A Formal Investigation of Mapping Language for Terminological Knowledge

Keywords: 1) semantic web 2) knowledge representation 3) ontologies

Abstract

The need for being able to talk about mappings between different ontologies has been recognized as a result of the fact that different ontologies may partially overlap or even represent the same domain from different points of view. Unlike for the case of ontology languages, work on mapping languages has not yet reached a state where a common understanding of the basic principles exists. In this paper we propose a formal comparison of existing mapping language by translating them into distributed first order logic. We analyze underlying assumptions and differences in the interpretation of mappings.

1 Motivation

The benefits of using ontologies as explicit models of the conceptualization underlying information sources has widely been recognized. Meanwhile, a number of logical languages for representing and reasoning about ontologies have been proposed and there are even langauge standards now that guarantee a stability and homogeneity on the language level. At the same time, the need for being able to talk about mappings between different ontologies has been recognized as a result of the fact that different ontologies may partially overlap or even represent the same domain from different points of view [Bouquet *et al.*, 2004]. As a result a number of proposal have been made for extending ontology languages with notions of mappings between different models. Unlike for the case of ontology languages, work on mapping languages has not yet reached a state where a common understanding of the basic principles exists. As a consequence, existing proposals show major differences concerning almost all possible aspects which makes it difficult to compare approaches and make a decision about the usefulness of a particular approach in a given situation.

The purpose of this work is to support a better understanding of the commonalities and differences of existing proposals for mapping languages. We restrict our attention to logic-based approaches that have been defined as extensions to existing formalisms for representing Terminological Knowledge. In particular, we chose approaches that extend description logics (DL) with notions of mappings between different T-boxes. The rationale for this choice is the fact that DLs are a widely agreed standard for describing terminological knowledge. In particular, because DLs have gained a lot of attention as a standardized way of representing ontologies on the Semantic Web [Horrocks *et al.*, 2003].

Approach and Contributions

We encode the different mapping languages in an extended version of distributed first-order logic (DFOL), a logical framework for representing distributed knowledge systems [Ghidini and Serafini, 2000]. DFOL consists of two components: a family of first order theories and a set of axioms describing the relations between these theories. As most proposals for mapping languages are based on a subset of first-order logic for describing local models and mappings with a particular semantics for the connections between models, these mapping language can be expressed in distributed first order logic in the following way:

- restrictions on the use of first order sentences for describing domain models
- the form of axioms that can be used for describing relations between domain models
- axioms describing the assumptions that are encoded in the specific semantics of mappings

Encoding the different mapping approaches in first-order logic in this way has several advantages with respect to an analysis and comparison of existing work. In particular it allows us to do a formal analysis and comparison of different approaches in a uniform logical framework. In the course of the investigations, we make the following contributions to the state of the art in distributed knowledge representation and reasoning:

- we show how DFOL formalism can be used to model relations between heterogeneous domains (Proposition 1)
- we encode existing mapping approaches in a common framework making them more comparable
- we make hidden assumptions explicit in terms of distributed first order logic axioms
- we provide first results on the relative expressiveness of the approaches and identify shared fragments

The paper is structured as follows. In section 2 we introduce distributed first order logic as a general model for describing distributed knowledge systems. We explain the intuition of the logic and introduce its syntax and semantics. In section 3 we describe how the different mapping approaches mentioned above can be encoded in distributed first order language. Here we will focus on the representation of mappings and the encoding of hidden assumptions. In section 4 we compare the different approaches based on their encoding in DFOL and discuss the issues such as relative expressiveness and compatibility of the different approaches and conclude with a summary of our findings and open questions.

2 Distributed First-Order Logic

This section introduces distributed first order logic as a basis for modeling distributed knowledge bases. More details about the language including a sound and complete calculus can be found in [Ghidini and Serafini, 2005].

Let $\{L_i\}_{i\in I}$ (in the following $\{L_i\}$) be a family of first order languages with equality defined over a non empty set I of indexes. Each language L_i is the language used by the i-th knowledge base (ontology). The signature of L_i is extended with a new set of symbols used to denote objects which are related with other objects in different ontologies. For each variable, and each index $j \in I$ with $j \neq i$ we have two new symbols $x^{\rightarrow j}$ and $x^{j\rightarrow}$, called *arrow variables*. Terms and formulas of L_i, also called i-*terms* and i-*formulas* and are defined in the usual way. Quantification on arrow variables is not permitted. The notation $\phi(\mathbf{x})$ is used to denote the formula ϕ and the fact that the free variables of ϕ are $\mathbf{x} = \{x_1, \ldots, x_n\}$. In order to distinguish occurrences of terms and formulas in different languages we label them with their index. The expression $i : \phi$ denotes the formula ϕ of the i-th knowledge base.

