■ Description Logics - Sintassi dei concetti

- $\mathcal{ALC}$

$$C, D \;\rightarrow\; C \sqcap D \mid C \sqcup D \mid \neg C \mid \forall R.C \mid \exists R.C \mid \top \mid \bot \mid A$$

A denota concetti atomici, R denota ruoli o attributi

- Esempi

$$\mathit{Rental} \sqcap \exists \mathit{object}.(\mathit{Car} \sqcap \mathit{Small})$$

$$\mathit{Rental} \sqcap \exists \mathit{object}.(\mathit{Ford} \sqcup \mathit{BMW})$$

$$\mathit{Rental} \sqcap \forall \mathit{object}.\neg \mathit{Ford}$$

$$\mathit{Rental} \sqcap \forall \mathit{object}.\bot$$

■ Semantica dei concetti

- Interpretazioni $\mathcal{I} = \langle\, \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \,\rangle$

- Vincoli su $\cdot^{\mathcal{I}}$

$$
\begin{array}{rcl}
A^{\mathcal{I}} & \subseteq & \Delta^{\mathcal{I}} \\
R^{\mathcal{I}} & \subseteq & \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} & = & C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} & = & C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(\neg C)^{\mathcal{I}} & = & \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} \\
(\forall R.C)^{\mathcal{I}} & = & \text{all } x \text{ such that for all } \langle\, x, y \,\rangle \in R^{\mathcal{I}},\ y \in C^{\mathcal{I}} \\
(\exists R.C)^{\mathcal{I}} & = & \text{all } x \text{ such that for some } \langle\, x, y \,\rangle \in R^{\mathcal{I}},\ y \in C^{\mathcal{I}} \\
\top^{\mathcal{I}} & = & \Delta^{\mathcal{I}} \\
\bot^{\mathcal{I}} & = & \emptyset
\end{array}
$$

■ Sintassi delle asserzioni

- Asserzioni: $C \sqsubseteq D$
  - abbreviazione: $C \equiv D$ per $\{C \sqsubseteq D, D \sqsubseteq C\}$

- TBox - Terminological Box: insieme di asserzioni
  - ontologie

- Esempio:

$$Person \equiv Man \sqcup Woman$$

$$Man \sqcap Woman \sqsubseteq \perp$$

$$\vdots$$

■ Semantica delle asserzioni - Implicazione

Definizione Una interpretazione $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ è un modello di (o soddisfa) $C \sqsubseteq D$ sse $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

Diciamo anche che $C \sqsubseteq D$ è vera in $\mathcal{I}$

Definizione Una interpretazione $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ è un modello di una TBox $\mathcal{T}$ sse $\mathcal{I}$ soddisfa tutte le asserzioni in $\mathcal{T}$

Definizione Una TBox $\mathcal{T}$ implica (entails) $C \sqsubseteq D$ sse ogni modello $\mathcal{I}$ di $\mathcal{T}$ soddisfa $C \sqsubseteq D$. In simboli:

$$C \sqsubseteq_{\mathcal{T}} D$$

Esempio di implicazione

- Sia $\mathcal{T}$

$$Person \equiv Man \sqcup Woman$$

$$Man \sqcap Woman \sqsubseteq \bot$$

allora

$$Person \sqcap \neg Man \sqsubseteq_{\mathcal{T}} Woman$$

mentre non vale

$$Person \sqcap \neg Man \sqsubseteq_{\emptyset} Woman$$

- Il difficile nella scrittura di ontologie sta nel redigere una assiomatizzazione "completa"

■ Estensioni - Ruoli inversi

- Sintassi: R^{-}

- Semantica: $(R^{-})^{\mathcal{I}} = \left\{ \langle\, x, y \,\rangle : \langle\, y, x \,\rangle \in R^{\mathcal{I}} \right\}$

- Esempio: $\exists \mathit{Figlio}^{-}.\mathit{Medico}$ per denotare individui con genitore medico

