonePoint}) & \mathsf{MAND}(\{\mathsf{HasPhonePoint\_\sharp.1}\}, \mathsf{PhonePoint}) \\
& \mathsf{FREQ}(\{\mathsf{HasPhonePoint\_\sharp.1}\}, \{\}, (1,1)) \\
\mathsf{TYPE}(\mathsf{HasPhonePoint\_\sharp.2}, \mathsf{PhonePoint\_\sharp}) & \mathsf{FREQ}(\{\mathsf{HasPhonePoint\_\sharp.2}\}, \{\}, (1,1)) \\
\mathsf{O\text{-}SET_{Tot}}(\{\mathsf{Landline}, \mathsf{Cell}\}, \mathsf{PhonePoint}) & \\
\mathsf{O\text{-}SET_{Ex}}(\{\mathsf{Landline}, \mathsf{Cell}\}, \mathsf{PhonePoint}) & \\
\mathsf{O\text{-}SET_{Ex}}(\{\mathsf{HomePoint}, \mathsf{Landline}\}, \mathsf{PhonePoint}) &
\end{array}
$$

$$
\begin{array}{l}
\mathsf{R\text{-}SET_{Sub}}(\{\mathsf{HasMOriginFrom.1}, \mathsf{HasMOriginFrom.2}\}, \{\}), (\{\mathsf{HasOriginFrom.1}, \mathsf{HasOriginFrom.2}\}, \{\}), \\
\qquad \{(\mathsf{HasMOriginFrom.1}, \mathsf{HasOriginFrom.1}), (\mathsf{HasMOriginFrom.2}, \mathsf{HasOriginFrom.2})\}
\end{array}
$$

$\mathsf{ORM2^{plus}}$ covers $\mathsf{ORM2}$[1]; in particular, *external* and *internal* forms are represented by means of different specialisations of the same constructs; the FREQ construct can now be applied to arbitrary role sequences no matter about the arity of involved relations, and the same holds for the $\mathsf{R\text{-}SET_H}$ constructs. Additional sequences of role pairs (see $\bowtie_{\mathbf{R}}$, and $\bowtie_{\mathbf{S}}$) are among the arguments of both FREQ and $\mathsf{R\text{-}SET_H}$, and used to specify the roles where the joins must be computed. Uniqueness constraints are viewed as frequency occurrence constraints with a fixed range $(1,1)$, and several other constraints can now be derived as shown in table 2. The 'strict' version of the subtyping relation, that is assumed as primitive in [2], is seen here as a derived constraint: Given a non-strict semantics for the subtyping relation, the strict one can be represented by a combination of partition, cardinality constraint, and the introduction of a a new fresh object type symbol ('equality' can also be expressed using a similar pattern, where the cardinality of the new introduced symbol is zero).

Note that the FOL semantics is based on a signature that perfectly matches the one of the linear syntax (n unary entity type predicates, m value type predicates, and l domain symbol predicates are introduced). Nonetheless, the following set of *background axioms* is needed in order to force the interpretation of the symbols in the FOL knowledge bases (KBs) to be correct w.r.t. the intended semantics of the corresponding $\mathsf{ORM2}$ symbols:

$$\forall x. E_i(x) \to \neg(D_1(x) \vee \cdots \vee D_l(x)),\ \text{for } 1 \leq i \leq n \tag{1}$$

$$\forall x. V_i(x) \to D_j(x),\ \text{for } 1 \leq i \leq m,\ \text{and some } j \tag{2}$$

$$\forall x. D_i(x) \leftrightarrow (x = d_1 \vee x = d_2 \vee \dots),\ \text{for all } d_i \in \Lambda_{D_i} \tag{3}$$

$$\forall x_1, \dots, x_n, z_1, \dots, z_n. \mathrm{ID}(\overline{\mathbf{x}}) = \mathrm{ID}(\overline{\mathbf{z}}) \leftrightarrow \overline{\mathbf{x}} = \overline{\mathbf{s}},\ \text{for } n = 1, \dots, n_{max} \tag{4}$$

[1] We consider here the standard ORM2 as defined in the book [2]; it is not hard to characterise in FOL, e.g., the additional constraints introduced by [5].

