A Faithful Translation from Entity-relationship Schemas to the Description Logic $\mathcal{ALENI}_+$

Yuxia Lei
School of Computer Science, Qufu Normal University, Yantai Street 80, Rizhao, China, 276826
Jingying Tian
School of Architectual Engineering, Rizhao Politechnic, Rizhao, China, 276826
Baoxiang Cao
School of Computer Science, Qufu Normal University, Rizhao, China, 276826
Email: {yx_lei, lwtree}@126.com

Abstract— **An interesting topic on designing entity relationship (ER) schemas is how to transform ER schemas into knowledge bases (KBs) in description logics (DLs). It is significance in translations that one can use automated DL reasoning services to support the development and maintenance of correct ER schemas. This paper proposed a faithful translation, which translates ER schemas and ER models into KBs in the description logic $\mathcal{ALENI}_+$. The faithfulness preserves the satisfiability and the unsatisfiability, and therefore the translation is sound. The translation allows us to reduce reasoning on ER schemas to finite models reasoning on $\mathcal{ALENI}_+$ KBs.**

Index Terms— **knowledge representation, ER schemas, ER models, $\mathcal{ALENI}_+$, faithful translations**

I. Introduction

Relational databases and description logics (DLs) are two important formalisms[1]. On the hand, the ER schema is the most widespread formalism for relational database schema design, which is usually defined using a graphical notation particularly useful for an easy visualization of the data dependencies [2-4]. On the other hand, DLs are equipped with capabilities to automatically reason on knowledge bases [5,6]. Thus, providing a formalization of the ER schema in terms of DLs will allow for supporting reasoning on the ER schema such as entity satisfiability, entity subsumption and consistency of the ER schema. Within this interesting topic, many researchers have already proposed some methods such as the formal framework for translating ER schemas [7], the translation from ER schemas into $\mathcal{ALUNI}$ knowledge bases [8] and the translation from ER schemas into $\mathcal{DLR}$ knowledge bases [9].

This paper mainly focuses on the following four important questions:

◇ How to transform ER schemas into knowledge bases in DLs;

◇ How to transform ER models;

◇ How to automatically decide whether a given ER model satisfies a correct ER schema; and

◇ How to ensure the soundness and the completeness of the translations. To represent ER models and ER schemas in description logics, an ER schema can be taken as a logical theory, and an ER model can be taken as a model for a logical theory. To translate the ER schemas into description logics, we can translate the ER models into models for the logical theories in description logics. In this translation, a set of entities is taken as a concept, and so is a set of relationships. The objects are either entities or relationships. The roles are classified into two kinds: the roles correspond to the attributes in the ER schemas, and the roles correspond to the ER-roles in the ER schemas.

Our main contributions in this paper are to propose a description logic called $\mathcal{ALENI}_+$ and a faithful translation from ER schemas into $\mathcal{ALENI}_+$ knowledge bases. The faithfulness ensures the preservations of the satisfiability and the unsatisfiability, which means that our translation is sound and complete. By this translation, one can not only design a correct ER schema, but also obtain whether a given ER model satisfies the ER schema. $\mathcal{ALENI}_+$ is quite expressive and includes a novel combination of constructs, including existential quantifications, existential number restrictions, and inverse roles. The significance of the translation is that it allows us to reduce reasoning on ER schemas to finite models reasoning on $\mathcal{ALENI}_+$ knowledge bases.

The paper is organized as follows: the next section introduces the description logic $\mathcal{ALENI}_+$, including its syntax and semantics; the third section takes ER schemas as logical theories and ER models as models for ER schemas, and translates them into $\mathcal{ALENI}_+$ knowledge bases, and proves the faithfulness of the translation; the last section concludes the paper.

Note that, in this paper, we respectively apply boldface, italic and typewriter to represent symbols in ER schemas, symbols in ER models, and symbols in $\mathcal{ALENI}_+$, for example, $\mathbf{E}$, E, $\mathtt{E}$.

This work is supported by the National Natural Science Foundation of China(60573063, 60573064 and 61035004) and the Natural Science Foundation of Shandong Province(No. ZR2011FQ026)

II. Description logic $\mathcal{ALENI}_+$

To represent ER schemas and ER models in terms of DLs knowledge bases, we introduce a DL called

doi:10.4304/jsw.8.2.296-301

$\mathcal{ALENI}_+$. In $\mathcal{ALENI}_+$, *concepts* are formed according to the following syntax:

constructor	syntax
atomic concept	$\mathtt{W}$
universal concept	$\top$
bottom concept	$\bot$
conjunction	$\mathtt{C}\sqcap\mathtt{D}$
universal quantification	$\forall\mathtt{r.C}$
existential quantification(E)	$\exists\mathtt{r.C}$
existential number restrictions(N)	$\exists^{\geq m}\mathtt{r.C}$
	$\exists^{\leq M}\mathtt{r.C}$
inverse role(I)	$\mathtt{r}^-$