■ Estensioni - Restrizioni numeriche

- Sintassi: $\exists^{\leq n} R$, $\exists^{\geq n} R$

- Semantica: $(\exists^{\leq n} R)^{\mathcal{I}} = \Big\{ x : \Big| \{ y : \langle\, x, y \,\rangle \in R^{\mathcal{I}} \} \Big| \leq n \Big\}$

- Esempio: $\mathit{Cattolico} \sqsubseteq \exists^{\leq 1} \mathit{Moglie}$

■ Estensioni - Individui (nominals)

- Sintassi: A appartenente ad insieme di simboli distinto
- Vincolo semantico: $\left|A^{\mathcal{I}}\right| = 1$
- Esempio: $Piero \sqsubseteq Person$

■ Forme di ragionamento

Consistenza concetti $C \sqsubseteq_{\mathcal{T}} \bot$?

Sussunzione $C \sqsubseteq_{\emptyset} D$?

Implicazione $C \sqsubseteq_{\mathcal{T}} D$?

Consistenza TBox Esiste un modello di $\mathcal{T}$?

In $\mathcal{ALC}$ tutte equivalenti

Classificazione Da $C_1, \ldots, C_n$ (e $\mathcal{T}$) costruire tassonomia, cioè grafo diretto delle loro relazioni, con archi (C_i, C_j) sse $C_i \sqsubseteq_{\mathcal{T}} C_j$ *e* non c'è alcun altro C_k tale che $C_i \sqsubseteq_{\mathcal{T}} C_k \sqsubseteq_{\mathcal{T}} C_j$

Utile per costruire *indici*

■ Complessità

- Da PTIME a NExpTime

- Utile per valutare ottimalità algoritmi

Complexity of DLs: Overview of the Complexity of Concept Consistency

P	(co-)NP	PSpace	ExpTime	NExpTime
	$\mathcal{ALUN}$ (NP) without $\exists$, only $\neg A$	$\mathcal{ALCN}$ (wrt acyc. TBoxes)	$\mathcal{ALC}_{reg}$ add regular roles $\mathcal{ALC}_{u}$ add universal role $\mathcal{ALC}$ wrt general TBoxes	+ $\mathcal{QI}$ still in ExpTime
$\mathcal{ALN}$ without $\sqcup$	$\mathcal{ALE}$ (co-NP) without $\sqcup$ and NRs, only $\neg A$	$\mathcal{ALCIQ}_{R^+}$		
subsumption of $\mathcal{FL}_0$ $\sqcap$ and $\forall$ only	subsumption of $\mathcal{FL}_0$ (co-NP) wrt acyc. TBoxes	$\mathcal{ALCNO}$ $\mathcal{ALCO}$	$\mathcal{ALCHIQ}_{R^+}$ add role hierarchies	
			$\mathcal{ALCIO}$	$\mathcal{ALCIQO}$
			$\mathcal{ALC}^{\neg}$	$\mathcal{ALC}^{\neg,\cap,\cup}$
		$\mathcal{ALCF}$		$\mathcal{ALCF}$ wrt acyc. TBoxes

$\mathcal{I}$ inverse roles: h-child$^-$
$\mathcal{N}$ NRs: $(\geq n$ h-child$)$
$\mathcal{Q}$ Qual. NRs: $(\geq n$ h-child Blond$)$
$\mathcal{O}$ nominals: "John" is a concept
$\mathcal{F}$ feature chain (dis)agreement
$\cdot_{R^+}$ declare roles as transitive
$\cdot^{\neg,\cap,\cup}$ Boolean ops on roles