Table 2. Linear Syntax (■) and FOL Semantics (□) table (contd)

■	$\mathsf{O\text{-}SET_H} \subseteq \wp(\mathcal{E} \cup \mathcal{V}) \times \mathcal{E} \cup \mathcal{V}$ where $\mathsf{H} = \{\mathsf{Isa}, \mathsf{Tot}, \mathsf{Ex}\}$
□	• If $\mathsf{O\text{-}SET_{Isa}}(\{O_1, \dots, O_n\}, O) \in \Sigma$ then $\forall y. O_i(y) \rightarrow O(y)$ for all $i = 1, \dots, n$ • If $\mathsf{O\text{-}SET_{Tot}}(\{O_1, \dots, O_n\}, O) \in \Sigma$ then $\begin{cases} \forall y. O_i(y) \rightarrow O(y) \\ \forall y. O(y) \rightarrow O_1(y) \vee \dots \vee O_n(y), \text{ for all } i = 1, \dots, n \end{cases}$ • If $\mathsf{O\text{-}SET_{Ex}}(\{O_1, \dots, O_n\}, O) \in \Sigma$ then $\begin{cases} \forall y. O_1(y) \rightarrow O(y) \wedge \neg O_2(y) \wedge \dots \wedge \neg O_n(y) \\ \forall y. O_2(y) \rightarrow O(y) \wedge \neg O_3(y) \wedge \dots \wedge \neg O_{n-1}(y) \\ \dots \\ \forall y. O_{n-1}(y) \rightarrow O(y) \wedge \neg O_1(y) \\ \forall y. O_n(y) \rightarrow O(y) \end{cases}$
■	$\mathsf{O\text{-}CARD} \subseteq (\mathcal{E} \cup \mathcal{V}) \times (\mathbb{N} \times (\mathbb{N} \cup \{\infty\}))$
□	If $\mathsf{O\text{-}CARD}(O) = (\mathsf{min}, \mathsf{max})) \in \Sigma$ then $\exists^{\geq \mathsf{min}} y. O(y) \wedge \exists^{\leq \mathsf{max}} y. O(y)$
■	$\mathsf{R\text{-}CARD} \subseteq \wp(\varrho) \times (\mathbb{N} \times (\mathbb{N} \cup \{\infty\}))$
□	If $\mathsf{R\text{-}CARD}(R.a) = (\mathsf{min}, \mathsf{max})) \in \Sigma$ then $\exists^{\geq \mathsf{min}} x_{\tau(R.a)}. R(x_1 \dots x_{\tau(R.a)} \dots x_n) \wedge \exists^{\leq \mathsf{max}} x_{\tau(R.a)}. R(x_1 \dots x_{\tau(R.a)} \dots x_n)$
■	$\mathsf{OBJ} \subseteq \mathcal{R} \times (\mathcal{E} \cup \mathcal{V})$
□	If $\mathsf{OBJ}(R, O) \in \Sigma$ then $\forall x. O(x) \leftrightarrow \exists \overline{y}. R(\overline{y}) \wedge \mathrm{ID}^{\lvert \varrho_R \rvert}(\overline{y}) = x$
■	$\mathsf{RING_J} \subseteq \wp(\varrho \times \varrho)$ where $\mathsf{J} = \{\mathsf{Irr}, \mathsf{Asym}, \mathsf{Trans}, \mathsf{Intr}, \mathsf{Antisym}, \mathsf{Acyclic}, \mathsf{Sym}, \mathsf{Ref}, \dots\}$
□	E.g. If $\mathsf{RING_{Irr}}(R.a, R.b) \in \Sigma$ then $\forall x_{\tau(R.a)}, x_{\tau(R.b)}. R(x_{\tau(R.a)}, x_{\tau(R.b)}) \rightarrow \neg R(x_{\tau(R.b)}, x_{\tau(R.a)})$
■	$\mathsf{V\text{-}VAL}: \mathcal{V} \rightarrow \wp(\Lambda_D)$ for some $\Lambda_D \in \Lambda$ (where $\Lambda_{(\cdot)}$ associates an extension to each domain symbol)
□	If $\mathsf{V\text{-}VAL}(V) = \{d_1, \dots, d_n\} \in \Sigma$ then $\forall x. V(x) \rightarrow (x = d_1) \vee \dots \vee (x = d_n)$