The semantics of DFOL is an extension of Local Models Semantics defined in [Ghidini and Giunchiglia, 2001]. Local models are defined in terms of first order models. To capture the fact that certain predicates are completely known by the i-th sub-system we select a sub-language of L_i containing the equality predicate, denoted as L_i^c we call the *complete fragment* of L_i. *Complete terms* and *complete formulas* are terms and formula of L_i^c and vice versa.

Definition 1 (Set of local Models)**.** A *set of local models* of L_i are a set of first order interpretations of L_i, on a domain $\mathbf{dom}_i$, which agree on the interpretation of L_i^c, the complete fragment of L_i.

As noted in [Franconi and Tessaris, 2004] there is a foundational difference between approaches that use epistemic states and approaches that use a classical model theoretic semantics. The two approaches differ as long as there is more than one model m. Using the notion of complete sublanguage L_c, however, we can force the set of local models is either a singleton or the empty set by enforcing that $L^c = L$. Under this assumption the two ways of defining the semantics of submodels are equivalent. Using this assumption, we are therefore able to simulate both kinds of semantics in DFOL.

Two or more models can carry information about the same portion of the world. In this case we say that they *semantically overlap*. Overlapping is unrelated to the fact that the same constant appears in two languages, as from the local semantics we have that the interpretation of a constant c in L_i is independent from the interpretation of the very same constant in L_j, with $i \neq j$. Overlapping is also unrelated to the intersection between the interpretation domains of two or more contexts. Namely if $\mathbf{dom}_1 \cap \mathbf{dom}_2 \neq \emptyset$ it does not mean that L_1 and L_j overlap. Instead, DFOL explicitly represent semantic overlapping via a domain relation.

Definition 2 (Domain relation)**.** A *domain relation* form $\mathbf{dom}_i$ and $\mathbf{dom}_j$ is a binary relations $r_{ij} \subseteq \mathbf{dom}_i \times \mathbf{dom}_j$.

Domain relation from i to j represents the capability of the j-th sub-system to represent in its domain the domain of the i-th subsystem. A pair $\langle d, d' \rangle$ being in r_{ij} means that, from the point of view of j, d in $\mathbf{dom}_i$ is the representation of d' in $\mathbf{dom}_j$. We use the functional notation $r_{ij}(d)$ to denote the set $\{d' \in \mathbf{dom}_j \,|\, \langle d, d' \rangle \in r_{ij}\}$. The domain relation r_{ij} formalizes j's subjective point of view on the relation between $\mathbf{dom}_i$ and $\mathbf{dom}_j$ and not an absolute objective point of view. Or in other words $r_{ij} \neq r_{ji}$ because of the non-symmetrical nature of mappings. Therefore $\langle d, d' \rangle \in r_{ij}$ must not be read as if d and d' were the same object in a domain shared by i and j. This facts would indeed be formalized by some observer which is external (above, meta) to both i and j. Using the notion of domain relation we can define the notion of a model for a set of local models.

Definition 3 (DFOL Model)**.** A DFOL *model*, $\mathcal{M}$ is a pair $\langle \{\mathcal{M}_i\}, \{r_{ij}\} \rangle$ where, for each $i \neq j \in I$: $\mathcal{M}_i$ is a set of local models for L_i, and r_{ij} is a domain relation from $\mathbf{dom}_i$ to $\mathbf{dom}_j$.

We extend the classical notion of assignment (e.g., the one given for first order logic) to deal with arrow variables using domain relations. In particular, an assignment a, provides for each system i, an interpretation for all the variable, and for *some* (by not necessarily all) arrow variables as the domain relations might be such that there is no consistent way to assign arrow variables. For instance if $a_i(x) = d$ and $r_{ij}(d) = \emptyset$, then a_j cannot assign anything to $x^{i\rightarrow}$.

Definition 4 (Assignment)**.** Let $\mathcal{M} = \langle \{\mathcal{M}_i\}, \{r_{ij}\} \rangle$ be a model for $\{L_i\}$. An *assignment* a is a family $\{a_i\}$ of partial functions from the set of variables and arrow variables to $\mathbf{dom}_i$, such that:

1. $a_i(x) \in \mathbf{dom}_i$;
2. $a_i(x^{j\rightarrow}) \in r_{ji}(a_j(x))$;
3. $a_j(x) \in r_{ij}(a_i(x^{\rightarrow j}))$;

An assignment a is *admissible for a formula* $i : \phi$ if a_i assigns all the arrow variables occurring in ϕ. Furthermore, a is admissible for a set of formulas Γ if it is admissible for any $j : \phi \in \Gamma$. An assignment a is *strictly admissible* for a set of formulas Γ if it is admissible for Γ and assigns only the arrow variables that occurs in Γ.

