Chapter 6

The $\mathcal{EL}$ family

The DL $\mathcal{EL}$ has the constructors

- existential restriction: $\exists r.C$; *no value restriction*
- conjunction: $C \sqcap D$; *no disjunction*
- the top concept: $\top$. *no negation, no* $\bot$

Every $\mathcal{EL}$ concept is satisfiable w.r.t. any $\mathcal{EL}$ TBox *Why?*
and thus satisfiability is not an interesting problem.

Subsumption in $\mathcal{EL}$ is non-trivial, and cannot be reduced to satisfiability in $\mathcal{EL}$.

We show that subsumption w.r.t. general TBoxes in $\mathcal{EL}$ can be decided in polynomial time.

Note: for the dual DL $\mathcal{FL}_0$, which uses $\forall r.C$ in place of $\exists r.C$, subsumption w.r.t. general TBoxes is ExpTime-complete.

6.1 Subsumption in $\mathcal{EL}$

w.r.t. general TBoxes

Without loss of generality we assume that the concepts tested for subsumption are concept names:

Lemma 6.1

Let $\mathcal{T}$ be a general $\mathcal{EL}$ TBox, C, D $\mathcal{EL}$ concepts, and A, B concept names not occurring in $\mathcal{T}$ or C, D. Then

$$\mathcal{T} \models C \sqsubseteq D \text{ iff } \mathcal{T} \cup \{A \sqsubseteq C, D \sqsubseteq B\} \models A \sqsubseteq B.$$

Proof: blackboard.

In addition, we assume that the TBox $\mathcal{T}$ is in normal form, i.e., all GCIs in $\mathcal{T}$ have one of the following forms:

$$\boxed{A \sqsubseteq B, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad A \sqsubseteq \exists r.B, \quad \text{or} \quad \exists r.A \sqsubseteq B,}$$

where A, A_1, A_2, B are concept names or the top concept $\top$ and r is a role name.

Normalisation of an $\mathcal{EL}$ TBox

One can transform a given TBox into a normalised one by applying the following normalisation rules:

$$\begin{array}{lrcl}
\mathsf{NF0} & \widehat{D} \sqsubseteq \widehat{E} & \longrightarrow & \widehat{D} \sqsubseteq A,\ \ A \sqsubseteq \widehat{E} \\
\mathsf{NF1}_r & C \sqcap \widehat{D} \sqsubseteq B & \longrightarrow & \widehat{D} \sqsubseteq A,\ \ C \sqcap A \sqsubseteq B \\
\mathsf{NF1}_\ell & \widehat{D} \sqcap C \sqsubseteq B & \longrightarrow & \widehat{D} \sqsubseteq A,\ \ A \sqcap C \sqsubseteq B \\
\mathsf{NF2} & \exists r.\widehat{D} \sqsubseteq B & \longrightarrow & \widehat{D} \sqsubseteq A,\ \ \exists r.A \sqsubseteq B \\
\mathsf{NF3} & B \sqsubseteq \exists r.\widehat{D} & \longrightarrow & A \sqsubseteq \widehat{D},\ \ B \sqsubseteq \exists r.A \\
\mathsf{NF4} & B \sqsubseteq D \sqcap E & \longrightarrow & B \sqsubseteq D,\ \ B \sqsubseteq E
\end{array}$$

where C, D, E denote arbitrary $\mathcal{EL}$ concepts,
$\widehat{D}, \widehat{E}$ denote $\mathcal{EL}$ concepts that are neither concept names nor $\top$,
B is a concept name, and
A is a new concept name.

Dresden

Normalisation

example

NF0	$\widehat{D} \sqsubseteq \widehat{E}$	$\longrightarrow$	$\widehat{D} \sqsubseteq A, \quad A \sqsubseteq \widehat{E}$
NF1$_r$	$C \sqcap \widehat{D} \sqsubseteq B$	$\longrightarrow$	$\widehat{D} \sqsubseteq A, \quad C \sqcap A \sqsubseteq B$
NF1$_\ell$	$\widehat{D} \sqcap C \sqsubseteq B$	$\longrightarrow$	$\widehat{D} \sqsubseteq A, \quad A \sqcap C \sqsubseteq B$
NF2	$\exists r.\widehat{D} \sqsubseteq B$	$\longrightarrow$	$\widehat{D} \sqsubseteq A, \quad \exists r.A \sqsubseteq B$
NF3	$B \sqsubseteq \exists r.\widehat{D}$	$\longrightarrow$	$A \sqsubseteq \widehat{D}, \quad B \sqsubseteq \exists r.A$
NF4	$B \sqsubseteq D \sqcap E$	$\longrightarrow$	$B \sqsubseteq D, \quad B \sqsubseteq E$

$$\exists r.A \sqcap \exists r.\exists s.A \sqsubseteq A \sqcap B \quad \rightsquigarrow_{\mathsf{NF0}} \quad \exists r.A \sqcap \exists r.\exists s.A \sqsubseteq B_0, \quad B_0 \sqsubseteq A \sqcap B,$$