Table 3. Derived constraints

Uniqueness:	$\mathsf{FREQ}(\{R^1.a_{11}, \dots, R^1.a_{1n}, \dots, R^k.a_{k1}, \dots, R^k.a_{km}\}, \bowtie_{\mathbf{R}}, \langle 1, 1 \rangle)$
Role value:	$\mathsf{TYPE}(R.a, V^*)$ where V^* is a new fresh value type symbol $\mathsf{V\text{-}VAL}(V^*) \subseteq \{v_1^D, \dots, v_n^D\}$
Equality:	$\mathsf{R\text{-}SET_{Sub}}((\{R^1.a_{11}, \dots, R^1.a_{1n}, \dots, R^k.a_{k1}, \dots, R^k.a_{km}\}, \bowtie_{\mathbf{R}}),$ $(\{S^1.b_{11}, \dots, S^1.b_{1v}, \dots, S^q.b_{q1}, \dots, S^q.b_{qw}\}, \bowtie_{\mathbf{S}}), \mu)$ $\mathsf{R\text{-}SET_{Sub}}((\{S^1.b_{11}, \dots, S^1.b_{1v}, \dots, S^q.b_{q1}, \dots, S^q.b_{qw}\}, \bowtie_{\mathbf{S}})$ $(\{R^1.a_{11}, \dots, R^1.a_{1n}, \dots, R^k.a_{k1}, \dots, R^k.a_{km}\}, \bowtie_{\mathbf{R}}), \mu^-)$
Exclusive-Or:	$\mathsf{MAND}(\{R^1.a_{11}, \dots, R^1.a_{1n}, \dots, R^k.a_{k1}, \dots, R^k.a_{km}\}, O)$ $\mathsf{R\text{-}SET_{Exc}}((\{R^1.a_{11}, \dots, R^1.a_{1n}\}), (\{R^2.a_{21}, \dots, R^2.a_{2n}\}), \mu_1)$ $\mathsf{R\text{-}SET_{Exc}}((\{R^1.a_{11}, \dots, R^1.a_{1n}\}), (\{R^3.a_{31}, \dots, R^3.a_{3n}\}), \mu_2), \cdots,$ $\mathsf{R\text{-}SET_{Exc}}((\{R^{k-1}.a_{k-11}, \dots, R^{k-1}.a_{k-1n}\}), (\{R^k.a_{k1}, \dots, R^k.a_{kn}\}), \mu_k)$
Partition:	$\mathsf{O\text{-}SET_{Tot}}(\{O_1, \dots, O_n\}, O)$ $\mathsf{O\text{-}SET_{Ex}}(\{O_1, \dots, O_n\}, O)$
Strict Subtyping:	$\mathsf{O\text{-}SET_{Tot}}(\{O_1, O^*\}, O)$ $\mathsf{O\text{-}SET_{Ex}}(\{O_1, O^*\}, O)$ $\mathsf{O\text{-}CARD}(O^*) = (1, \mathsf{inf})$ where O^* is a new fresh object type symbol

where, axiom (1) forces the interpretation of each entity type to be disjoint from the interpretation of the domain symbols; axiom (2) says that objects in the interpretation of a value type must be also in the interpretation of a specific domain symbol; axiom (3) forces the interpretation of a domain symbols to be among the set of values predefined by $\Lambda_{(\cdot)}$, while axiom (4) captures the injective nature of each ID function and the fact that tuples of different length will never agree on the same identifier (where $n_{max} = \mathsf{max}\{|\varrho_R| | R \in \mathcal{R}\}$). A FO interpretation is a **model** for an ORM2$^{\mathsf{plus}}$ schema if it satisfies the background axioms and the corresponding FOL KB built as described in tables 1 and 2. We can prove that, when the schema is restricted to a NIAM schema, the models of the corresponding ORM2$^{\mathsf{plus}}$ schema are the same as the FO models of the NIAM schema as specified in [1].