Using the notion of an admissible assignment given above, satisfiability in distributed first order logic is defined as follows:

Definition 5 (Satisfiability)**.** Let $\mathcal{M} = \langle \{\mathcal{M}_i\}, \{r_{ij}\} \rangle$ be a model for $\{L_i\}$, $m \in \mathcal{M}_i$, and a an assignment. An i-formula ϕ is *satisfied* by m, w.r.t, a, in symbols $m \models_{\text{D}} \phi[a]$ if

$a)$	$\mathcal{M} \models i:P(x^{\rightarrow j}) \rightarrow j:Q(x)$	iff	For all $d \in \|P\|_i$ and <u>for all $d' \in r_{ij}(d)$</u>, $d' \in \|Q\|_j$
$b)$	$\mathcal{M} \models i:P(x) \rightarrow j:Q(x^{i\rightarrow})$	iff	For all $d \in \|P\|_i$ <u>there is a $d' \in r_{ij}(d)$</u>, s.t., $d' \in \|Q\|_j$
$c)$	$\mathcal{M} \models j:Q(x^{i\rightarrow}) \rightarrow i:P(x)$	iff	For all $d \in \|Q\|_j$ and <u>for all d'</u> <u>with $d \in r_{ij}(d')$</u>, $d' \in \|P\|_i$
$d)$	$\mathcal{M} \models j:Q(x) \rightarrow i:P(x^{\rightarrow j})$	iff	For all $d \in \|Q\|_i$ <u>there is a d'</u> <u>with $d \in r_{ij}(d')$</u>, s.t., $d' \in \|P\|_i$

Figure 1: Implicit Quantification of Arrow Variables in Interpretation Constraints

1. a is admissible for $i:\phi$ and
2. $m \models \phi[a_i]$, according to the definition of satisfiability for first order logic.

$\mathcal{M} \models \Gamma[a]$ if for all $i:\phi \in \Gamma$ and $m \in \mathcal{M}_i$, $m \models_{\mathrm{D}} \phi[a_i]$[1].

Mappings between different knowledge bases are formalized in DFOL by a new form of constraints that involves more than one knowledge bases. These formulas that will be the basis for describing different mapping approaches are called interpretation constraints and defined as follows:

Definition 6 (Interpretation constraint)**.** An *interpretation constraint from* $i_1, \ldots, i_n$ *to* i with $i_k \neq i$ for $1 \leq k \leq n$ is an expression of the form

$$i_1:\phi_1, \ldots, i_n:\phi_n \rightarrow i:\phi \quad (1)$$

The interpretation constraint (1) can be consider as an axiom that restrict the set of possible DFOL models to those which satisfies it. Therefore we need to define when a DFOL model satisfies an interpretation constraint.

Definition 7 (Satisfiability of interpretation constraints)**.** A model $\mathcal{M}$ *satisfies* the interpretation constraint (1), in symbols $\mathcal{M} \models i_1:\phi_1, \ldots, i_n:\phi_n \rightarrow i:\phi$ if for any assignment a strictly admissible for $\{i_1:\phi_1, \ldots, i_n:\phi_n\}$, if $\mathcal{M} \models i_k:\phi_k[a]$ for $1 \leq k \leq n$, then a can be extended to an assignment a' admissible for $i:\phi$ and such that $\mathcal{M} \models i:\phi[a']$.

Notice that, depending on the fact that an arrow variable $x^{\rightarrow}$ occurs on the left or on the right side of the constraint, $x^{\rightarrow}$ has a universal or an existential reading. Figure 1 summarizes the different possible readings that will reoccur later. Notationally for any predicate P, $\|P\|_i = \bigcap_{m \in \mathcal{M}_i} m(P)$, where $m(P)$ is the interpretation of P in m.

By means of interpretation constraints on equality, we can formalize possible relations between heterogeneous domains.

$$\begin{aligned}
\mathsf{F}_{ij} &= \{i:x^{\rightarrow j} = y^{\rightarrow j} \rightarrow j:x = y\} \\
\mathsf{INV}_{ij} &= \left\{ \begin{array}{l} i:x = y^{j\rightarrow} \rightarrow j:x^{i\rightarrow} = y \\ j:x = y^{i\rightarrow} \rightarrow i:x^{j\rightarrow} = y \end{array} \right\} \\
\mathsf{OD}_{ij} &= \mathsf{F}_{ij} \cup \mathsf{F}_{ji} \cup \mathsf{INV}_{ij} \\
\mathsf{ED}_{ij} &= \mathsf{OD}_{ij} \cup \{i:x = x \rightarrow j:x^{i\rightarrow} = x^{i\rightarrow}\} \\
\mathsf{ID}_{ij} &= \mathsf{ED}_{ij} \cup \mathsf{ED}_{ji} \\
\mathsf{RD}_{ij} &= \left\{ \begin{array}{l} i:x = c \rightarrow j:x^{i\rightarrow} = c \\ j:x = c \rightarrow i:x^{j\rightarrow} = c \end{array} \middle| \, c \in L_i \cap L_j \right\} \\
\mathsf{IP}_{ij} &= i:\bot \rightarrow j:\bot
\end{aligned}$$

[1]Since it will be clear from the context, in the rest we will use the classical satisfiability symbol $\models$ instead of $\models_{\mathrm{D}}$ and we will write $m \models \phi[a]$ to mean that an i-formula ϕ is satisfied by m. In writing $m \models \phi[a]$ we always mean that of a is admissible for $i:\phi$ (in addition to the fact that m classically satisfies ϕ under the assignment a)

Proposition 1. Let $\mathcal{M}$ be a DFOL model and $i \neq j \in I$.