$$\exists r.A \sqcap \exists r.\exists s.A \sqsubseteq B_0 \quad \rightsquigarrow_{\mathsf{NF1}_\ell} \quad \exists r.A \sqsubseteq B_1, \quad B_1 \sqcap \exists r.\exists s.A \sqsubseteq B_0,$$

$$B_1 \sqcap \exists r.\exists s.A \sqsubseteq B_0 \quad \rightsquigarrow_{\mathsf{NF1}_r} \quad \exists r.\exists s.A \sqsubseteq B_2, \quad B_1 \sqcap B_2 \sqsubseteq B_0,$$

$$\exists r.\exists s.A \sqsubseteq B_2 \quad \rightsquigarrow_{\mathsf{NF2}} \quad \exists s.A \sqsubseteq B_3, \quad \exists r.B_3 \sqsubseteq B_2,$$

$$B_0 \sqsubseteq A \sqcap B \quad \rightsquigarrow_{\mathsf{NF4}} \quad B_0 \sqsubseteq A, \quad B_0 \sqsubseteq B.$$

Normalisation

terminates

Lemma 6.2

Any $\mathcal{EL}$ TBox $\mathcal{T}$ can be transformed into a normalised $\mathcal{EL}$ TBox $\mathcal{T}'$ by a linear number of applications of the normalisation rules.

In addition, the size of the resulting TBox $\mathcal{T}'$ is linear in the size of $\mathcal{T}$.

Proof: Show that the abnormality degree of a TBox decreases with each rule application

Normalisation terminates

Abnormal occurrence of a concept $\widehat{D}$ within a general $\mathcal{EL}$ TBox :

(i) $\widehat{D}$ is the left-hand side of a GCI $\widehat{D} \sqsubseteq \widehat{E}$ where $\widehat{D}$, $\widehat{E}$ are neither concept names nor $\top$; or

(ii) $\widehat{D}$ is neither concept name nor $\top$, and this occurrence is under a conjunction or an existential restriction operator; or

(iii) the occurrence of $\widehat{D}$ is under a conjunction operator on the right-hand side of a GCI.

The abnormality degree of a general $\mathcal{EL}$ TBox is the number of abnormal occurrences of a concept in this TBox:

- the abnormality degree of a TBox is bounded by the size of the TBox,
- a TBox with abnormality degree 0 is normalised.

Proof continued on blackboard.

Dresden

Normalisation

correctness

original TBox $\mathcal{T}$ ⟺ normalised TBox $\mathcal{T}'$

appropriate semantic relationship

subsumption hierarchy for the concept names occurring in $\mathcal{T}$ ← classification of $\mathcal{T}'$

yields

Note:

$\mathcal{T}$ and $\mathcal{T}'$ are not equivalent in the sense that they have the same models due to the introduction of new concept names by the normalisation rules.

However, $\mathcal{T}'$ is a conservative extension of $\mathcal{T}$.

Conservative extension

definition

Definition 6.3

For a given general $\mathcal{EL}$ TBox $\mathcal{T}_0$, its signature $sig(\mathcal{T}_0)$ consists of the concept and role names occurring in the GCIs of $\mathcal{T}_0$.

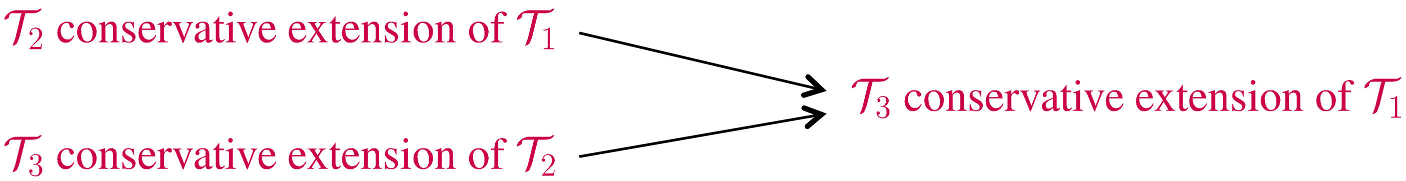
Given general $\mathcal{EL}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, we say that $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ if

- $sig(\mathcal{T}_1) \subseteq sig(\mathcal{T}_2)$,
- every model of $\mathcal{T}_2$ is a model of $\mathcal{T}_1$, and
- for every model $\mathcal{I}_1$ of $\mathcal{T}_1$ there exists a model $\mathcal{I}_2$ of $\mathcal{T}_2$ such that $\mathcal{I}_1$ and $\mathcal{I}_2$ coincide on $sig(\mathcal{T}_1) \cup \{\top\}$, i.e.,
  - $\Delta^{\mathcal{I}_1} = \Delta^{\mathcal{I}_2}$,
  - $A^{\mathcal{I}_1} = A^{\mathcal{I}_2}$ for all concept names $A \in sig(\mathcal{T}_1)$, and
  - $r^{\mathcal{I}_1} = r^{\mathcal{I}_2}$ for all role names $r \in sig(\mathcal{T}_1)$.