3 Encoding in $\mathcal{ALCQI}$

With the main aim of relying on available tools to reason in an effective way on ORM2 schemas, we present here the encoding in the logic $\mathcal{ALCQI}$ for which tableaux-based reasoning algorithms with a tractable computational complexity have been developed [3]. $\mathcal{ALCQI}$ corresponds to the basic DL $\mathcal{ALC}$ equipped with *qualified cardinality restrictions* and *inverse roles*, and it is a fragment of OWL2. The difficulty implied by the absence of n-ary relations has been overcome by means of *reification*: For each relation R with arity $n \geq 2$, a new atomic concept A_R and n functional roles $\tau(R.a_1), \ldots, \tau(R.a_n)$ one for each component of R are introduced. Due to the tree-model property of $\mathcal{ALCQI}$, the reification process provides a sound and complete translation w.r.t. concept satisfiability, such that each instance of the new introduced concept is a representative of one and only one tuple of R. Given as such, the tree-model property guarantees the *correctness* of the $\mathcal{ALCQI}$ encoding w.r.t. the reasoning services over ORM2. Besides reification, we can prove that the expressiveness of $\mathcal{ALCQI}$ does not allow to fully capture ORM2: $\mathcal{ALCQI}$ does not admit neither arbitrary set-comparison assertions on relations, nor external uniqueness or uniqueness involving more than one role, nor arbitrary frequency occurrence constraints. The analysis of these restrictions thus led to identification of the fragment ORM2$^{\mathsf{zero}}$ that is maximal with respect to the expressiveness of $\mathcal{ALCQI}$, and still expressive enough to capture the most frequent usage patterns of the modelling community in ORM2 [2]. Let ORM2$^{\mathsf{zero}}$ = $\{\mathsf{TYPE}, \mathsf{FREQ}^-, \mathsf{MAND}, \mathsf{R\text{-}SET}^-, \mathsf{O\text{-}SET}_{\mathsf{Isa}}, \mathsf{O\text{-}SET}_{\mathsf{Tot}}, \mathsf{O\text{-}SET}_{\mathsf{Ex}}, \mathsf{OBJ}\}$ be the fragment of ORM2 where: (i) FREQ^- can only be applied to single roles, and (ii) $\mathsf{R\text{-}SET}^-$ applies either to relations of the same arity or to two single roles. The encoding of the semantics of ORM2$^{\mathsf{zero}}$ shown in table 3 is based on the $\mathcal{S}^{\mathcal{ALCQI}}$ signature made of: (i) A set $E_1, E_2, \ldots, E_n$ of concepts for *entity types*; (ii) a set $V_1, V_2, \ldots, V_m$ of concepts for *value types*; (iii) a set $A_{R_1}, A_{R_2}, \ldots, A_{R_k}$ of concepts for objectified n-ary *relations*; (iv) a set $D_1, D_2, \ldots, D_l$ of concepts for *domain symbols*; (v) $1, 2, \ldots, n_{max} + 1$ roles.

