Unification in the Description Logic $\mathcal{EL}$

Franz Baader and Barbara Morawska

TU Dresden, Germany

UNIF 2009
**UNIF 2008** Unification in $\mathcal{EL}$ is of type zero.

**UNIF 2009** Unification in $\mathcal{EL}$ is decidable and is in NP. Unification problem in $\mathcal{EL}$ is NP-complete.
Outline

1. Introduction
2. $\mathcal{EL}$-unification
3. Towards a decision procedure
   - Reductions and reduced form
   - Subsumption order and its inverse
   - Minimal Unifiers
4. Decision Procedure
   - Computing minimal unifiers
   - Complexity
5. Conclusion
Description Logic $\mathcal{EL}$

- Concept names: City, Cathedral,
- Top concept: $\top$,
- Conjunction: $\sqcap$,
- Existential restriction: $\exists \text{has-location}. \top$

Example (concept term)

$\text{City} \sqcap \exists \text{location. East-South of Germany} \sqcap \exists \text{university.} \top$
Description Logic $\mathcal{EL}$

### Semantics

$(\Delta, \mathcal{I})$ is an interpretation, where:

- Concepts are sets: if $A \in N_C$, $A^\mathcal{I} \subseteq \Delta$;
- Roles are binary relations: if $r \in N_R$, $r^\mathcal{I} \subseteq \Delta \times \Delta$;
- $\top$ is the domain: $\top^\mathcal{I} = \Delta$;
- Conjunction is intersection: $(C \sqcap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$;
- $(\exists r . C)^\mathcal{I} = \{ c \in \Delta \mid \exists b \in \Delta. (c, b) \in r^\mathcal{I} \text{ and } b \in C^\mathcal{I} \}$

### Subsumption and equivalence

- Subsumption:
  $C \sqsubseteq D$ iff for all interpretations $C^\mathcal{I} \subseteq D^\mathcal{I}$.
- Equivalence:
  $C \equiv D$ iff $C \sqsubseteq D$ and $D \sqsubseteq C$.
We define a set of variables $N_V$ as a subset of $N_C$.

Idea: concept names in $N_V$ may be defined differently by different users or developers of a given ontology.

Concepts from $N_V$ can be substituted with concept terms, concepts from $N_C$ cannot be substituted.
Example:

- City \n  \n  \exists\ \text{location. East-South of Germany}\n  \n  \exists\ \text{size. (more-than-500000 \n  less-than-1000000)}

- Settlement \n  \n  \exists\ \text{has. Cathedral}\n  \n  \exists\ \text{location.Saxony}\n  \n  \exists\ \text{size. middle}
**EL-Unification**

**EL-Unification Problem**

is a set of equalities, $C_1 \equiv D_1, \ldots, C_n \equiv D_n$, where $C_i, D_i$ are $\mathcal{EL}$-concept terms.

**A substitution $\sigma$ is an $\mathcal{EL}$-unifier (solution)**

of an $\mathcal{EL}$-unification problem $C_1 \equiv D_1, \ldots, C_n \equiv D_n$ if $\sigma(C_1) \equiv \sigma(D_1), \ldots, \sigma(C_n) \equiv \sigma(D_n)$. 


SLmO – semilattices with monotone operators

\[ SLmO = \{ \begin{align*}
x \land (y \land z) &= (x \land y) \land z, \\
x \land y &= y \land z, \\
x \land x &= x, \\
x \land 1 &= x, \\
\{ f_i(x \land y) \land f_i(y) &= f_i(x \land y) \mid 1 \leq i \leq n \}
\end{align*} \]

- \( \land \) is associative, commutative and idempotent,
- \( \top \) is a unit for \( \land \)
- \( \exists r_i. (C \land D) \land \exists r_i. D \equiv \exists r_i. (C \land D) \)

Existential restriction is not a homomorphism:
\[ \exists r. (A \land B) \nsubseteq \exists r. A \land \exists r. B \]
What are the unifiers of the following goal:

\[ \exists R. Y \sqsubseteq X \]

For example:

- \([X \mapsto \exists R.Z_1, \ Y \mapsto Z_1]\)
- \([X \mapsto \exists R.Z_1 \sqcap \exists R.Z_2, \ Y \mapsto Z_1 \sqcap Z_2]\)
- \([X \mapsto \exists R.Z_1 \sqcap \exists R.Z_2 \sqcap \exists R.Z_3, \ Y \mapsto Z_1 \sqcap Z_2 \sqcap Z_3]\)
- \(\ldots\)
Reduction rules are applied to concept terms modulo $AC$

- $C \sqcap T \rightarrow C$
- $A \sqcap A \rightarrow A$
- if $D \subseteq C$, then $\exists r.D \sqcap \exists r.C \rightarrow \exists r.D$
Equivalence of reduced concepts

Theorem (Küsters)

\[ C \equiv D \quad \text{iff} \quad \hat{C} = \mathcal{A}_C \hat{D} \]

where \( C \rightsquigarrow \hat{C}, \ D \rightsquigarrow \hat{D} \)
Subsumption order: $C_1 > C_2$ iff $C_1 \sqsupseteq C_2$.
Subsumption order is not well founded.