Conservative extension

properties

The notion of a conservative extensions is transitive:



$\mathcal{T}_2$ conservative extension of $\mathcal{T}_1$

$\mathcal{T}_3$ conservative extension of $\mathcal{T}_2$

$\mathcal{T}_3$ conservative extension of $\mathcal{T}_1$

Lemma 6.4

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be general $\mathcal{EL}$ TBoxes such that $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$, and C, D $\mathcal{EL}$ concepts containing only concept and role names from $sig(\mathcal{T}_1)$.

Then $\mathcal{T}_1 \models C \sqsubseteq D$ iff $\mathcal{T}_2 \models C \sqsubseteq D$.

Proof: blackboard.

Conservative extension

application

Proposition 6.5

Assume that $\mathcal{T}_2$ is obtained from $\mathcal{T}_1$ by applying one of the normalisation rules. Then $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$.

Proof: blackboard.

Corollary 6.6

Let $\mathcal{T}$ be a general $\mathcal{EL}$ TBox and $\mathcal{T}'$ the normalised TBox obtained from $\mathcal{T}$ using the normalisation rules, as described in the proof of Lemma 6.2.

Then we have

$$\mathcal{T} \models A \sqsubseteq B \quad \text{iff} \quad \mathcal{T}' \models A \sqsubseteq B$$

for all concept names $A, B \in sig(\mathcal{T})$.

Proof: immediate consequence of Proposition 6.5, transitivity, and Lemma 6.4.

Conservative extension

application

subsumption hierarchy for the concept names occurring in $\mathcal{T}$ $\xleftarrow{\text{yields}}$ classification of $\mathcal{T}'$

Corollary 6.6

Let $\mathcal{T}$ be a general $\mathcal{EL}$ TBox and $\mathcal{T}'$ the normalised TBox obtained from $\mathcal{T}$ using the normalisation rules, as described in the proof of Lemma 6.2.

Then we have

$$\mathcal{T} \models A \sqsubseteq B \quad \text{iff} \quad \mathcal{T}' \models A \sqsubseteq B$$

for all concept names $A, B \in sig(\mathcal{T})$.

Proof: immediate consequence of Proposition 6.5, transitivity, and Lemma 6.4.

Classification procedure for $\mathcal{EL}$

We assume that the input TBox $\mathcal{T}$ is a general $\mathcal{EL}$ TBox in normal form.

The procedure starts with the GCIs in $\mathcal{T}$ and adds implied GCIs using appropriate inference rules.

All the GCIs generated in this way are of a specific form:

Definition 6.7

A $\mathcal{T}$-sequent is a GCI of the form

$$A \sqsubseteq B, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad A \sqsubseteq \exists r.B, \quad \text{or} \quad \exists r.A \sqsubseteq B,$$

where A, A_1, A_2, B are concept names in $sig(\mathcal{T})$ or the top concept $\top$, and r is a role name in $sig(\mathcal{T})$.

Note:

- The overall number of $\mathcal{T}$-sequents is polynomial in the size of $\mathcal{T}$.
- Every GCI in $\mathcal{T}$ is a $\mathcal{T}$-sequent.
- A set of $\mathcal{T}$-sequents consists of GCIs, and thus is a TBox.

Classification rules

for $\mathcal{EL}$

CR1 $\dfrac{}{A \sqsubseteq A}$ CR2 $\dfrac{}{A \sqsubseteq \top}$

CR3 $\dfrac{A_1 \sqsubseteq A_2 \quad A_2 \sqsubseteq A_3}{A_1 \sqsubseteq A_3}$ CR4 $\dfrac{A \sqsubseteq A_1 \quad A \sqsubseteq A_2 \quad A_1 \sqcap A_2 \sqsubseteq B}{A \sqsubseteq B}$

CR5 $\dfrac{A \sqsubseteq \exists r.A_1 \quad A_1 \sqsubseteq B_1 \quad \exists r.B_1 \sqsubseteq B}{A \sqsubseteq B}$

The rules given above are, of course, not concrete rules, but rule schemata.

Concrete instance: replace meta-variables A, A_1, A_2, B, B_1 by concrete $\mathcal{EL}$ concepts and meta-variable r by a concrete role name.

Only instantiations are allowed for which all the GCIs occurring in the rule are $\mathcal{T}$-sequents!

Dresden

Classification rules

for $\mathcal{EL}$

CR1 $\dfrac{}{A \sqsubseteq A}$ CR2 $\dfrac{}{A \sqsubseteq \top}$

CR3 $\dfrac{A_1 \sqsubseteq A_2 \quad A_2 \sqsubseteq A_3}{A_1 \sqsubseteq A_3}$ CR4 $\dfrac{A \sqsubseteq A_1 \quad A \sqsubseteq A_2 \quad A_1 \sqcap A_2 \sqsubseteq B}{A \sqsubseteq B}$

CR5 $\dfrac{A \sqsubseteq \exists r.A_1 \quad A_1 \sqsubseteq B_1 \quad \exists r.B_1 \sqsubseteq B}{A \sqsubseteq B}$

Rule application:

if all the $\mathcal{T}$-sequents above the line occur in the current TBox $\mathcal{T}'$,

add the $\mathcal{T}$-sequent below the line to $\mathcal{T}'$

unless it already belongs to $\mathcal{T}'$.