The correctness of the introduced encoding is guaranteed by the following theorem (whose complete proof is available online at [4]):

Theorem 1. *Let Σ^{zero} be an* $\mathsf{ORM2}^{\mathsf{zero}}$ *conceptual schema and $\Sigma^{\mathcal{ALCQI}}$ the $\mathcal{ALCQI}$ KB constructed as described above. Then an object type O is consistent in Σ^{zero} if and only if the corresponding concept O is satisfiable w.r.t. $\Sigma^{\mathcal{ALCQI}}$.*

As a matter of fact, all the reasoning tasks for a conceptual schema can be reduced to object type consistency. Let us conclude this section with some observation about the complexity of reasoning on ORM2 conceptual schemas. Undecidability of the ORM2 object type consistency problem can be proved by showing that arbitrary combinations of subset constraints between n-ary relations and uniqueness constraints over single roles are allowed [6] (*a fortiori*, $\mathsf{ORM2}^{\mathsf{plus}}$ is also undecidable). As for $\mathsf{ORM2}^{\mathsf{zero}}$, one can conclude that object type consistency is ExpTime-complete: The upper bound is established by reducing the $\mathsf{ORM2}^{\mathsf{zero}}$ problem to concept satisfiability w.r.t. $\mathcal{ALCQI}$ KBs (which is known to be ExpTime-hard) [7], the lower bound by reducing concept satisfiability w.r.t. $\mathcal{ALC}$ KBs (which is known to be ExpTime-complete) to object consistency w.r.t. $\mathsf{ORM2}^{\mathsf{zero}}$ schemas [8]. Therefore, we obtain the following result:

Theorem 2. *Reasoning over* $\mathsf{ORM2}^{\mathsf{zero}}$ *schemas is* ExpTime-*complete.*

Table 4. $\mathcal{ALCQI}$ encoding

Background domain axioms:	$E_i \sqsubseteq \neg(D_1 \sqcup \cdots \sqcup D_l)$ for $i \in \{1, \ldots, n\}$ $V_i \sqsubseteq D_j$ for $i \in \{1, \ldots, m\}$, and some j with $1 \leq j \leq l$ $D_i \sqsubseteq \sqcap_{j=i+1}^{l} \neg D_j$ for $i \in \{1, \ldots, l\}$ $\top \sqsubseteq A_{\top_1} \sqcup \cdots \sqcup A_{\top_{n_{max}}}$ $\top \sqsubseteq (\leq 1 i.\top)$ for $i \in \{1, \ldots, n_{max}\}$ $\forall i.\bot \sqsubseteq \forall i+1.\bot$ for $i \in \{1, \ldots, n_{max}\}$ $A_{\top_n} \equiv \exists 1.A_{\top_1} \sqcap \cdots \sqcap \exists n.A_{\top_1} \sqcap \forall n+1.\bot$ for $n \in \{2, \ldots, n_{max}\}$ $A_R \sqsubseteq A_{\top_n}$ for each atomic relation R of arity n $A \sqsubseteq A_{\top_1}$ for each atomic concept A
$\mathsf{TYPE}(R.a, O)$	$\exists\tau(R.a)^-.A_R \sqsubseteq O$
$\mathsf{FREQ}^-(R.a, \langle \mathsf{min}, \mathsf{max} \rangle)$	$\exists\tau(R.a)^-.A_R \sqsubseteq\ \geq \mathsf{min}\ \tau(R.a)^-.A_R \sqcap \leq \mathsf{max}\ \tau(R.a)^-.A_R$
$\mathsf{MAND}(\{R^1.a_1, \ldots, R^1.a_n, \ldots, R^k.a_1, \ldots, R^k.a_m\}, O)$	$O \sqsubseteq \exists\tau(R^1.a_1)^-.A_{R^1} \sqcup \cdots \sqcup \exists\tau(R^1.a_n)^-.A_{R^1} \sqcup \cdots \sqcup \exists\tau(R^k.a_1)^-.A_{R^k} \sqcup \cdots \sqcup \exists\tau(R^k.a_m)^-.A_{R^k}$
$^{(A)}$ $\mathsf{R\text{-}SET}^-_{\mathsf{Sub}}(A, B)$ $^{(A)}$ $\mathsf{R\text{-}SET}^-_{\mathsf{Exc}}(A, B)$	$A_R \sqsubseteq A_S$ $\quad$ $^{(A)}\ A = \{R.a_1, \ldots, R.a_n\}, B = \{S.b_1, \ldots, S.b_n\}$ $A_R \sqsubseteq A_{\top_n} \sqcap \neg A_S$
$^{(B)}$ $\mathsf{R\text{-}SET}^-_{\mathsf{Sub}}(A, B)$ $^{(B)}$ $\mathsf{R\text{-}SET}^-_{\mathsf{Exc}}(A, B)$	$\exists\tau(R.a_i)^-.A_R \sqsubseteq \exists\tau(S.b_j)^-.A_S$ $\quad$ $^{(B)}\ A = \{R.a_i\}, B = \{S.b_j\}$ $\exists\tau(R.a_i)^-.A_R \sqsubseteq A_{\top_n} \sqcap \neg\exists\tau(S.b_j).A_S$
$\mathsf{O\text{-}SET}_{\mathsf{Isa}}(\{O_1, \ldots, O_n\}, O)$	$O_1 \sqcup \cdots \sqcup O_n \sqsubseteq O$
$\mathsf{O\text{-}SET}_{\mathsf{Tot}}(\{O_1, \ldots, O_n\}, O)$	$O \sqsubseteq O_1 \sqcup \cdots \sqcup O_n$
$\mathsf{O\text{-}SET}_{\mathsf{Ex}}(\{O_1, \ldots, O_n\}, O)$	$O_1 \sqcup \cdots \sqcup O_n \sqsubseteq O$ and $O_i \sqsubseteq \sqcap_{j=i+1}^{n} \neg O_j$ for each $i = 1, \ldots, n$
$\mathsf{OBJ}(R, O)$	$O \equiv A_R$