Terminological KR and Ontologies

Initial motivation for work on **FaCT** system was **Galen** project

- General requirement for medical terminologies
- Static lists/taxonomies difficult to build and maintain
  - Need to be very **large** and highly interconnected
  - Inevitably contain many **errors** and **omissions**
- Galen project aims to replace static hierarchy with DL
  - **Describe** concepts (e.g., spiral fracture of left femur)
  - Use DL classifier to **build taxonomy**
- Needed expressive DL **and** efficient reasoning
  - Descriptions use transitive/inverse roles, GCIs etc.
  - Very large KBs (tens of thousands of concepts)
    - Even prototype KB is very large (≈3,000 concepts)
    - Existing (incomplete) classifier took **≈24 hours** to classify KB
    - FaCT system (sound and complete) takes **≈60 seconds**

■ Descrizione semantica di servizi

- Per ricerca basata su funzionalità, riconfigurazione automatica, composizione automatica (almeno)

- Le semantiche formali dei linguaggi di programmazione (operazionale, denotazionale, ...) sono generiche risp. ai contenuti delle strutture dati

  - non adatte a ricerca/riconfigurazione/composizione automatiche

- Le logiche descrittive non esprimono direttamente le relazioni tra parametri ed effetto del servizio

  - in effetti un motivo c'è...

■ Service description logics (SDL)

- Estendono description logics con primitive per la descrizione di servizi

- In fase di sviluppo
  - per ora solo servizi funzionali, niente side-effects
  - per ora solo semantica, niente modalità d'uso, stime di costo ed efficacia,...
  - appropriato per retrieval, composition, non per scelte
  - per ora solo definizioni di base e "studio di fattibilità" teorico (decidibilità, complessità)

■ Outline ultima parte

- Requisiti per una SDL
- Sintassi e semantica di $\mathcal{SDL}(\mathcal{X})$
- Risultati di indecidibilità
- Risultati di decidibilità per $\mathcal{X}$-TBoxes
- Conclusioni

■ Minimal requisites for SDL (I)

- Reference example: Map Service
  - given an address, the service returns a map of the surrounding area

- The underlying ontology $\mathcal{O}$ defines *place* with attributes *address* and *map*. The address is a structured object, with attributes *state*, *city*, *street* and *number*

- A SDL should support expressions similar to

$$\{\, \mathit{address.state}, \ldots, \mathit{address.number} \mapsto \mathit{map} \mid \mathit{place} \,\}$$

  called service abstractions, to define mappings using the underlying ontology

■ Minimal requisites for SDL (II)

- If a service $\mathcal{S}$ is not available, support replacement with alternative services $\mathcal{S}'$
- *Service strengthening*: more mappings, i.e., $\mathcal{S}'$ is more general than $\mathcal{S}$
  - e.g., a map service $\mathcal{S}'$ covers more countries than $\mathcal{S}$
- *Service weakening*: less mappings, i.e., $\mathcal{S}'$ is less general than $\mathcal{S}$
  - e.g., a sharper search engine
- A SDL inferences should be able to test both service strengthening and service weakening

■ Minimal requisites for SDL (III)

- Sometimes, $\mathcal{S}$ and its replacement $\mathcal{S}'$ are only required to be comparable on a restricted domain
  - example: $\mathcal{S}'$ covers only european addresses, but this is enough if we are planning a travel to France
  - technically, we need to check whether $\mathcal{S}'$ is a strengthening of the *restriction* of $\mathcal{S}$ to French addresses
- Analogous considerations hold for service range
- SDL should express domain and range restrictions

■ Minimal requisites for SDL (IV)

- Even if $\mathcal{S}$ is not immediately available, an equivalent service can be obtained by composing simpler services
  - e.g., the map service can be reconstructed from a service $\mathcal{S}'$ mapping addresses onto latitude and longitude, plus a service $\mathcal{S}''$ mapping latitude and longitude onto a map
- SDL should model service composition (in a functional sense)
- Composition may require a flexible way of connecting outputs to inputs
- SDL should provide means for restructuring inputs and outputs—say—by replicating them, changing their order (if any), and ignoring them

■ Minimal requisites for SDL (V)

- Services often operate on data sources that do not cover the whole extension of a concept or relation. Together, different services may extend the coverage
  - example: a portal that provides users with a unified view of two map services, one for Europe and one for the US, hiding the details of the composition

- SDL should support a form of service union

■ Minimal requisites for SDL (VI)

- The above constructs are simple and minimal, but powerful enough to capture a number of typical web services
  - library catalogues, wheather reports, express mail delivery status, comparison shopping, bank information, financial and insurance information services,...
- In some applications, it is even possible to model services with side effects (provided that the temporal sequence is irrelevant)
  - plane/train/hotel reservations
- The new constructs are powerful enough to affect decidability...