Dresden

Classification rules

for $\mathcal{EL}$

CR1 $$\frac{}{A \sqsubseteq A}$$

CR2 $$\frac{}{A \sqsubseteq \top}$$

CR3 $$\frac{A_1 \sqsubseteq A_2 \quad A_2 \sqsubseteq A_3}{A_1 \sqsubseteq A_3}$$

CR4 $$\frac{A \sqsubseteq A_1 \quad A \sqsubseteq A_2 \quad A_1 \sqcap A_2 \sqsubseteq B}{A \sqsubseteq B}$$

CR5 $$\frac{A \sqsubseteq \exists r.A_1 \quad A_1 \sqsubseteq B_1 \quad \exists r.B_1 \sqsubseteq B}{A \sqsubseteq B}$$

Example 6.8

$$\begin{aligned}\mathcal{T}_1 = \{&A \sqsubseteq \exists r.A,\\ &\exists r.B \sqsubseteq B_1,\\ &\top \sqsubseteq B,\\ &A \sqsubseteq B_2,\\ &B_1 \sqcap B_2 \sqsubseteq C\}.\end{aligned}$$

$$\begin{aligned}\mathcal{T}_2 = \{&A \sqsubseteq \exists r.A,\\ &\exists r.A \sqsubseteq B\}.\end{aligned}$$

Dresden

Classification rules

for $\mathcal{EL}$

CR1 $\dfrac{}{A \sqsubseteq A}$ CR2 $\dfrac{}{A \sqsubseteq \top}$

CR3 $\dfrac{A_1 \sqsubseteq A_2 \quad A_2 \sqsubseteq A_3}{A_1 \sqsubseteq A_3}$ CR4 $\dfrac{A \sqsubseteq A_1 \quad A \sqsubseteq A_2 \quad A_1 \sqcap A_2 \sqsubseteq B}{A \sqsubseteq B}$

CR5 $\dfrac{A \sqsubseteq \exists r.A_1 \quad A_1 \sqsubseteq B_1 \quad \exists r.B_1 \sqsubseteq B}{A \sqsubseteq B}$

Saturation of $\mathcal{T}$:

- apply the classification rules exhaustively to the input TBox $\mathcal{T}$
- the resulting TBox $\mathcal{T}^*$ is called the saturated TBox

Lemma 6.9

The saturated TBox $\mathcal{T}^*$ is uniquely determined by $\mathcal{T}$, and it can be computed by a polynomial number of rule applications.

Proof: blackboard.

Dresden

Classification procedure for $\mathcal{EL}$

To show that polynomial-time saturation of $\mathcal{EL}$ TBoxes yields a polynomial-time classification procedure, it is sufficient to prove the following equivalence:

$$\boxed{\mathcal{T} \models A \sqsubseteq B \text{ iff } A \sqsubseteq B \in \mathcal{T}^*}$$

Soundness of the classification procedure (i.e, the if-direction of the equivalence) is an easy consequence of the next lemma:

Lemma 6.10 (Soundness)

If all the GCIs in $\mathcal{T}'$ follow from $\mathcal{T}$

and

the $\mathcal{T}$-sequents above the line of one of the rules belong to $\mathcal{T}'$

then

the $\mathcal{T}$-sequent below the line also follows from $\mathcal{T}$

Proof: blackboard.

Classification procedure for $\mathcal{EL}$

$$\boxed{\mathcal{T} \models A \sqsubseteq B \text{ iff } A \sqsubseteq B \in \mathcal{T}^*}$$

Completeness: instead of showing the only-if direction of the equivalence directly, we prove its contrapositive:

$$\text{if } A \sqsubseteq B \notin \mathcal{T}^* \text{ then } \mathcal{T} \not\models A \sqsubseteq B.$$

For this purpose, we use $\mathcal{T}^*$ to construct a canonical model of $\mathcal{T}$ that

- does not satisfy the GCI $A \sqsubseteq B$
- in case $A \sqsubseteq B \notin \mathcal{T}^*$.

Dresden

Canonical model

Definition 6.11

Let $\mathcal{T}$ be a general $\mathcal{EL}$ TBox in normal form and $\mathcal{T}^*$ the saturated TBox obtained by exhaustive application of the classification rules.

The canonical interpretation $\mathcal{I}_{\mathcal{T}^*}$ induced by $\mathcal{T}^*$ is defined as follows:

$$\Delta^{\mathcal{I}_{\mathcal{T}^*}} = \{A \mid A \text{ is a concept name in } sig(\mathcal{T})\} \cup \{\top\},$$

$$A^{\mathcal{I}_{\mathcal{T}^*}} = \{B \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid B \sqsubseteq A \in \mathcal{T}^*\} \text{ for all concept names } A \in sig(\mathcal{T}),$$

$$r^{\mathcal{I}_{\mathcal{T}^*}} = \{(A, B) \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \times \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid A \sqsubseteq \exists r.B \in \mathcal{T}^*\} \text{ for all role names } r \in sig(\mathcal{T}).$$

Note:

- By definition, we have $B \in A^{\mathcal{I}_{\mathcal{T}^*}}$ iff $B \sqsubseteq A \in \mathcal{T}^*$ for all concept names $A \in sig(\mathcal{T})$.
- The same is actually true for $A = \top$.