1. $\mathcal{M} \models \mathsf{F}_{ij}$ iff r_{ij} is a partial function.
2. $\mathcal{M} \models \mathsf{INV}_{ij}$ iff r_{ij} is the inverse of r_{ji}.
3. $\mathcal{M} \models \mathsf{OD}_{ij}$ if $r_{ij}(= r_{ji}^{-1})$ is an isomorphism between a subset of $\mathbf{dom}_i$ and a subset of $\mathbf{dom}_j$. I.e., $\mathbf{dom}_i$ and $\mathbf{dom}_j$ (isomorphically) overlap.
4. $\mathcal{M} \models \mathsf{ED}_{ij}$ iff $r_{ij}(= r_{ji}^{-1})$ is an isomorphism between $\mathbf{dom}_i$ and a subset of $\mathbf{dom}_j$. I.e., $\mathbf{dom}_i$ is (isomorphically) embedded in $\mathbf{dom}_j$
5. $\mathcal{M} \models \mathsf{ID}_{ij}$ iff $r_{ij}(= r_{ji}^{-1})$ is an isomorphism between $\mathbf{dom}_i$ and $\mathbf{dom}_j$. I.e., $\mathbf{dom}_i$ is isomorphic to $\mathbf{dom}_j$.
6. $\mathcal{M} \models \mathsf{RD}$, if for every constant c of L_i and L_j, if c is interpreted in d for all $m \in \mathcal{M}_i$ then c is interpreted in $r_{ij}(d)$ for all models of $m \in \mathcal{M}_j$, and vice-versa. I.e., the constant c is rigidly interpreted by i and j in two corresponding objects.
7. Finally $\mathcal{M} \models \mathsf{IP}_{ij}$ iff $\mathcal{M}_i = \emptyset$ implies that $\mathcal{M}_j = \emptyset$. I.e., inconsistency propagates from i to j.

3 Modeling Mapping Languages in DFOL

Mapping languages formalisms are based on four main parameters: local languages and local semantics used to specify the local knowledge, and mapping languages and semantics for mappings, used to specify the semantic relations between the local knowledge. In this section we focus on the second pairs and as far as local languages and local semantics it is enough to notice that

Local languages In all approaches local knowledge is expressed by a suitable fragment of first order languages.

Local semantics with the notable exception of [Franconi and Tessaris, 2004], where authors propose an *epistemic approach* to information integration, all the other formalisms for ontology mapping assume that each local knowledge is interpreted in a (partial) state of the world and not into an epistemic state. This formally corresponds to the fact that each local knowledge base is associated with *at most one* FOL interpretation.

The first assumption is naturally captured in DFOL, by simply considering L_i to be an adequately restricted FOL language. As far as the local semantics, in DFOL models each L_i is associates with a *set of interpretations*. To simulate the single local model assumption, in DFOL it is enough to declare each L_i to be a *complete* language. This implies that all the $m \in M_i$ have to agree on the interpretation of L_i-symbols.

Notationally, $\phi, \psi, \ldots$ will be used to denote both DL expressions and FOL open formulas. If ϕ is a DL concept, $\phi(x)$ (or $\phi(x_1, \ldots, x_n)$) will denote the corresponding translation of ϕ in FOL as described in [Borgida, 1996]. If ϕ is a role R then $\phi(x, y)$ denotes its translation $P(x, y)$, and if ϕ is a constant c, then $\phi(x)$ denote its translation $x = c$. Finally we use $\mathbf{x}$ to denote a set $x_1, \ldots, x_n$ of variables.

3.1 Distributed Description Logics/C-OWL

The approach presented in [Borgida and Serafini, 2003] extends DL with a local model semantics similar to the one introduced above and so-called bridge rules to define semantic relations between different T-Boxes. A distributed interpretation for DDL on a family of DL language $\{L_i\}$, is a family $\{\mathcal{I}_i\}$ of interpretations, one for each L_i plus a family $\{r_{ij}\}_{i \neq j \in I}$ of domain relations. While the original proposal only considered subsumption between concept expressions, the model was extended to a set of five semantic relations discussed below. The Semantics of the five semantic relations defined in C-OWL is the following:

Definition 8 ([Bouquet *et al.*, 2004]). Let ϕ and ψ be either concepts, or individuals, or roles of the descriptive languages L_i and L_j respectively[2].