Concepts are interpreted as subset of a domain and roles as binary relations over that domain. $\mathtt{C}\sqcap\mathtt{D}$ represents the *conjunction* of two concepts and is interpreted as set intersection. Consequently, $\top$ represents the whole domain, and $\bot$ the empty set. $\forall\mathtt{r.C}$ is called *universal quantification over roles* and is used to denote those elements of the interpretation domain that are connected through role $\mathtt{r}$ only to instances of the concept $\mathtt{C}$. $\exists^{\geq m}\mathtt{r.C}$ and $\exists^{\leq M}\mathtt{r.C}$ are called *existential number restrictions*, and impose in their instances restrictions on the minimum and maximum number of objects in concept $\mathtt{C}$ they are connected to through role $\mathtt{r}$, which are mainly different to the description logic $\mathcal{ALUNI}$, where $\mathtt{r}$ and $\mathtt{C}$ are role name and concept description, respectively. More formally, an interpretation $I=(\Delta,\cdot^I)$ consists of an interpretation domain Δ and an interpretation function $\cdot^I$ that maps every concept $\mathtt{C}$ to a subset $\mathtt{C}^\mathtt{I}$ of Δ and every role $\mathtt{r}$ to a subset $\mathtt{r}^\mathtt{I}$ of $\Delta\times\Delta$ according to the following semantic rules:

syntax	semantics
$\mathtt{W}$	$\mathtt{W}^I\subseteq\Delta$
$\top$	Δ
$\bot$	$\emptyset$
$\mathtt{C}\sqcap\mathtt{D}$	$\mathtt{C}^\mathcal{I}\cap\mathtt{D}^\mathcal{I}$
$\forall\mathtt{r.C}$	$\{x\in\Delta\|\forall y((x,y)\in\mathtt{r}^I\rightarrow y\in\mathtt{C}^I)\}$
$\exists\mathtt{r.C}$	$\{x\in\Delta\|\exists y(x,y)\in\mathtt{r}^I\wedge y\in\mathtt{C}^I\}$
$\exists^{\geq m}\mathtt{r.C}$	$\{x\in\Delta\|\|\{y\in\Delta:(x,y)\in\mathtt{r}^I\wedge y\in\mathtt{C}^I\}\|\geq m\}$
$\exists^{\leq M}\mathtt{r.C}$	$\{x\in\Delta\|\|\{y\in\Delta:(x,y)\in\mathtt{r}^I\wedge y\in\mathtt{C}^I\}\|\leq M\}$
$\mathtt{r}^-$	$\{(y,x)\|(x,y)\in\mathtt{r}^I\}$

Similar to other description logics, an $\mathcal{ALENI}_+$ knowledge base also consists of $TBox$ and $ABox$[9-11]. $TBox$ is a set of the following statements: $\mathtt{C}\sqsubseteq\mathtt{D}$, where $\mathtt{C},\mathtt{D}$ are concepts. $ABox$ is a set of the following statements: $\mathtt{C}(\mathtt{e})$ or $\mathtt{r}(\mathtt{e}_1,\mathtt{e}_2)$, where $\mathtt{e},\mathtt{e}_1,\mathtt{e}_2$ are constant symbols, $\mathtt{C}$ is a concept. Given a knowledge base $KB=(TBox,ABox)$, for any statement $\mathtt{C}\sqsubseteq\mathtt{D}$ in TBox, an interpretation I satisfies the statement $\mathtt{C}\sqsubseteq\mathtt{D}$ if $\mathtt{C}^I\subseteq\mathtt{D}^I$, denoted by $I\models\mathtt{C}\sqsubseteq\mathtt{D}$. An interpretation I is a model for a $TBox$ if I satisfies all the statements in the $TBox$. For any statement $\mathtt{C}(\mathtt{e})$ or $\mathtt{r}(\mathtt{e}_1,\mathtt{e}_2)$ in $ABox$, an interpretation I satisfies the statement $\mathtt{C}(\mathtt{e})$ if $\mathtt{e}^\mathcal{I}\in\mathtt{C}^I$. An interpretation I satisfies the statement $\mathtt{r}(\mathtt{e}_1,\mathtt{e}_2)$ if $(\mathtt{e}_1^I,\mathtt{e}_2^I)\in\mathtt{r}^I$. An interpretation I is a model for a $ABox$ if I satisfies all the statements in the $ABox$. An interpretation I is a model for a KB if it is both a model for $TBox$ and $ABox$.

III. ER SCHEMAS/ER MODELS TAKEN AS LOGICAL THEORIES/MODELS

In this section, we can logically represent the connection between ER models and ER schemas, by taking ER schemas as logical theories and ER models as the models for ER schemas, and then provide a faithful translation from ER schemas into $\mathcal{ALENI}_+$ knowledge bases. Generally, setting up a translation from one formalism to another formalism is usually taken into account the following logic properties: the soundness and the completeness. However, the two properties do not immediately lead to the preservation of the unsatisfiability. In order to preserve the satisfiability and the unsatisfiability, we define the faithfulness of the translation, which implies that the translation is sound and complete.