4 Related Works

In the last few years, several papers addressed the issue of encoding ORM2 conceptual schema into DL KBs [9,10,11,12]. Among those proposals, [9] goes through the encoding with a formal perspective. [9] introduces an encoding of a fragment of ORM2 into the logic $\mathcal{DLR}_{ifd}$, an extension of $\mathcal{DLR}$ with *identification assertions* on concepts, and *functional dependencies* assertions on relations [6]. Except for the presence of uniqueness constraints spanning over arbitrary sequence of n roles of the same relation, and external uniqueness over 2 roles, that are represented in the paper by means of suitable identification assertions, $\mathsf{ORM2}^{\mathsf{zero}}$ and the fragment identified in [9] agree on the same expressive power. Unfortunately, the proposal is somehow sloppy, and it is wrong w.r.t. the semantics specified in [1] (e.g., the semantics of 'objectification' and frequency occurrence constraints is wrong).

As regards to [10], already [9] claimed its limits. It should be also noticed that [11,12] suffer from the same formal inconsistencies and limitations of [10]. In particular, [10] is sloppy with respect to the underlying DL formalism: distinct extensions of the adopted logic (e.g. $\mathcal{DLR}$ plus $\mathcal{DLR}$-*Lite*) and distinct DL languages (e.g. $\mathcal{DLR}$, plus $\mathcal{DLR}$-*Lite*, plus $\mathcal{SROIQ}$, plus 'role composition' operator) are mixed together. No semantics nor complexity results are is provided for these combinations.

In [13], a list of nine 'unsatisfiability constraint patterns' is introduced with the aim of supporting the automatic detection of unsatisfiable schema elements. The patterns discussed in the paper represent a subset of all the possible sources of inconsistency that can occur in a conceptual schema, and it is therefore severely *incomplete*.

Similar to [13], a paper focused on the encoding of ORM2 in OWL has been recently published [14]. The paper introduces a set of informal 'rules' devoted to the mapping of a subset of the ORM2 constructs into OWL. Unfortunately, the paper is wrong in several respects: (i) the OWL EquivalentTO, instead of the SubClassOf, is erroneously introduced; (ii) optionality of uniqueness constraints is lost. In general, the paper covers a fragment that is smaller than $\mathsf{ORM2}^{\mathsf{zero}}$, and the proposed mapping mostly remains formally unjustified.