1. $\mathfrak{I} \models i:\phi \xrightarrow{\sqsubseteq} j:\psi$ if $r_{ij}(\phi^{\mathcal{I}_i}) \subseteq \psi^{\mathcal{I}_j}$;
2. $\mathfrak{I} \models i:\phi \xrightarrow{\sqsupseteq} j:\psi$ if $r_{ij}(\phi^{\mathcal{I}_i}) \supseteq \psi^{\mathcal{I}_j}$;
3. $\mathfrak{I} \models i:\phi \xrightarrow{\equiv} j:\psi$ if $r_{ij}(\phi^{\mathcal{I}_i}) = \psi^{\mathcal{I}_j}$;
4. $\mathfrak{I} \models i:\phi \xrightarrow{\perp} j:\psi$ if $r_{ij}(\phi^{\mathcal{I}_i}) \cap \psi^{\mathcal{I}_j} = \emptyset$;
5. $\mathfrak{I} \models i:\phi \xrightarrow{*} j:\psi$ if $r_{ij}(\phi^{\mathcal{I}_i}) \cap \psi^{\mathcal{I}_j} \neq \emptyset$;

A interpretation for a context space is a model for it if all the bridge rules are satisfied.

¿From the above satisfiability condition one can see that the mapping $i:\phi \xrightarrow{\equiv} j:\psi$ is equivalent to the conjunction of the mappings $i:\phi \xrightarrow{\sqsubseteq} j:\psi$ and $i:\phi \xrightarrow{\sqsupseteq} j:\psi$. The mapping $i:\phi \xrightarrow{\perp} j:\psi$ is equivalent to $i:\phi \xrightarrow{\sqsubseteq} j:\neg\psi$. And finally the mapping $i:\phi \xrightarrow{*} j:\psi$ is the negation of the mapping $i:\phi \xrightarrow{\sqsubseteq} j:\psi$. Therefore for the translation we will consider only the primitive mappings. As the underlying notion of a model is the same as for DFOL, we can directly try to translate bridge rules into interpretation constraints. In particular, there are no additional assumptions about the nature of the domains that have to be modeled. The translation is the following:

C-OWL	DFOL
$i:\phi \xrightarrow{\sqsubseteq} j:\psi$	$i:\phi(x^{\rightarrow j}) \rightarrow j:\psi(x)$
$i:\phi \xrightarrow{\sqsupseteq} j:\psi$	$j:\psi(x) \rightarrow i:\phi(x^{\rightarrow j})$
$i:\phi \overset{\sqsubseteq}{\not\longrightarrow} j:\psi$	No translation

We see that a bridge rule basically corresponds to the interpretation a) and d) in Figure 1. The different semantic relations correspond to the usual reads of implications. Finally negative information about mappings (i.e., $i:\phi \overset{\sqsubseteq}{\not\longrightarrow} j:\psi$ is not representable by means of DFOL interpretation constraints.

[2] In this definition, to be more homogeneous, we consider the interpretations of individuals to be sets containing a single object rather than the object itself.

3.2 Ontology Integration Framework (OIS)

Calvanese and colleagues in [Calvanese *et al.*, 2002b] propose a framework for mappings between ontologies that generalizes existing work on view-based schema integration [Ullman, 1997] and subsumes other approaches on connecting DL models with rules. In particular, they distinguish global centric, local centric and the combined approach. Differences between these approaches are in the types of expressions connected by mappings. With respect to the semantics of mappings, they are the same and are therefore treated as one.

OIS assumes the existence of a global model g into which all local models s are mapped. On the semantic level, the domains of the local models are assumed to be embedded in a global domain. Further, in OIS constants are assumed to rigidly designate the same objects across domain. Finally, global inconsistency is assumed, in the sense that the inconsistency of a local knowledge makes the whole system inconsistent. As shown in Proposition 1, we can capture these assumptions by the set of interpretation constraints ED_{sg}, RD_{sg}, and IP_{sg}, where s is the index of any source ontology and g the index of the global ontology.

According to these assumptions mappings are described in terms of correspondences between a local and the global model. The interpretation of these correspondences are defined as follows:

Definition 9 ([Calvanese *et al.*, 2002b]). Correspondences between source ontologies and global ontology are of the following four forms

1. $\mathcal{I}$ satisfies $\langle\phi, \psi, \textit{sound}\rangle$ w.r.t. the local interpretation $\mathcal{D}$, if all the tuples satisfying ψ in $\mathcal{D}$ satisfy ϕ in $\mathcal{I}$
2. $\langle\phi, \psi, \textit{complete}\rangle$ w.r.t. the local interpretation $\mathcal{D}$, if no tuple other than those satisfying ψ in $\mathcal{D}$ satisfies ϕ in $\mathcal{I}$,
3. $\langle\phi, \psi, \textit{exact}\rangle$ w.r.t. the local interpretation $\mathcal{D}$, if the set of tuples that satisfies ψ in $\mathcal{D}$ is exactly the set of tuples satisfying ϕ in $\mathcal{I}$.