A. ER schemas

An ER schema $\mathbf{S}$ is constructed starting from pairwise disjoint of entity name symbols, relationship name symbols, ER-role name symbols, attribute symbols, and domain symbols. Formally, an ER schema $\mathbf{S}$ is a septuple $(E,R,A,\rho,k,U,\mathbf{isa})$, where

$\diamond$ E is a set of entity set names, where its elements are denoted by $\mathbf{E}_1,...,\mathbf{E}_m$;

$\diamond$ R is a set of relationship set names, where its elements are denoted by $\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n)$, where $n\leq m$; $\mathbf{r}_i$ is the ER-role name of $\mathbf{E}_i$ in $\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n)$, where $1\leq i\leq n$;

$\diamond$ A is a set of the attributes, such that for each attribute $a\in A$ there is a non-empty domain D_a;

$\diamond$ ρ is a function such that for any $\mathbf{E}\in\mathbf{E}$, $\rho(\mathbf{E})\subseteq\mathbf{A}$; and for any $\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n)$,

$$\rho(\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n))\subseteq A,\ \text{and}$$

$$\rho(\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n))\supseteq\rho(\mathbf{E}_1)\cup\cdots\cup\rho(\mathbf{E}_n);$$

$\diamond$ k is a function such that for any $\mathbf{E}\in\mathbf{E}$, $k(\mathbf{E})\subseteq\rho(\mathbf{E})$; and for any $\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n)$,

$$k(\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n))\subseteq\rho(\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n)),\ \text{and}$$

$$k(\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n))\supseteq k(\mathbf{E}_1)\cup\cdots\cup k(\mathbf{E}_n);$$

$\diamond$ U is a set of participation constraints of the form $(\mathbf{E},m,M,\mathbf{R})$; and

$\diamond$ **isa** is a binary relation on E, that is, $\mathbf{isa}\subseteq E\times E$, which is irreflexive, antisymmetric and transitive.

Example 1 [12]. An ER schema of some college database is shown in the following Figure 1. The ER schema uses the notions of entity, relationship and attribute. Entities can be described as distinct objects that need to be represented in the database; relationships reflect interactions between entities, and propertied of entities and relationships are described by attributes. For example, the set of entities **Students** of the college database has the attributes student identification number (*stno*), student name (*name*), street address(*addr*), city(*city*), state of residence(*state*) and zip code (*zip*).

The ER schema can be formalized as follows:

1

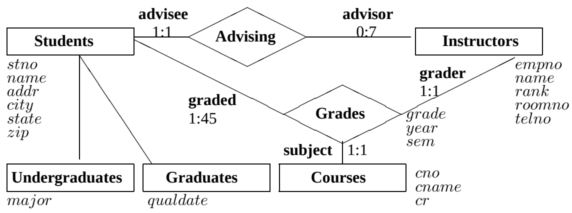

Figure 1: An ER schema of some college database

Figure 1. An ER schema of some college database. Entity sets are represented by rectangles and sets of relationship by diamonds, and attributes of sets of entities or relationships are listed close to the graphical representations of the those sets.

$\mathbf{E}_1 =$ **Students**,
$\mathbf{E}_2 =$ **Instructors**,
$\mathbf{E}_3 =$ **Courses**,
$\mathbf{E}_4 =$ **Graduates**,
$\mathbf{E}_5 =$ **Undergraduates**;
$\mathbf{R}_1(\mathbf{E}_2, \mathbf{E}_1) =$ **Advising**(**Instructors**, **Students**),
$\mathbf{R}_2(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3) =$ **Grades**(**Students**, **Instructors**, **courses**);

The ER-role names of $\mathbf{E}_1$ and $\mathbf{E}_2$ in $\mathbf{R}_1$ is respectively $\mathbf{r}_{12}$ and $\mathbf{r}_{11}$;

The ER-role names of $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ in $\mathbf{R}_2$ is respectively $\mathbf{r}_{21}$, $\mathbf{r}_{22}$ and $\mathbf{r}_{23}$;

$\rho(\mathbf{E}_1) = \{stno, name, addr, city, state, zip\}$,
$\rho(\mathbf{E}_2) = \{empno, name, rank, roomno, telno\}$,
$\rho(\mathbf{E}_3) = \{cno, cname, credits\}$;
$\rho(\mathbf{R}_1) = \{stno, empno\}$,
$\rho(\mathbf{R}_2) = \{stno, empno, cno, sem, year, grade\}$;
$k(\mathbf{E}_1) = \{stno\}$,
$k(\mathbf{E}_2) = \{empno\}$,
$k(\mathbf{E}_3) = \{cno\}$;
$k(\mathbf{R}_1) = \{stno, empno\}$,
$k(\mathbf{R}_2) = \{stno, empno, cno, sem, year, grade\}$;
U=$\{(\mathbf{E}_1, 1, 1, \mathbf{R}_1)$, $(\mathbf{E}_2, 0, 7, \mathbf{R}_1)$, $(\mathbf{E}_1, 1, 45, \mathbf{R}_2)$, $(\mathbf{E}_2, 1, 1, \mathbf{R}_2)$, $(\mathbf{E}_3, 1, 1, \mathbf{R}_2)\}$; and
isa=$\{$(**Graduates**, **Students**), (**Undergraduates**, **Students**)$\}$.

B. The ER schemas taken as logical theories

Let L be a logical language containing the following symbols:

◇ A set $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ of entity set names;
◇ A set $\{\mathbf{R}_1, ..., \mathbf{R}_n\}$ of relationship set names;
◇ A set $\{a_1, ..., a_n\}$ of attributes and a symbol A for set $\{a_1, ..., a_n\}$, such that for each attribute a, there is a domain D_a of attribute a;
◇ Two function symbols ρ, k; and
◇ Two binary relation symbols **isa**, $\in$.