Dresden

Canonical model

Lemma 6.12

The canonical interpretation induced by $\mathcal{T}^*$ is a model of the saturated TBox $\mathcal{T}^*$.

Proof: blackboard.

Lemma 6.13 (Completeness)

Let $\mathcal{T}$ be a general $\mathcal{EL}$ TBox in normal form and $\mathcal{T}^*$ the saturated TBox obtained by exhaustive application of the classification rules. Then

$$\mathcal{T} \models A \sqsubseteq B \text{ implies } A \sqsubseteq B \in \mathcal{T}^*.$$

Proof: blackboard.

Theorem 6.14

Subsumption in $\mathcal{EL}$ w.r.t. general TBoxes is decidable in polynomial time.

Proof: blackboard.

Dresden

6.2 Subsumption in $\mathcal{ELI}$

w.r.t. general TBoxes

Inverse roles: if r is a role, then r^- denotes its inverse

$$(r^-)^\mathcal{I} := \{(e,d) \mid (d,e) \in r^\mathcal{I}\}$$

$\mathcal{I}$

As usual, we will use r^- to denote s if $r = s^-$ for a role name s.

In contrast to the case of $\mathcal{EL}$, subsumption in $\mathcal{ELI}$ w.r.t. general TBoxes is no longer polynomial, but ExpTime-complete.

One reason for the higher complexity of subsumption in $\mathcal{ELI}$ is that it can express a restricted form of value restrictions, and thus comes close to $\mathcal{FL}_0$:

$\exists r^-.C \sqsubseteq D$ has the same models as $C \sqsubseteq \forall r.D$

In the following, we will show the ExpTime-upper bound.

Dresden

Normalisation of an $\mathcal{ELI}$ TBox

We say that the general $\mathcal{ELI}$ TBox $\mathcal{T}$ is in i.normal form (or i.normalised) if all its GCIs are of one of the following forms:

$$A \sqsubseteq B, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad A \sqsubseteq \exists r.B, \quad \text{or} \quad A \sqsubseteq \forall r.B,$$

where A, A_1, A_2, B are concept names or the top-concept $\top$ and r is a role name or the inverse of a role name.

Corollary 6.15

Given a general $\mathcal{ELI}$ TBox $\mathcal{T}$, we can compute in polynomial time an i.normalised $\mathcal{ELI}$ TBox $\mathcal{T}'$ that is a conservative extension of $\mathcal{T}$.

In particular, we have

$$\mathcal{T} \models A \sqsubseteq B \;\text{ iff }\; \mathcal{T}' \models A \sqsubseteq B$$

for all concept names $A, B \in sig(\mathcal{T})$.

Proof: blackboard.

Dresden

Classification procedure for $\mathcal{ELI}$

We assume that the input TBox $\mathcal{T}$ is a general $\mathcal{ELI}$ TBox in i.normal form.

The higher complexity of subsumption in $\mathcal{ELI}$ necessitates the use of an extended notion of sequents:

Definition 6.16

A $\mathcal{T}$-i.sequent is an expression of the form

$$K \sqsubseteq \{A\}, \quad K \sqsubseteq \exists r.K', \quad \text{or} \quad K \sqsubseteq \forall r.\{A\},$$

where K, K' are sets of concept names in $sig(\mathcal{T})$, A is a concept name in $sig(\mathcal{T})$, and r is a role name in $sig(\mathcal{T})$ or the inverse of a role name in $sig(\mathcal{T})$.

Note:

- The overall number of $\mathcal{T}$-i.sequents is exponential in the size of $\mathcal{T}$.
- A set in a $\mathcal{T}$-i.sequent stands for the conjunction of its element.

↑ empty conjunction is $\top$

Dresden

Classification procedure for $\mathcal{ELI}$

We assume that the input TBox $\mathcal{T}$ is a general $\mathcal{ELI}$ TBox in i.normal form.

The higher complexity of subsumption in $\mathcal{ELI}$ necessitates the use of an extended notion of sequents:

Definition 6.16

A $\mathcal{T}$-i.sequent is an expression of the form

$$K \sqsubseteq \{A\}, \quad K \sqsubseteq \exists r.K', \quad \text{or} \quad K \sqsubseteq \forall r.\{A\},$$

where K, K' are sets of concept names in $sig(\mathcal{T})$, A is a concept name in $sig(\mathcal{T})$, and r is a role name in $sig(\mathcal{T})$ or the inverse of a role name in $sig(\mathcal{T})$.

Note:

- The overall number of $\mathcal{T}$-i.sequents is exponential in the size of $\mathcal{T}$.
- A set in a $\mathcal{T}$-i.sequent stands for the conjunction of its element.
- $\mathcal{T}$-i.sequents are GCIs, and a set of $\mathcal{T}$-i.sequents is a general $\mathcal{ELI}$ TBox.
- Every GCI in the i.normalised TBox $\mathcal{T}$ is either equivalent to a $\mathcal{T}$-i.sequent or a tautology, i.e., satisfied in every interpretation.