■ $\mathcal{SDL}(\mathcal{X})$

- A parametric service description logic

- $\mathcal{X}$ is the underlying concept description logic

- New constructs for service description

$$
\begin{array}{rcl}
F, G & \to & \{\, p_1, \ldots, p_k \to q_1, \ldots, q_m \mid C \,\} \mid C_{\to} \mid R_{\to} \mid \\
 & & \mid F \sqcup G \mid F \circ_i G \mid F[i_1, \ldots, i_m] \mid F(i_1, \ldots, i_k/n) \\
 & & F \Downarrow (C_1 * \ldots * C_k \to D_1 * \ldots * D_m) \mid F^* \mid \mathsf{Id}\,.
\end{array}
$$

■ Semantics

$$
\begin{aligned}
\{\vec{p} \to \vec{q} \mid C\}^{\mathcal{I}} &= \left\{\vec{d} \mapsto \vec{d'} \mid \exists e \in C^{\mathcal{I}}.\ \bigwedge_i \langle\, e, d_i \,\rangle \in p_i^{\mathcal{I}} \wedge \bigwedge_j \langle\, e, d'_j \,\rangle \in q_j^{\mathcal{I}}\right\} \\
C_{\to}^{\mathcal{I}} &= \left\{\mapsto d \mid d \in C^{\mathcal{I}}\right\} \\
R_{\to}^{\mathcal{I}} &= \left\{d_1 \mapsto d_2 \mid \langle\, d_1, d_2 \,\rangle \in R^{\mathcal{I}}\right\} \\
(F \sqcup G)^{\mathcal{I}} &= F^{\mathcal{I}} \cup G^{\mathcal{I}} \\
(F \circ_i G)^{\mathcal{I}} &= \left\{d_1, \ldots, d_{i-1}, \vec{d'}, \vec{d''} \mapsto \vec{e} \mid \vec{d'} \mapsto d_i \in G^{\mathcal{I}} \wedge\right. \\
&\qquad \left.\wedge d_1, \ldots, d_{i-1}, d_i, \vec{d''} \mapsto \vec{e} \in F^{\mathcal{I}}\right\} \\
F[i_1, \ldots, i_m]^{\mathcal{I}} &= \left\{\vec{d} \mapsto d'_{i_1}, \ldots, d'_{i_m} \mid \vec{d} \mapsto d'_1, \ldots, d'_n \in F^{\mathcal{I}}\right\} \\
F(i_1, \ldots, i_k/n)^{\mathcal{I}} &= \left\{d_1, \ldots, d_n \mapsto \vec{d'} \mid d_{i_1}, \ldots, d_{i_k} \mapsto \vec{d'} \in F^{\mathcal{I}}\right\} \\
F \Downarrow (C_1 * \ldots * C_k \to C_{k+1} * \ldots * C_m)^{\mathcal{I}} &= \\
&= \{d_1, \ldots, d_k \mapsto d_{k+1}, \ldots, d_m \in F^{\mathcal{I}} \mid d_i \in C_i \ (1 \leq i \leq m)\} \\
(F^*)^{\mathcal{I}} &= \bigcup_{i \geq 0} (F^i)^{\mathcal{I}}, \text{ where } F^0 = \mathrm{Id} \text{ and } F^{i+1} = F \circ F^i
\end{aligned}
$$