A statement ϕ in L is defined as follows:

$$\begin{array}{rcll} \phi & = & \mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n) & \text{or} \\ & & a \in \rho(\mathbf{E}) & \text{or} \\ & & a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)) & \text{or} \\ & & a \in k(\mathbf{E}) & \text{or} \\ & & a \in k(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n) & \text{or} \\ & & (\mathbf{E}, m, M, \mathbf{R}) & \text{or} \\ & & \mathbf{E} \textbf{ isa } \mathbf{E}'. & \end{array}$$

An ER schema S is a set of statements in the language L. Formally,

$$S = \{\phi : \phi \text{ is a statement in the language } L\}.$$

C. The ER models taken as models for logic theories

An ER model M is a quadruple $(\Sigma, \mathcal{A}, \{D_a : a \in \mathcal{A}\}, I)$ such that Σ is a non-empty universe; $\mathcal{A}$ is a set of attributes, such that for each $a \in \mathcal{A}$, there is a non-empty attribute domain D_a; I is an interpretation such that

◇ For each entity set name $\mathbf{E}$, $I(\mathbf{E}) \subseteq \Sigma$;
◇ For each relationship set name $\mathbf{R}$, $I(\mathbf{R}) \subseteq \Sigma^n$;
◇ For each attribute $a \in \mathcal{A}$,

$$I(a) \subseteq (\Sigma \cup \bigcup_{i \in \omega} \Sigma^i) \times \bigcup_{a \in \mathcal{A}} D_a,$$

where ω is a set of some natural numbers; and

◇ $I(\textbf{isa}) =\subseteq$, which means that **isa** is a subconcept-superconcept relation; and there is a function ι such that for each attribute $a \in \rho(\mathbf{E}) \subseteq \mathcal{A}$ and $e \in I(\mathbf{E}) \subseteq \Sigma$,

$$\iota(e, a) = v \in D_a;$$

and for each attribute $a \in \rho(\mathbf{R})$ and $e_1, ..., e_n \in \Sigma$,

$$\iota((e_1, ..., e_n), a) = v \in D_a.$$

A statement ϕ is satisfied in the ER model M, denoted by $M \models \phi$, if

Case 1: if $\phi = \mathbf{E} \textbf{ isa } \mathbf{E}'$, then $I(\mathbf{E}) \subseteq I(\mathbf{E}')$;
Case 2: if $\phi = a \in \rho(\mathbf{E})$, then $a \in \rho(I(\mathbf{E}))$;
Case 3: if $\phi = a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))$,
then $a \in \rho(I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$;
Case 4: if $\phi = a \in k(\mathbf{E})$, then $a \in k(I(\mathbf{E}))$;
Case 5: if $\phi = a \in k(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))$,
then $a \in k(I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$;
Case 6: if $\phi = \mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$,
then $\forall e_1, ..., e_n \in \Sigma((e_1, ..., e_n) \in I(\mathbf{R}) \Rightarrow e_1 \in I(\mathbf{E}_1) \& \cdots \& e_n \in I(\mathbf{E}_n))$;
Case 7: if $\phi = (\mathbf{E}, m, M, \mathbf{R})$, where $\mathbf{E} = \mathbf{E}_i$ for some i in $\mathbf{R}(\mathbf{E_1}, ..., \mathbf{E_n})$,
then $\forall e \in I(\mathbf{E})(m \leq |\Phi| \leq M)$, where $\Phi = \{t : t = (e_1, ..., e_{i-1}, e, e_{i+1}, ..., e_n) \in I(\mathbf{R})\}$.

An ER model M satisfies the ER schema S, denoted by $M \models S$, if for each statement ϕ in $S, M \models \phi$.

D. The translation σ from the ER schemas into $\mathcal{ALENI}_+$ knowledge bases

Let σ be the translation translating statements in an ER schema into statements in $\mathcal{ALENI}_+$. Then, σ is defined as follows:

Let L be the logical language for the description logic, which contains the following symbols:

⋄ An atomic concept name $\mathtt{E}$ for each entity set name $\mathbf{E} \in \{\mathbf{E}_1, ..., \mathbf{E}_n\}$;

⋄ An atomic concept name $\mathtt{R}$ for each relationship set name $\mathbf{R} \in \{\mathbf{R}_1, ..., \mathbf{R}_n\}$;

⋄ A role name $\mathtt{a}$ for each attribute $a \in A$;

⋄ An atomic concept name $\mathtt{D_a}$ for each $a \in A$; and

⋄ For each relationship set $\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$ and each $1 \leq i \leq n$, there is a ER-role $\mathtt{r_i}$.