Inverse of subsumption order: $C_1 >_{is} C_2$ iff $C_1 \sqsubseteq C_2$.

**Lemma**

*There is no infinite sequence $C_0, C_1, C_2, \ldots$ of $\mathcal{EL}$-concept terms such that $C_0 \sqsubseteq C_1 \sqsubseteq C_2 \sqsubseteq \ldots$.*
Monotonicity of $>_{is}$

Lemma

$C$ is a reduced concept term and contains $D$,

$$D >_{is} D'$$

Then:

$$C >_{is} C'$$

where $C'$ is obtained from $C$ by replacing an occurrence of $D$ by $D'$.

Proof

Induction on size of $C$.

1. $C = D$, obvious.
2. $C = \exists R.C_1$ and $D$ occurs in $C_1$ (induction).
3. $C = C_1 \sqcap \cdots \sqcap C_n$ and $D$ occurs in $C_i$. 
Monotonicity of $>_i$s

Proof of the case where $C = C_1 \sqcap \cdots \sqcap C_n$ and $D$ occurs in $C_1$.

$$C_1 \sqcap \cdots \sqcap C_n \rightsquigarrow C_1' \sqcap C_2 \sqcap \cdots \sqcap C_n$$

By induction $C_1 >_i{s} C_1'$, i.e. $C_1 \sqsubseteq C_1'$.

and by monotonicity of $\sqsubseteq$:

$$C_1 \sqcap \cdots \sqcap C_n \sqsubseteq C_1' \sqcap C_2 \sqcap \cdots \sqcap C_n$$

Hence

$$C_1 \sqcap \cdots \sqcap C_n \not>_i{s} C_1' \sqcap C_2 \sqcap \cdots \sqcap C_n$$

means $C_1 \sqcap \cdots \sqcap C_n \equiv C_1' \sqcap C_2 \sqcap \cdots \sqcap C_n$

$C_1 \not\equiv C_1'$, there is $i \neq 1$, such that

$$C_1 \sqsubset C_1' \equiv C_i.$$  

But this means that $C_1$ “eats up” $C_i$ in $C$, and thus $C$ is not reduced. Contradiction.
Minimal unifiers

$>_i s$ is well-founded
its multiset extension $>_m$ is well-founded.

$S(\sigma)$ as a multiset of all $\sigma(X), X \in Var(\Gamma)$.

**Definition**

$\sigma > \gamma$ iff $S(\sigma) >_m S(\gamma)$.
$\sigma, \theta$ are ground, reduced unifiers of $\Gamma$.

The ground, reduced unifier $\sigma$ of $\Gamma$ is **minimal** iff there is no unifer $\theta$, such that $\sigma > \theta$.

Obviously, a goal is unifiable iff it has a minimal ground reduced unifier.
Atoms and flat goals

A concept term is an atom iff it is a constant or of form $\exists r. C$.

A flat atom is an atom which is a constant or $\exists r. C$, where $C$ is constant, variable or $\top$.

A goal $\Gamma$ is flat iff it only contains the equations of the form:

- $X \equiv ? C$ with $X$ a variable and $C$ a non-variable flat atom,
- $X_1 \sqcap \ldots \sqcap X_m \equiv ? Y_1 \sqcap \ldots \sqcap Y_n$, where $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are variables.
Atoms of a unifier $\sigma$

$$At(\sigma) = \bigcup_{X \in \text{Var}(\Gamma)} At(\sigma(X))$$

**Definition**

For every concept term $C$, define $At(C)$:

- if $C = \top$, then $At(C) = \emptyset$,
- if $C$ is a constant, then $At(C) = \{C\}$,
- if $C = \exists r.D$, then $At(C) = \{C\} \cup At(D)$,
- if $C = D_1 \cap D_2$, then $At(C) = At(D_1) \cup At(D_2)$. 
Locality of a minimal ground reduced unifier

\( \gamma \) is a minimal reduced ground unifier of \( \Gamma \)

**Lemma**

*If \( C \) is an atom of \( \gamma \),
then there is a non-variable atom \( D \) in \( \Gamma \),
such that \( C \equiv \gamma(D) \)*

**Proof by contradiction.**

**Idea:** If \( C \) is maximal w. r. t. \( \sqsubseteq \) and violates the lemma, we construct a smaller unifier \( \gamma' \) – contradiction.