Finally, in [15] the *converse* (incomplete) encoding of OWL-DL into ORM2 is presented. In the paper, the authors claim that 'universal restrictions' of the form $A \sqsubseteq \forall R.B$ cannot be translated in ORM2 in a way that preserves the semantics of the original constructs. But, this is not the case: A viable translation into ER has been introduced in [8], where covering (i.e. total subtyping) and disjointness (i.e. exclusive subtyping) between relationships are used, and a second one into UML can be found in [7], making use of reification of roles. Both translations can be straightforwardly rephrased into ORM2.

A detailed account of all the problems in the related work can be found in [4].

5 Automated Reasoning Support Tool

With the main goal of providing automated reasoning services facilitating the conceptual modelling activity, a prototype of $\mathsf{ORM2}^{\mathsf{zero}}$ modelling support tool has been implemented for NORMA. The prototype takes an ORM2 schema produced by NORMA as input, and encodes it into the linear syntax using an XSLT script. By relying on existing OWL2 reasoners (e.g. HermiT, FaCT++), the tool provides the following functionality:

- **Implicit constraints deduction**. Derived implicit $\mathsf{ORM2}^{\mathsf{zero}}$ constraints, including inconsistent object types and fact types, are displayed in distinct pop-up windows. The computation is complete, but only cognitively relevant constraints are visualised, e.g. redundant transitive links are not visualised.
- **Translation into OWL2 ontology**. In order to facilitate web-data exchange and to make conceptual schemas readily accessible to automated processes, the prototype features a translator from $\mathsf{ORM2}^{\mathsf{zero}}$ schema into OWL2 ontology, which can then be saved in various formats.

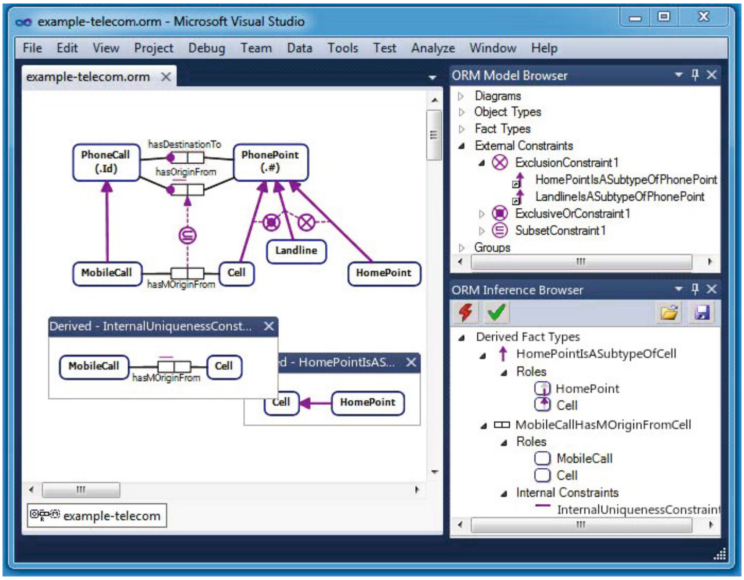

Fig. 1. Graphical interface of the prototype

Let us now illustrate the essential functionality of the prototype using the example introduced on fig. 1. Among the key constraints, the schema involves the uniqueness constraint imposed on the origin of a phone call as well as the hierarchical constraints describing the nature of possible phone points. Therefore, with a single click we can obtain the following relevant deductions for the given conceptual schema: $\mathsf{FREQ}(\{\mathsf{hasMOriginFrom.1}\}, \{\}, (1, 1))$ and $\mathsf{O\text{-}SET}_{\mathsf{Isa}}(\{\mathsf{HomePoint}\}, \mathsf{Cell})$, i.e. it is true that any home point is also a cell point, and each mobile call may have an origin from at most one cell point. In order to understand why this is true, consider the following. The class of home points is a sub class of all the phone points, and it is disjoint from the class of landline points. Since any phone point is either a cell point or a landline point, then any home point should necessarily be a cell point. The hasMOriginFrom binary relation is included in the hasOriginFrom binary relation. Since each call participates exactly once as first argument to the hasOriginFrom, if we take a generic sub class of calls, such as the class of mobile calls, and a sub relationship of the hasOriginFrom relation, such as hasMOriginFrom, then we can conclude that necessarily each mobile call participates at most once as first argument to the hasMOriginFrom relation. The full list of the inferred constraints is displayed in the ORM2 *Inference Browser* window while selected deductions are illustrated by relevant fragments of the inferred schema in pop-up windows over the initial schema.