At the language level, σ translates entity set names and relationship set names into concept names, attributes into role names, attribute domains into concept names; and **isa** into $\sqsubseteq$. Precisely, for any entity set name $\mathbf{E}$, relationship set name $\mathbf{R}$, any attribute a, any ER-role $\mathbf{r}_i$,

$$\begin{aligned}
&\sigma(\mathbf{E}) = \mathtt{E}, \text{atomic concept name}\\
&\sigma(\mathbf{R}) = \mathtt{R}, \text{atomic concept name}\\
&\sigma(a) = \mathtt{a}, \text{atomic role name}\\
&\sigma(D_a) = \mathtt{D_a}, \text{atomic concept name}\\
&\sigma(\mathbf{r}_i) = \mathtt{r_i}, \text{atomic role name}\\
&\sigma(\mathbf{isa}) = \sqsubseteq .
\end{aligned}$$

At the syntactical level, σ translates statements into concepts and statements in the description logic:

$$\begin{aligned}
&\sigma(\mathbf{E}\ \mathbf{isa}\ \mathbf{E}') = \mathtt{E} \sqsubseteq \mathtt{E}';\\
&\sigma(a \in \rho(\mathbf{E})) = \mathtt{E} \sqsubseteq \exists \mathtt{a}.\mathtt{D_a};\\
&\sigma(a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))) = \mathtt{R}(\mathtt{E_1}, ..., \mathtt{E_n}) \sqsubseteq \exists \mathtt{a}.\mathtt{D_a};\\
&\sigma(a \in k(\mathbf{E})) = \mathtt{E} \sqsubseteq \exists \mathtt{a}.\mathtt{D_a};\\
&\sigma(a \in k(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))) = \mathtt{R}(\mathtt{E_1}, ..., \mathtt{E_n})) \sqsubseteq \exists \mathtt{a}.\mathtt{D_a};\\
&\sigma(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)) = \mathtt{R}(\mathtt{E_1}, \ldots, \mathtt{E_n})\\
&\sqsubseteq \forall \mathtt{r_1}.\mathtt{E_1} \sqcap \cdots \sqcap \forall \mathtt{r_n}.\mathtt{E_n} \sqcap \sqcap_{a \in \rho(\mathbf{R}(\mathbf{E}_1,...,\mathbf{E}_n))} \exists \mathtt{a}.\mathtt{D_a};\\
&\sigma((\mathbf{E}, m, M, \mathbf{R})) = \mathtt{E} \sqsubseteq \exists^{\geq \mathtt{m}} \mathtt{r_i^-}.\mathtt{R} \sqcap \exists^{\leq \mathtt{M}} \mathtt{r_i^-}.\mathtt{R},\\
&\text{where } \mathbf{E} \text{ is some } \mathbf{E}_i.
\end{aligned}$$

Example 2. Let $KB = (ABox, TBox) = \sigma(S)$, where $TBox$ is the set of statements, and $ABox = \emptyset$. For example, by applying the translation presented above to the ER schema in Example 1, we obtain the following $\mathcal{ALENI}_+$ knowledge base $KB = (TBox, ABox) = \sigma(S)$, where $\sigma(S)$ contains the following statements:

$$\begin{aligned}
&\mathtt{E_1} \sqsubseteq \exists \mathtt{stno}.\mathtt{D_{stno}} \sqcap \exists \mathtt{name}.\mathtt{D_{name}} \sqcap \exists \mathtt{addr}.\mathtt{D_{addr}},\\
&\mathtt{E_1} \sqsubseteq \exists \mathtt{city}.\mathtt{D_{city}} \sqcap \exists \mathtt{state}.\mathtt{D_{state}} \sqcap \exists \mathtt{zip}.\mathtt{D_{zip}};\\
&\mathtt{E_2} \sqsubseteq \exists \mathtt{cno}.\mathtt{D_{cno}} \sqcap \exists \mathtt{cname}.\mathtt{D_{cname}} \sqcap \exists \mathtt{credits}.\mathtt{D_{credits}};\\
&\mathtt{E_3} \sqsubseteq \exists.\mathtt{empno}.\mathtt{D_{empno}} \sqcap \exists \mathtt{name}.\mathtt{D_{name}} \sqcap \exists \mathtt{rank}.\mathtt{D_{rank}},\\
&\mathtt{E_3} \sqsubseteq \exists \mathtt{roomno}.\mathtt{D_{roomno}} \sqcap \exists \mathtt{telno}.\mathtt{D_{telno}};\\
&\mathtt{R_1} \sqsubseteq \exists \mathtt{stno}.\mathtt{D_{stno}} \sqcap \exists \mathtt{empno}.\mathtt{D_{empno}};\\
&\mathtt{R_2} \sqsubseteq \exists \mathtt{stno}.\mathtt{D_{stno}} \sqcap \exists \mathtt{empno}.\mathtt{D_{empno}} \sqcap \exists \mathtt{cno}.\mathtt{D_{cno}},\\
&\mathtt{R_2} \sqsubseteq \exists \mathtt{sem}.\mathtt{D_{sem}} \sqcap \exists \mathtt{year}.\mathtt{D_{year}} \sqcap \exists \mathtt{grade}.\mathtt{D_{grade}};
\end{aligned}$$