¿From the above semantic conditions, $\langle\phi, \psi, \textit{exact}\rangle$ is equivalent to the conjunction of $\langle\phi, \psi, \textit{sound}\rangle$ and $\langle\phi, \psi, \textit{complete}\rangle$. It's therefore enough to provide the translation of the first two correspondences. The definitions 1 and 2 above can directly be expressed into interpretation constraints (compare Figure 1) resulting in the following translation:

GLAV Correspondence	DFOL
$\langle\phi, \psi, \textit{sound}\rangle$	$s:\psi(\mathbf{x}) \rightarrow g:\phi(\mathbf{x}^{s\rightarrow})$
$\langle\phi, \psi, \textit{complete}\rangle$	$g:\phi(\mathbf{x}) \rightarrow s:\psi(\mathbf{x}^{\rightarrow g})$

The translation shows that there is a fundamental difference in the way mappings are interpreted in C-OWL and in OIS. While C-OWL mappings correspond to a universally quantified reading (Figure 1 a), OIS mappings have an existentially quantified readings (Figure 1 b/d). We will come back to this difference later.

3.3 DL for Information Integration (DLII)

A slightly different approach to the integration of different DL models is described in [Calvanese *et al.*, 2002a]. This approach assumes a partial overlap between the domains of the models M_i and M_j rather than a complete embedding of them in a global domain. This is captured by the interpretation constraint OD_{ij}. The other assumptions (rigid designators and global inconsistency) are the same as for OIS.

An interpretation $\mathcal{I}$ associates to each M_i a domain Δ_i. These different models are connected by inter-schema assertions. Satisfiability of interschema assertions is defined as follows [3]

Definition 10 (Satisfiability of interschema assertions)**.** If $\mathcal{I}$ is an interpretation for M_i and M_j we say that $\mathcal{I}$ satisfies the interschema assertion

$\phi \sqsubseteq_{ext} \psi$, if $\phi^{\mathcal{I}} \subseteq \psi^{\mathcal{I}}$ $\quad$ $\phi \not\sqsubseteq_{ext} \psi$, if $\phi^{\mathcal{I}} \not\subseteq \psi^{\mathcal{I}}$
$\phi \equiv_{ext} \psi$, if $\phi^{\mathcal{I}} = \psi^{\mathcal{I}}$ $\quad$ $\phi \not\equiv_{ext} \psi$, if $\phi^{\mathcal{I}} \neq \psi^{\mathcal{I}}$
$\phi \sqsubseteq_{int} \psi$, if $\phi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij} \subseteq \psi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij}$
$\phi \equiv_{int} \psi$, if $\phi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij} = \psi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij}$
$\phi \not\sqsubseteq_{int} \psi$, if $\phi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij} \not\subseteq \psi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij}$
$\phi \not\equiv_{int} \psi$, if $\phi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij} \neq \psi^{\mathcal{I}} \cap \top^{\mathcal{I}}_{nij}$

As before $\equiv_{est}$ and $\equiv_{int}$ are definable as conjunctions of $\sqsubseteq_{est}$ and $\sqsubseteq_{int}$, so we can ignore them for the DFOL translation. Furthermore, a distinction is made between extensional and intentional interpretation of inter-schema assertions, which leads to different translations into DFOL.

inter-schema assertions	**DFOL**
$\phi \sqsubseteq_{ext} \psi$	$i:\phi(\mathbf{x}) \rightarrow j:\psi(\mathbf{x}^{i\rightarrow})$
$\phi \not\sqsubseteq_{ext} \psi, \phi \not\equiv_{ext} \psi$	No translation
$\phi \sqsubseteq_{int} \psi$	$i:\phi(\mathbf{x}^{\rightarrow j}) \rightarrow j:\psi(\mathbf{x})$
$\phi \not\sqsubseteq_{int} \psi, \phi \not\equiv_{int} \psi$	No translation

While the extensional interpretation corresponds to the semantics of mappings in OIS, the intentional interpretation corresponds to the semantics of mappings in C-OWL. Thus using the distinction made in this approach we get an explanation of different conceptualizations underlying the semantics of C-OWL and OIS that use an extensional and an intentional interpretation, respectively.

3.4 ϵ-connections

A different approach for defining relations between DL knowledge bases has emerged from the investigation of so-called ϵ-connections between abstract description systems [Kutz *et al.*, 2004]. Originally intended to extend the decidability of DL models by partitioning it into a set of models that use a weaker logic, the approach has recently been proposed as a framework for defining mappings between ontologies [Grau *et al.*, 2004].

In the ϵ-connections framework, for every pair of ontologies ij there is a set ϵ_{ij} of *links*, which represents binary relations between the domain of the i-th ontology and the domain of the j-th ontology. Links from i to j can be used to define i concepts, in a way that is analogous to how roles are used to define concepts. In the following table we report the syntax and the semantics of i-concepts definition based on links. (E denotes a link from i to j and C denotes a concept in j. The only assumption about the relation between domains is global inconsistency (see above).