Dresden

Classification rules for $\mathcal{ELI}$

i.CR1 $\dfrac{}{K \sqsubseteq \{A\}}$ if $A \in K$ and K occurs in $\mathcal{T}'$

i.CR2 $\dfrac{M \sqsubseteq \{B\} \text{ for all } B \in K \quad K \sqsubseteq C}{M \sqsubseteq C}$ if M occurs in $\mathcal{T}'$

i.CR3 $\dfrac{M_2 \sqsubseteq \exists r.M_1 \quad M_1 \sqsubseteq \forall r^-.\{A\}}{M_2 \sqsubseteq \{A\}}$

i.CR4 $\dfrac{M_1 \sqsubseteq \exists r.M_2 \quad M_1 \sqsubseteq \forall r.\{A\}}{M_1 \sqsubseteq \exists r.(M_2 \cup \{A\})}$

The rules given above are, again, not concrete rules, but rule schemata.

Concrete instance: replace K, M, M_1, M_2 by sets of concept names in $sig(\mathcal{T})$,
A by a concept name in $sig(\mathcal{T})$,
r by a role name or inverse of a role name in $sig(\mathcal{T})$,
C by any admissible right-hand side of a $\mathcal{T}$-i.sequent.

Dresden

Classification rules

explanations

$$\text{i.CR1} \quad \frac{}{K \sqsubseteq \{A\}} \quad \text{if } A \in K \text{ and } K \text{ occurs in } \mathcal{T}'$$

$$\text{i.CR2} \quad \frac{M \sqsubseteq \{B\} \text{ for all } B \in K \quad K \sqsubseteq C}{M \sqsubseteq C} \quad \text{if } M \text{ occurs in } \mathcal{T}'$$

In i.CR1, only instantiations are allowed for which K actually occurs explicitly in some $\mathcal{T}$-i.sequent in the current TBox $\mathcal{T}'$.

Reason:

Otherwise, the procedure would always generate an exponential number of $\mathcal{T}$-i.sequents.

The analogous restriction on M in rule i.CR2 is needed in the case where $K = \emptyset$.

condition "$M \sqsubseteq \{B\}$ for all $B \in K$" trivially satisfied for all sets M

Classification rules

explanations

$$\text{i.CR1} \quad \frac{}{K \sqsubseteq \{A\}} \quad \text{if } A \in K \text{ and } K \text{ occurs in } \mathcal{T}'$$

$$\text{i.CR2} \quad \frac{M \sqsubseteq \{B\} \text{ for all } B \in K \quad K \sqsubseteq C}{M \sqsubseteq C} \quad \text{if } M \text{ occurs in } \mathcal{T}'$$

Example 6.17

$$\mathcal{T} = \{A \sqsubseteq B\} \cup \{A_i \sqsubseteq A_i \mid 1 \leq i \leq n\}$$

We have $\mathcal{T} \models M \cup \{A\} \sqsubseteq \{B\}$ for all (exponentially many) sets $\emptyset \neq M \subseteq \{A_1, \ldots, A_n\}$.

None of these $\mathcal{T}$-i.sequents is actually generated by the rules when applied to

$$\mathcal{T}' = \{\{A\} \sqsubseteq \{B\}\} \cup \{\{A_i\} \sqsubseteq \{A_i\} \mid 1 \leq i \leq n\}.$$

Classification rules

explanations

i.CR1 $$\frac{}{K \sqsubseteq \{A\}}$$ if $A \in K$ and K occurs in $\mathcal{T}'$

i.CR2 $$\frac{M \sqsubseteq \{B\} \text{ for all } B \in K \quad K \sqsubseteq C}{M \sqsubseteq C}$$ if M occurs in $\mathcal{T}'$

Example 6.18 (i.CR1 and i.CR2 in action)

$$\mathcal{T} = \{A \sqsubseteq \exists r.(A_1 \sqcap A_2 \sqcap A_3), \exists r.(A_1 \sqcap A_2) \sqsubseteq B\}$$

Blackboard.

Dresden

Classification rules

explanations

i.CR1 $$\frac{}{K \sqsubseteq \{A\}} \quad \text{if } A \in K \text{ and } K \text{ occurs in } \mathcal{T}'$$

i.CR2 $$\frac{M \sqsubseteq \{B\} \text{ for all } B \in K \quad K \sqsubseteq C}{M \sqsubseteq C} \quad \text{if } M \text{ occurs in } \mathcal{T}'$$

i.CR3 $$\frac{M_2 \sqsubseteq \exists r.M_1 \quad M_1 \sqsubseteq \forall r^-.\{A\}}{M_2 \sqsubseteq \{A\}}$$

i.CR4 $$\frac{M_1 \sqsubseteq \exists r.M_2 \quad M_1 \sqsubseteq \forall r.\{A\}}{M_1 \sqsubseteq \exists r.(M_2 \cup \{A\})}$$

Due to the occurrence restrictions, the rules i.CR1 and i.CR2 cannot introduce new sets of concept names into $\mathcal{T}'$.