and

$$\begin{aligned}
&\mathtt{E_1} \sqsubseteq \exists^{=1} \mathtt{r_{11}}.\mathtt{R_1} \equiv \exists^{\geq 1} \mathtt{r_{11}}.\mathtt{R_1} \sqcap \exists^{\leq 1} \mathtt{r_{11}}.\mathtt{R_1};\\
&\mathtt{E_2} \sqsubseteq \exists^{\geq 0} \mathtt{r_{12}}.\mathtt{R_1} \sqcap \exists^{\leq 7} \mathtt{r_{12}}.\mathtt{R_1};\\
&\mathtt{E_1} \sqsubseteq \exists^{\geq 1} \mathtt{r_{21}}.\mathtt{R_2} \sqcap \exists^{\leq 45} \mathtt{r_{21}}.\mathtt{R_2};\\
&\mathtt{E_2} \sqsubseteq \exists^{\geq 1} \mathtt{r_{22}}.\mathtt{R_2} \sqcap \exists^{\leq 1} \mathtt{r_{22}}.\mathtt{R_2};\\
&\mathtt{E_3} \sqsubseteq \exists^{\geq 1} \mathtt{r_{23}}.\mathtt{R_2} \sqcap \exists^{\leq 1} \mathtt{r_{23}}.\mathtt{R_2};\\
&\mathtt{E_4} \sqsubseteq \mathtt{E_1},\\
&E_5 \sqsubseteq E_1.
\end{aligned}$$

Given an ER model $M = (\Sigma, \mathcal{A}, \{D_a : a \in \mathcal{A}\}, I)$, we define the translated model $\sigma(M) = (\Delta, I')$ of M as follows: $\Delta = \Sigma \cup \bigcup_{a \in \mathcal{A}} D_a$, and I' is an interpretation such that

⋄ For any entity set name $\mathbf{E}$, $I'(\mathtt{E})$=$I'(\sigma(\mathbf{E})) = I(\mathbf{E})$;

⋄ For any relationship set name $\mathbf{R}$, $I'(\mathtt{R})$=$I'(\sigma(\mathbf{R})) = I(\mathbf{R})$;

⋄ For any attribute $a \in \mathcal{A}$, $I'(\mathtt{a})$=$I'(\sigma(a)) = I(a)$; and

⋄ For any ER-role $\mathtt{r_i}$ in $\mathbf{R}$, $I'(\mathtt{r_i})$=$\{(t, e) \in \Delta \times \Delta : t \in I(\mathbf{R}) \wedge r_i(t) = e\}$, $i \in \{1, 2, \ldots, n\}$.

E. The Complexity of the Translation

In order to describe the complexity of the translation, we only take into the syntactical-level translations consideration. Let S be an ER schema, based on the septuple $(E, R, A, \rho, k, U, \mathbf{isa})$. We distinguish the following four cases: for the statements with the form $\mathbf{E}$ **isa** $\mathbf{E}'$, the total time of transformations is the number of elements in **isa**, that is, $|\mathbf{isa}|$; for the statements with the form $\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$, the total time is the number of elements in R; for the statements with the form $(\mathbf{E}, m, M, \mathbf{R})$, the total time is the number of elements in U; and for other statements, the total time is at most the number of elements in A. Thus, the complexity of the translation is $|\mathbf{isa}| + |R| + |U| + |A|$, which means that the algorithm is linear.

F. The Faithfulness of the Translation

In this section, we firstly define the faithfulness, which preserves the satisfiability and the unsatisfiability, and then show that our translation is faithful, which implies that the translation is sound and complete.

Definition 1. Let σ be a translation from an ER schema S into an $\mathcal{ALENI}_+$ knowledge base $\sigma(S)$. For any ER model M, if σ satisfies the following condition: M satisfies S if and only if $\sigma(M)$ satisfies $\sigma(S)$, that is, $M \models S$ if and only if $\sigma(M) \models \sigma(S)$, then σ is faithful.

Proposition 1. Let σ be a faithful translation from an ER schema S into an $\mathcal{ALENI}_+$ knowledge base. Then for any statement $\phi \in S$, ϕ is satisfiable in S if and only if $\sigma(\phi)$ is satisfiable in $\sigma(S)$, that is, there is a ER model M such that $M \models \phi$ if and only if $\sigma(M) \models \sigma(\phi)$.

Proposition 2. For any ER schema $S = \{\phi : \phi \text{ is a statement in } L\}$ and ER model $M = (\Sigma, \mathcal{A}, \{D_a : a \in \mathcal{A}\}, I)$,

$$M \models S \text{ if and only if } \sigma(M) \models \sigma(S).$$

Proof: ($\Rightarrow$) Assume that $M \models S$. We show that $\sigma(M) \models \sigma(S)$. By the construction of $\sigma(S)$, we distinguish seven cases to prove this proposition as follows:

Case 1: If **E isa E**$'$, then $I' \models \mathtt{E} \sqsubseteq \mathtt{E}'$. By the definition of I' and $M \models S$, $I'(\mathtt{E}) = I(\mathbf{E}) \subseteq I(\mathbf{E}') = I'(\mathtt{E}')$;

Case 2: If $a \in \rho(\mathbf{E})$, then $I' \models \mathtt{E} \sqsubseteq \exists \mathtt{a}.\mathtt{D_a}$. By the definition of I', $I'(\mathtt{E}) = I(\mathbf{E})$. By $M \models S$, $a \in \rho(I(\mathbf{E}))$. For any element $e \in I(\mathbf{E})$, because M is an ER model, $\iota(e, a) = v \in D_a$, and further $(e, v) \in I(a) = I'(\mathtt{a})$, $v \in D_a$, that is, $I'(\mathtt{E}) \sqsubseteq I'(\exists \mathtt{a}.\mathtt{D_a})$;