[3] To simplify the definition we introduce the notation $\top^{\mathcal{I}}_{nij} = \top^{\mathcal{I}}_{ni} \cap \top^{\mathcal{I}}_{nj}$ for any $n \geq 1$. Notice that $\top^{\mathcal{I}}_{nij} = \Delta^n_i \cap \Delta^n_j$.

In DFOL we have only one single relation between from i to j, while in ϵ-connection there are many possible relation. However, we can use a similar trick as used in [Borgida and Serafini, 2003] to map relations on inter-schema relations: each of the relation in ϵ_{ij} acts as a r_{ij}. To represent ϵ-connection it is therefore enough to label each arrow variable with the proper link name. The arrow variable $x^{\xrightarrow{Own} j}$ is read as the arrow variable $x^{\rightarrow i}$ where r_{ij} is intended to be the interpretation of Own_{ij}. With this syntactic extension of DFOL concepts definitions based on links (denoted as E) can be codified in DFOL as follows:

ϵ-connections	**DFOL**
$\phi \sqsubseteq \exists E.\psi$	$i:\phi(x) \rightarrow j:\psi(x^{i \xrightarrow{E}})$
$\phi \sqsubseteq \forall E.\psi$	$i:\phi(x^{\xrightarrow{E} j}) \rightarrow j:\psi(x)$
$\phi \sqsubseteq \geq nE.\psi$	$i:\bigwedge_{k=1}^{n} \phi(x_1) \rightarrow$ $j:\bigwedge_{k \neq h=1}^{n} \psi(x_k^{i \xrightarrow{E}}) \wedge x_k \neq x_h$
$\phi \sqsubseteq \leq nE.\psi$	$i:\phi(x) \wedge \bigwedge_{k=1}^{n+1} x = x_k^{\xrightarrow{E} j} \rightarrow$ $j:\bigvee_{k=1}^{n+1} \left(\psi(x_k) \supset \bigvee_{h \neq k} x_h = x_k\right)$

We see that like OIS, links in the ϵ-connections framework have an extensional interpretation. The fact, that the framework distinguishes between different types of domain relations, however, makes it different from all other approaches.

Another difference to the previous approaches is that new links can be defined on the bases of existing links similar to complex roles in DL. Syntax and semantics for link constructors is defined in the usual way: $(E^-)^I = \left(E^{\mathcal{I}}\right)^{-1}$ (Inverse), $(E \sqcap F)^I = E^{\mathcal{I}} \cap F^{\mathcal{I}}$ (Conjunction), $(E \sqcup F)^I = E^{\mathcal{I}} \cup F^{\mathcal{I}}$ (Disjunction), and $(\neg E)^I = (\Delta_i \times \Delta_j) \setminus E^{\mathcal{I}}$ (Complement). Notice that, by means of inverse link we can define mapping of the b and d type. E.g., the e-connection statement $\phi \sqsubseteq \exists E^- \psi$, encodes corresponds to the DFOL bridge rules $\phi(x) : i \rightarrow \psi x^{i \rightarrow} j$: which is of type b). Similarly the e-connection $\phi \sqsubseteq \forall E^- \psi$ corresponds to a mapping of type d).

As the distinctions between different types of links is only made on the model theoretic level, it is not possible to model Boolean combinations of links. Inverse links, however, can be represented by the following axiom:

$$i:y = x^{\xrightarrow{E} j} \rightarrow j:y^{\xrightarrow{E^-} i} = x$$
$$j:y^{\xrightarrow{E^-} i} = x \rightarrow i:y = x^{\xrightarrow{E} j}$$

Finally the inclusion axioms between links, i.e., axioms of the form $E \sqsubseteq F$ where E and F are homogeneous links, i.e., links of the same ϵ_{ij}, can be translated in DFOL as follows:

$$i:x = y^{\xrightarrow{E} j} \rightarrow j:x^{i \xrightarrow{F}} = y$$

We can say that the ϵ-connections framework significantly differs from the other approaches in terms of the possibilities to define and combine mappings of different types.

4 Discussion and Conclusions

The encoding of different mapping approaches in a common framework has two immediate advantages. The first one is the ability to reason across the different frameworks. This can be done on the basis of the DFOL translation of the different approaches using the sound and complete calculus for DFOL [Ghidini and Serafini, 2000]. As there are not always complete translations, this approach does not cover all aspects of the different approaches, but as shown above, we can capture the most aspects. There are only two aspects which cannot be represented in DFOL, namely "non mappings" ($i : \phi \xrightarrow{*} j : \psi$ in C-OWL, $\phi \not\sqsubseteq_{int} \psi$ etc. in DLII) and "complex mappings" such as complex links in ϵ-connection. The second benefit is the possibility to compare the expressiveness of the approaches. We have several dimensions along which the framework can differs:

Arity of mapped items[4] C-OWL allows only to align constants, concepts and roles (2-arity relations), ϵ-connection allows to align only 1-arity items, i.e., concepts, while DLII and OIS allow to integrate n-arity items.