The same is obviously true (without any restriction) for i.CR3.

In contrast, rule i.CR4 can generate new sets, and thus may cause an exponential blowup.

Classification rules

explanations

i.CR1 $$\frac{}{K \sqsubseteq \{A\}} \quad \text{if } A \in K \text{ and } K \text{ occurs in } \mathcal{T}'$$

i.CR2 $$\frac{M \sqsubseteq \{B\} \text{ for all } B \in K \quad K \sqsubseteq C}{M \sqsubseteq C} \quad \text{if } M \text{ occurs in } \mathcal{T}'$$

i.CR3 $$\frac{M_2 \sqsubseteq \exists r.M_1 \quad M_1 \sqsubseteq \forall r^-.\{A\}}{M_2 \sqsubseteq \{A\}}$$

i.CR4 $$\frac{M_1 \sqsubseteq \exists r.M_2 \quad M_1 \sqsubseteq \forall r.\{A\}}{M_1 \sqsubseteq \exists r.(M_2 \cup \{A\})}$$

Example 6.19 (exponential blowup)

$$\mathcal{T} := \{A \sqsubseteq \exists r.\top\} \cup \{\exists r^-.A \sqsubseteq A_i \mid i = 1, \ldots, n\}$$

i.normalisation: $\mathcal{T}' := \{\{A\} \sqsubseteq \exists r.\emptyset\} \cup \{\{A\} \sqsubseteq \forall r.\{A_i\} \mid i = 1, \ldots, n\}$.

Classification algorithm

i.Saturation of $\mathcal{T}$:

- apply the classification rules exhaustively to the input TBox $\mathcal{T}$
- the resulting TBox $\mathcal{T}^*$ is called the i.saturated TBox

The i.saturated TBox $\mathcal{T}^*$ is again uniquely determined by $\mathcal{T}$.

Proposition 6.20 (soundness and completeness)

For all concept names A, B in $sig(\mathcal{T})$ such that $\{A\}$ occurs in $\mathcal{T}^*$ we have

$$\mathcal{T} \models A \sqsubseteq B \text{ iff } \{A\} \sqsubseteq \{B\} \in \mathcal{T}^*.$$

Condition $\{A\}$ occurs in $\mathcal{T}^*$:

can easily be satisfied by adding $A \sqsubseteq A$ to the input TBox.

$\mathcal{T}$-i.sequent $\{A\} \sqsubseteq \{A\}$

Classification algorithm

soundness

Soundness, i.e. the if direction of Proposition 6.20, is an easy consequence of the next lemma and the fact that any GCI in $\mathcal{T}$ follows from $\mathcal{T}$.

Lemma 6.21 (soundness)

Assume that

- all the GCIs in $\mathcal{T}'$ follow from $\mathcal{T}$ and
- the $\mathcal{T}$-i.sequents above the line of one of the classification rules belong to $\mathcal{T}'$.

Then the $\mathcal{T}$-i.sequent below the line also follows from $\mathcal{T}$.

Proof: blackboard.

Classification algorithm

completeness

To show completeness, i.e. the only-if direction of Proposition 6.20, we construct an appropriate canonical interpretation.

Definition 6.22 (canonical interpretation)

Let $\mathcal{T}$ be a general $\mathcal{ELI}$ TBox in i.normal form and $\mathcal{T}^*$ the i.saturated TBox obtained by exhaustive application of the classification rules.

The canonical interpretation $\mathcal{I}_{\mathcal{T}^*}$ induced by $\mathcal{T}^*$ is defined as follows:

$$\Delta^{\mathcal{I}_{\mathcal{T}^*}} = \{M \mid M \text{ is a set of concept names in } sig(\mathcal{T}) \text{ that occurs in } \mathcal{T}^*\},$$

$$A^{\mathcal{I}_{\mathcal{T}^*}} = \{M \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid M \sqsubseteq \{A\} \in \mathcal{T}^*\},$$

$$\begin{aligned} s^{\mathcal{I}_{\mathcal{T}^*}} = \{&(M,N) \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \times \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid M \sqsubseteq \exists s.N \in \mathcal{T}^* \text{ and } N \text{ is maximal,} \\ &\text{i.e., there is no } N' \supsetneq N \text{ such that } M \sqsubseteq \exists s.N' \in \mathcal{T}^*\} \cup \\ \{&(N,M) \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \times \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid M \sqsubseteq \exists s^-.N \in \mathcal{T}^* \text{ and } N \text{ is maximal,} \\ &\text{i.e., there is no } N' \supsetneq N \text{ such that } M \sqsubseteq \exists s^-.N' \in \mathcal{T}^*\}. \end{aligned}$$

A concept name in $sig(\mathcal{T})$; s role name in $sig(\mathcal{T})$.