Case 3: If $a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))$, then

$$I' \models \mathtt{R}(\mathtt{E_1}, ..., \mathtt{E_n}) \sqsubseteq \exists \mathtt{a}.\mathtt{D_a}.$$

By the definition of I', $I'(\mathtt{R}) = I(\mathbf{R})$, and by $M \models S$, $a \in \rho(I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$. For any element $t = (e_1, \ldots, e_n) \in I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))$, because M is a ER model, $\iota(t, a) = v \in D_a$, and further $(t, v) \in I(a)$, $v \in D_a$, that is,

$$I'(\mathtt{R}(\mathtt{E_1}, ..., \mathtt{E_n})) \sqsubseteq \mathtt{I}'(\exists \mathtt{a}.\mathtt{D_a});$$

Case 4: If $a \in k(\mathbf{E})$, then $I' \models \mathtt{E} \sqsubseteq \exists \mathtt{a}.\mathtt{D_a}$. The process of proof is similar to the proof of Case 3;

Case 5: If $a \in k(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))$, then

$$I' \models \mathtt{R}(\mathtt{E_1}, ..., \mathtt{E_n}) \sqsubseteq \exists \mathtt{a}.\mathtt{D_a},$$

which follows immediately from Case 2;

Case 6: Assume that $\mathbf{r}_i$ is the ER-role name of $\mathbf{E}_i$ in $\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$, $i \in \{1, 2, \ldots, n\}$. For any $t = (e_1, \ldots, e_n) \in I'(\mathtt{R}(\mathtt{E_1}, ..., \mathtt{E_n}) = I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$, by $M \models S$ and the definition of I', $e_i \in I(\mathbf{E}_i) = I'(\mathtt{E_i})$, where $i \in \{1, 2, \ldots, n\}$. That is, if $(t, e_i) \in I'(\mathtt{r_i})$, then $e_i \in I'(\mathtt{E_i})$, and hence

$$I' \models \mathtt{R} \sqsubseteq \forall \mathtt{r_1}.\mathtt{E_1} \sqcap \cdots \sqcap \forall \mathtt{r_n}.\mathtt{E_n}.$$

By Case 3, $I' \models \mathtt{R} \sqsubseteq \sqcap_{a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))} \exists \mathtt{a}.\mathtt{D_a}$. Hence,

$$I' \models \forall \mathtt{r_1}.\mathtt{E_1} \sqcap \cdots \sqcap \forall \mathtt{r_n}.\mathtt{E_n} \sqcap \sqcap_{a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))} \exists \mathtt{a}.\mathtt{D_a};$$

Case 7: Given $(\mathbf{E}, m, M, \mathbf{R})$, assume that $\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$, and $\mathbf{E} = \mathbf{E}_i$, and $\mathbf{r}_i$ is the ER-role name of $\mathbf{E}_i$ in $\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$. By $M \models S$,

$$\forall e \in I(\mathbf{E})(m \leq |\Phi| \leq M),$$

where $\Phi = \{(e_1, ..., e_n) : (e_1, ..., e, ..., e_n) \in I(\mathbf{R})\}$. Hence, there are at least m and at most M elements $t \in I'(\mathtt{R})$ such that $r_i(t) = e$. In other words, there exist at least m and at most M pairs in $I'(\mathtt{r_i})$ that have e as their second component, and moreover all the first components are elements in $I'(\mathtt{R})$. Therefore,

$$I' \models \mathtt{E} \sqsubseteq \exists^{\geq \mathtt{m}} \mathtt{r_i^-}.\mathtt{R} \sqcap \exists^{\leq \mathtt{M}} \mathtt{r_i^-}.\mathtt{R}.$$

($\Leftarrow$) Let $\sigma(M) \models \sigma(S)$. For any statement ϕ in S, we have to show that $M \models \phi$.

Case 1: $\phi =$ **E isa E**$'$. By $\sigma(M) \models \sigma(S)$ and the definition of I', $I(\mathbf{E}) = I'(\sigma(\mathbf{E})) \subseteq I'(\sigma(\mathbf{E}')) = I(\mathbf{E}')$, that is, $M \models \phi$;

Case 2: $\phi = a \in \rho(\mathbf{E})$. By $\sigma(M) \models \sigma(S)$ and the definition of I', $I(\mathbf{E}) = I'(\sigma(\mathbf{E})) \subseteq I'(\exists \mathtt{a}.\mathtt{D_a})$. For any $e \in I(\mathbf{E})$, there exists an element $v \in D_a$ such that $(e, v) \in I(a)$, that is, all elements in $I(\mathbf{E})$ have values on the attribute a, and hence $a \in \rho(I(\mathbf{E}))$;

Case 3: If $\phi = a \in \rho(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n))$, then $M \models a \in \rho(I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$, which follows directly from Case 2;

Case 4: If $\phi = a \in k(\mathbf{E})$, then $M \models a \in \rho(I(\mathbf{E}))$. The result follows directly from Case 2;