References

1. Halpin, T.: A Logical Analysis of Information Systems: Static Aspects of the Data-oriented Perspective. PhD thesis, Department of Computer Science, University of Queensland (1989)
2. Halpin, T., Morgan, T.: Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design, 2nd edn. Morgan Kaufmann (2001)
3. 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, New York (2003)
4. Franconi, E., Mosca, A.: The formalisation of ORM2 and its encoding in OWL2. Technical Report KRDB12-2, KRDB Research Centre, Free University of Bozen-Bolzano (2012), `http://www.inf.unibz.it/krdb/pub/TR/KRDB12-2.pdf`
5. Curland, M., Halpin, T.A., Stirewalt, K.: A Role Calculus for ORM. In: Meersman, R., Herrero, P., Dillon, T. (eds.) OTM 2009 Workshops. LNCS, vol. 5872, pp. 692–703. Springer, Heidelberg (2009)
6. Calvanese, D., De Giacomo, G., Lenzerini, M.: Identification constraints and functional dependencies in description logics. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence, pp. 155–160 (2001)
7. Berardi, D., Cali, A., Calvanese, D., Giacomo, G.D.: Reasoning on UML class diagrams. Artificial Intelligence 168 (2003)
8. Artale, A., Calvanese, D., Kontchakov, R., Ryzhikov, V., Zakharyaschev, M.: Reasoning over Extended ER Models. In: Parent, C., Schewe, K.-D., Storey, V.C., Thalheim, B. (eds.) ER 2007. LNCS, vol. 4801, pp. 277–292. Springer, Heidelberg (2007)
9. Keet, M.: Mapping the Object-Role Modeling language ORM2 into description logic language $\mathcal{DLR}_{ifd}$. Technical Report KRDB07-2, KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano (2007)
10. Jarrar, M.: Towards Automated Reasoning on ORM Schemes. Mapping ORM into the $\mathcal{DLR}_{ifd}$ Description Logic. In: Parent, C., Schewe, K.-D., Storey, V.C., Thalheim, B. (eds.) ER 2007. LNCS, vol. 4801, pp. 181–197. Springer, Heidelberg (2007)
11. Jarrar, M.: Mapping ORM into the $\mathcal{SHOIN}$/OWL description logic. In: Proc. of the International Workshop on Object-Role Modeling (ORM 2007), pp. 729–741. Springer (2007)
12. Hodrob, R., Jarrar, M.: ORM to OWL2 DL mapping. In: International Conference on Intelligent Semantic Web: Applications and Services. ACM (2010)
13. Jarrar, M., Heymans, S.: Unsatisfiability reasoning in orm conceptual schemes. In: Proc. of the IFIP-2.6 International Conference on Semantics of a Networked World, pp. 517–534. Springer (2006)
14. Wagih, H.M., ElZanfaly, D.S., Kouta, M.M.: Mapping Object Role Modeling 2 schemes to OWL2 ontologies. In: Proc. of the 3rd IEEE International Conference on Computer Research and Development (ICCRD), pp. 126–132. IEEE Press (2011)
15. Bach, D.B., Meersman, R., Spyns, P., Trog, D.: Mapping OWL-DL into ORM/RIDL. In: Proceedings of the 2007 OTM Confederated International Conference, OTM 2007, pp. 742–751 (2007)