Positive/negative mappings Most approaches state positive facts about mapping, e.g that two elements are equivalent. The DLII and C-OWL frameworks also allow to state that two elements do not map ($\phi \not\equiv \psi$).

Domain relations The approaches make different assumptions about the nature of the domain. While C-OWL and ϵ-connections do not assume any relation between the domains, DLII assumes overlapping domains and OIS local domains that are embedded in a global domain.

Multiple mappings Only ϵ-connection approach supports form the definition of different types of mappings between ontologies that partition the inter-domain relations.

Local inconsistency Some approaches provide a consistent semantics also in the case in which some of the ontologies or mappings are inconsistent.

We summarize the comparison in the following table.

	Int. constr. (cf. fig. 1)				Mapping type			Domain relation	Arity	Local $\bot$
	a)	b)	c)	d)	Pos.	Neg.	Mult.			
C-OWL	×			×	×	×		Het.	2	×
OIS		×		×	×			Incl.	n	
DLII	×	×			×	×		Emb.	n	
ϵ-Conn.	×	×	×	×	×	×	×	Het.	1	

We conclude that existing approaches make choices along a number of dimensions. These choices are obviously influenced by the intended use. Approaches intended for database integration for example will support the mapping of n-ary items that correspond to tuples in the relational model. Despite this fact, almost no work has been done on charting the landscape of choices to be made when designing a mapping approach and for adapting the approach to the requirement of the application. The work reported in this paper provides the basis for this kind of work by identifying the possible choices on a formal level. An important topic of future work is to identify possible combinations of features for mapping languages on a formal level in order to get a more complete picture of the design space of mapping languages.

[4]Due to limited space we did not discuss the encoding of mapped items in this paper

References

[Borgida and Serafini, 2003] A. Borgida and L. Serafini. Distributed description logics: Assimilating information from peer sources. *Journal of Data Semantics*, 1:153–184, 2003.

[Borgida, 1996] A. Borgida. On the relative expressiveness of description logics and predicate logics. *Artificial Intelligence*, 82:353–367, 1996. Research Note.

[Bouquet *et al.*, 2004] P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. Contextualizing ontologies. *Journal on Web Semantics*, 1(4):xx–xx, 2004.

[Calvanese *et al.*, 2002a] Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. Description logics for information integration. In A. Kakas and F. Sadri, editors, *Computational Logic: Logic Programming and Beyond*, volume 2408 of *Lecture Notes in Computer Science*, pages 41–60. Springer, 2002.

[Calvanese *et al.*, 2002b] Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. A framework for ontology integration. In Isabel Cruz, Stefan Decker, Jerome Euzenat, and Deborah McGuinness, editors, *The Emerging Semantic Web*, pages 201–214. IOS Press, 2002.

[Franconi and Tessaris, 2004] E. Franconi and S. Tessaris. Rules and queries with ontologies: a unified logical framework. In *Workshop on Principles and Practice of Semantic Web Reasoning (PPSWR'04)*, 2004.

[Ghidini and Giunchiglia, 2001] C. Ghidini and F. Giunchiglia. Local model semantics, or contextual reasoning = locality + compatibility. *Artificial Intelligence*, 127(2):221–259, 2001.

[Ghidini and Serafini, 2000] C. Ghidini and L. Serafini. Distributed first order logics. In D.M. Gabbay and M. De Rijke, editors, *Frontiers of Combining Systems 2*, pages 121–139. Research Studies Press Ltd., 2000.

[Ghidini and Serafini, 2005] Chiara Ghidini and Luciano Serafini. Distributed first order logic - revised semantics. Technical report, ITC-irst, January 2005.

[Grau *et al.*, 2004] Bernardo Cuenca Grau, Bijan Parsia, and Evren Sirin. Working with multiple ontologies on the semantic web. In *Proceedings of the Third Internatonal Semantic Web Conference (ISWC2004)*, volume 3298 of *Lecture Notes in Computer Science*, 2004.

[Horrocks *et al.*, 2003] Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen. From shiq and rdf to owl: The making of a web ontology language. *Journal of Web Semantics*, 1(1):7–26, 2003.

[Kutz *et al.*, 2004] O. Kutz, C. Lutz, F. Wolter, and M. Zakharyaschev. E-connections of abstract description systems. *Artificial Intelligence*, 156(1):1–73, 2004.

[Ullman, 1997] Jeffrey D. Ullman. Information integration using logical views. In *Proceedings of the 6th International Conference on Database Theory*, volume 1186 of *Lecture Notes In Computer Science*, pages 19 – 40, 1997.