- \( C \) is a constant \( A \).
- \( C \) is of the form \( \exists r.C_1 \).
Proof of the case where $C$ is of the form $\exists r . C_1$


$D_1, \ldots, D_n$ are all atoms in $\Gamma$, such that 

$C \sqsubseteq \gamma(D_1), \ldots, C \sqsubseteq \gamma(D_n)$. 

$C \sqsubseteq \gamma(D_1) \sqcap \cdots \sqcap \gamma(D_n)$. 

Obtain $\gamma'$ by replacing $C$ with reduced form of $\gamma(D_1) \sqcap \cdots \sqcap \gamma(D_n)$. 

Check if $\gamma'$ is also a unifier of $\Gamma$

- $X \equiv? E$, 
- $X_1 \sqcap \cdots \sqcap X_m \equiv? Y_1 \sqcap \cdots \sqcap Y_n$, 

$\Gamma$
\( \gamma(X_1) \sqcap \cdots \sqcap \gamma(X_m) \equiv \gamma(Y_1) \sqcap \cdots \sqcap \gamma(Y_n) \)
\( \gamma(X_1) \sqcap \cdots \sqcap \gamma(X_m) \leadsto [U]_{AC} \iff \gamma(Y_1) \sqcap \cdots \sqcap \gamma(Y_n) \)

We show that all these reductions are preserved if \( C \) is replaced by reduced \( \gamma(D_1) \sqcap \cdots \sqcap \gamma(D_n) \).

The most interesting reduction is:

\[ \exists r. E_1 \sqcap \exists r. E_2 \leadsto \exists r. E_1 \]

if \( E_1 \sqsubseteq E_2 \)

Assume that \( C \) is in \( E_1 \) and there is \( C' \) in \( E_2 \), such that \( C \sqsubseteq C' \).

- \( C = C' \), (easy, both are replaced by \( \gamma(D_1) \sqcap \cdots \sqcap \gamma(D_n) \)),
- \( C \sqsubseteq C' \)

In the second case \( C' = \top \) or \( C' \) is \( \gamma(D_i) \), and \( \gamma(D_1) \sqcap \cdots \sqcap \gamma(D_n) \sqsubseteq C' \).
Corollary

\(\Gamma\) – a flat goal
\(\gamma\) – minimal reduced ground unifier of \(\Gamma\)
\(X \in \text{Var}(\Gamma)\)

Then \(\gamma(X) = \top\) or there are non-variable atoms \(D_1, \ldots, D_n\) \((n \geq 1)\) of \(\Gamma\) such that \(\gamma(X) \equiv \gamma(D_1) \sqcap \cdots \sqcap \gamma(D_n)\).
Algorithm

1. For each $X$ in $\Gamma$ guess a set $S_X$ of non-variable atoms in $\Gamma$.
2. Define: $X$ depends on $Y$ if $Y$ occurs in $S_X$. 
   Fail if there are circular dependencies in the transitive closure of depends.
3. Define a substitution
   - If $S_X$ is empty, then $\sigma(X) = T$,
   - otherwise, $S_X = \{D_1, \ldots, D_n\}$ and
     $\sigma(X) = \sigma(D_1) \cap \cdots \cap \sigma(D_n)$.
4. Check if $\sigma$ is a unifier of $\Gamma$. 
Theorem

$\mathcal{EL}$-unification is NP-complete.

Proof.

The problem is NP-hard, because $\mathcal{EL}$-matching is NP-hard.

Consider the algorithm:

Present the substitution $\sigma$ as a sequence of equations, a TBox $T_\sigma$. (Hence the definition of $\sigma$ is polynomial)

For each $C \equiv ? D \in \Gamma$, $\sigma(C) \equiv \sigma(D)$ iff $C \equiv_{T_\sigma} D$.

In $\mathcal{EL}$ subsumption (and thus equivalence) modulo acyclic TBoxes is polynomial.

(What is a minimal unifier of the "type-zero" problem?)
Conclusion

We have shown

Unification in $\mathcal{EL}$ is $NP$-complete.

What next?

- Implementation...
- Unification in $\mathcal{EL}$ but without $T$...
- Unification in extensions of $\mathcal{EL}$, e.g. $\forall r. C$. 