Case 5: if $\phi = a \in k(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$, then $M \models a \in k(I(\mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)))$, which follows directly from Case 3;

Case 6: $\phi = \mathbf{R}(\mathbf{E}_1, ..., \mathbf{E}_n)$. For any $x_1, \ldots, x_n \in \Sigma$, let $t = (x_1, \ldots, x_n) \in I(\mathbf{R}) = I'(\mathtt{R})$. Because

$$I' \models \mathtt{R} \sqsubseteq \forall \mathtt{r_1}.\mathtt{E_1} \sqcap \cdots \sqcap \forall \mathtt{r_n}.\mathtt{E_n},$$

$x_i \in I'(\mathtt{E_i}) = I(\mathbf{E}_i)$, where $i \in \{1, 2, \ldots, n\}$;

Case 7: $\phi = (\mathbf{E}, m, M, \mathbf{R})$, let $\mathbf{E} = \mathbf{E}_i$. For any $e \in I(\mathbf{E}) = I'(\mathtt{E})$, by

$$I' \models \mathtt{E} \sqsubseteq \exists^{\geq \mathtt{m}} \mathtt{r_i^-}.\mathtt{R} \sqcap \exists^{\leq \mathtt{M}} \exists \mathtt{r_i^-}.\mathtt{R}$$

and the definition of $I'(\mathtt{r_i})$,

$$m \leq |\Phi| \leq M,$$

where $\phi = \{t = (e_1, \ldots, e_{i-1}, e, e_{i+1}, \ldots, e_n) : r_i(t) = e \wedge t \in I(\mathbf{R})$.

Hence, we have the following commutative diagram:

$$\begin{tikzcd}
\text{ER schema} \arrow[r, "\sigma"] \arrow[d, "\text{interpretation}"'] & KB \arrow[d, "\text{interpretation}"] \\
\text{ER model} \arrow[r, "\sigma"'] & \text{model for } KB
\end{tikzcd}$$

IV. Conclusion

This paper proposed a faithful translation from ER schemas into $\mathcal{ALENI}_+$ knowledge bases, which allows us to reduce reasoning on ER schemas to finite models reasoning on $\mathcal{ALENI}_+$ KBs. However, several problems remain unsolved. One unsolved problem is how to transform ER schemas and ER models with imprecise information. Future works focus on these questions.

Acknowledgment

The authors would like to thank the anonymous referees for their careful reading and helpful comments, which led to an improved presentation of the manuscript. The idea came from ICT, CAS.

References

[1] Z. Ma, F. Zhang, "Extracting knowledge from fuzzy relational databases with description logic", Integrated Computer-Aided Engineering, vol.18, pp.181-200, 2011.

[2] B. Thalheim, Entity-relationship modelling-Fundamentals of database technology. Springer, Berlin, 2000.

[3] O. Rauh, E. Stickel, "Standard transformations for the normalization of ER schemata", Information Systems, vol.21, pp.187-208, 1996.

[4] B. Thalheim, Entity-Relationship Modeling: Foundations of Database Technology, Springer-Verlag, New York, 2000.
[5] F. Baader, J. Hladik, C. Lutz, et al.,"From tableaux to automata for description logics", Fundamenta Informaticae, vol.57, pp. 1-33, 2003.
[6] G. Hao, S. Ma, Y. Sui, et al., "An unified dynamic description logic model for databases: relational data, relational operations and queries", Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling, vol.83, pp. 121-126, 2007.
[7] P. McBrien, A. Poulovassilis, "A formal framework for ER schema transformation", Conceptual Modelling-ER'97, 16th International Conference on Conceptual Modelling, LNCS 1331, pp. 408-421, 1997.
[8] D. Calvanese, M. Lenzerini, D. Nardi, "Unifying class-based representation formalisms", Journal of Artificial Intelligence Research, vol.11, pp.199-240, 1999.
[9] A. Borgida, M. Lenzerini, R. Rosati, "Description logics for data bases", F. Baader, D. Calvanese, et al. (Eds.), The Description Logic Handbook, Cambridge University Press, pp. 472-494, 2002.
[10] F. Baader, Description logics, S. Tessaris et al.(Eds.): Reasoning Web 2009, LNCS 5689, pp.1-39, 2009.
[11] Y. Jiang, Y. Tang, et al., "Extending Soft Sets with Description Logics", Computers and Mathematics with Application, vol.59, pp.2087-2096, 2010.
[12] Dan A. Simovici, Richard L. Tenney, relational database sytems, Academic Press, 1995.

Yuxia Lei received his PhD degree in computer software and theory from the Institute of Computing Technology, CAS, in 2010, his MS degree in mathematics from the Institute of software, CAS, in 2002.

He is an associate professor at school of computer science, Qufu Normal University, China. His current research interests include concept lattice, artificial intelligence and machine learning.

Jingying Tian received her MS degree in architectual engineering from Shandong University of Science and Technology, China, in 2004.

She is a lecture at school of architectual engineering, Rizhao Politechnic. Her research interests include computer applications in the construction.

Baoxiang Cao is a professor at school of computer science, Qufu Normal University, China. His current research interests include software engineering and artificial intelligence.