Dresden

Classification algorithm

completeness

Lemma 6.23

Let r be a role name or the inverse of a role name. Then

$$\begin{aligned} r^{\mathcal{I}_{\mathcal{T}^*}} = \{(M,N) \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \times \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid M \sqsubseteq \exists r.N \in \mathcal{T}^*,\ N \text{ maximal}\} \cup \\ \{(N,M) \in \Delta^{\mathcal{I}_{\mathcal{T}^*}} \times \Delta^{\mathcal{I}_{\mathcal{T}^*}} \mid M \sqsubseteq \exists r^-.N \in \mathcal{T}^*,\ N \text{ maximal}\}. \end{aligned}$$

Proof: blackboard.

Lemma 6.24

The canonical interpretation induced by $\mathcal{T}^*$ is a model of the i.saturated TBox $\mathcal{T}^*$.

Proof: blackboard.

Dresden

Classification algorithm

completeness

Example (maximality condition needed)

Consider Example 6.19, where all the $\mathcal{T}$-i.sequents

$$\{A\} \sqsubseteq \exists r.M \;\text{ for }\; M \subseteq \{A_1, \ldots, A_n\}$$

belong to $\mathcal{T}^*$.

We have $(\{A\}, \{A_1, \ldots, A_n\}) \in r^{\mathcal{I}}_{\mathcal{T}^*}$,

but $(\{A\}, M) \notin r^{\mathcal{I}}_{\mathcal{T}^*}$ for any strict subset $M \subset \{A_1, \ldots, A_n\}$.

In fact, such a role relationship would violate one of the GCIs $\{A\} \sqsubseteq \forall r.\{A_i\}$.

Dresden

Classification algorithm

Lemma 6.25 (completeness)

Let A, B in $sig(\mathcal{T})$ be such that $\{A\}$ occurs in $\mathcal{T}^*$.
Then $\mathcal{T} \models A \sqsubseteq B$ implies $\{A\} \sqsubseteq \{B\} \in \mathcal{T}^*$.

Proof: blackboard.

Theorem 6.26

Subsumption in $\mathcal{ELI}$ w.r.t. general TBoxes is decidable in exponential time.

Proof: blackboard.

Classification algorithm

for $\mathcal{ELI}$ applied to $\mathcal{EL}$

We can show that the algorithm for $\mathcal{ELI}$ runs in polynomial time if it receives a general $\mathcal{EL}$ TBox as input.

$\mathcal{EL}$-$\mathcal{T}$-i.sequents are $\mathcal{T}$-i.sequents satisfying the following restrictions:

1. the only sets occurring in them are the empty set and singleton sets,
2. value restrictions in these $\mathcal{T}$-i.sequents are only w.r.t. inverses of role names;
3. existential restrictions in these $\mathcal{T}$-i.sequents are only w.r.t. role names.

If we start with an $\mathcal{EL}$ TBox $\mathcal{T}_0$, then the corresponding i.normalised TBox $\mathcal{T}$ (written as a set of $\mathcal{T}$-i.sequents) contains only $\mathcal{EL}$-$\mathcal{T}$-i.sequents.

Classification algorithm

for $\mathcal{ELI}$ applied to $\mathcal{EL}$

Lemma 6.27

There are only polynomially many $\mathcal{EL}$-$\mathcal{T}$-i.sequents in the size of $\mathcal{T}$.

In addition, applying a classification rule for $\mathcal{ELI}$ to a set $\mathcal{T}'$ of $\mathcal{EL}$-$\mathcal{T}$-i.sequents yields a set of $\mathcal{EL}$-$\mathcal{T}$-i.sequents.

Proof: blackboard.

Proposition 6.28

The subsumption algorithm for $\mathcal{ELI}$ yields a polynomial-time decision procedure for subsumption in $\mathcal{EL}$.

Proof: blackboard.

Classification algorithm for $\mathcal{ELI}$ is exponential

In Example 6.19, the i.saturated TBox $\mathcal{T}^*$ contains exponentially many $\mathcal{T}$-i.sequents,

In the following example, one needs to derive exponentially many $\mathcal{T}$-i.sequents before the consequence $\{A\} \sqsubseteq \{B\}$ can be derived.

Example 6.29 (unavoidable exponential blowup)

$$\begin{array}{rcl}
\{A\} & \sqsubseteq & \{\overline{X}_i\} \text{ for } 0 \leq i \leq n-1, \\
\emptyset & \sqsubseteq & \exists r.\emptyset, \\
\{\overline{X}_i, X_0, \ldots, X_{i-1}\} & \sqsubseteq & \forall r.\{X_i\} \text{ for } 0 \leq i \leq n-1, \\
\{X_i, X_0, \ldots, X_{i-1}\} & \sqsubseteq & \forall r.\{\overline{X}_i\} \text{ for } 0 \leq i \leq n-1, \\
\{\overline{X}_i, \overline{X}_j\} & \sqsubseteq & \forall r.\{\overline{X}_i\} \text{ for } 0 \leq j < i \leq n-1, \\
\{X_i, \overline{X}_j\} & \sqsubseteq & \forall r.\{X_i\} \text{ for } 0 \leq j < i \leq n-1, \\
\{X_0, \ldots, X_{n-1}\} & \sqsubseteq & \{B\}, \\
\{B\} & \sqsubseteq & \forall r^-.\{B\}.
\end{array}$$

Dresden