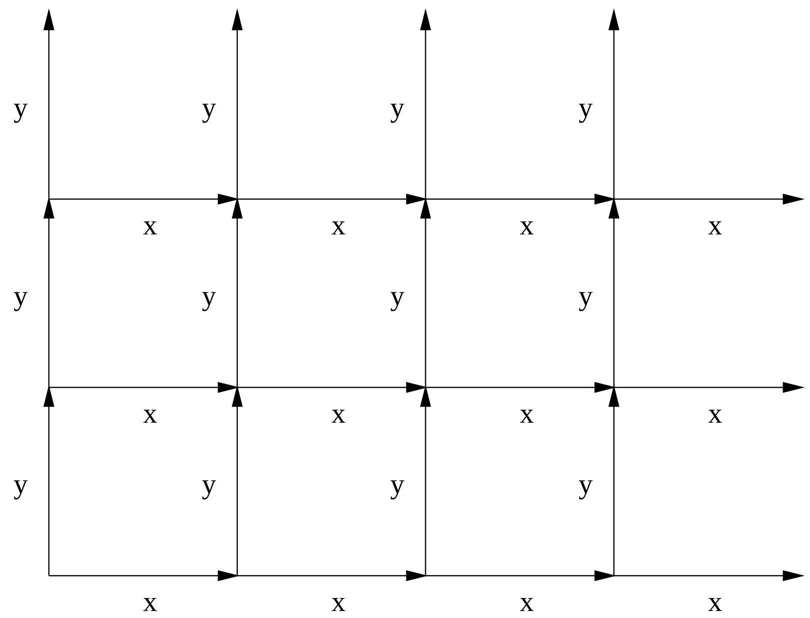
$$\mathsf{Id}^{\mathcal{I}} \;=\; \left\{ d \mapsto d \mid d \in \Delta^{\mathcal{I}} \right\}$$

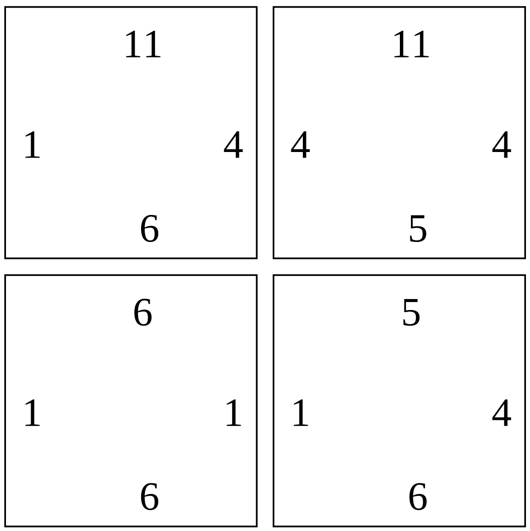
■ Undecidability results

- Checking whether $F \sqsubseteq_{\mathcal{T}} G$, for given F, G and $\mathcal{T}$, is undecidable in $\mathcal{SDL}(\mathcal{ALCI}_f)$

- Proof: by reducing *domino problems* (undecidable) to entailment

■ Domino problems





Domino system $\mathcal{D} = (D, H, V)$, where D is a finite set of *tile types* and $H, V \subseteq D^2$. A mapping $\tau : \mathrm{N}^2 \to D$ is a *solution* of $\mathcal{D}$ if

1. if $\tau(x, y) = d$ and $\tau(x + 1, y) = d'$ then $(d, d') \in H$, and
2. if $\tau(x, y) = d$ and $\tau(x, y + 1) = d'$ then $(d, d') \in V$.

■ Reduction

$$\{\, x.y \to y.x \mid \top \,\} \equiv \mathsf{Id} \tag{1}$$

$$\left\{ x^- \to y \mid \top \right\} \equiv \left\{ y^- \to x \mid \top \right\} \tag{2}$$

$$\top \sqsubseteq \left(\bigsqcup_{d \in \mathcal{D}} D_d \right) \sqcap \bigsqcap_{d \in \mathcal{D}} \; \bigsqcap_{d' \in \mathcal{D} \setminus \{d\}} \neg (D_d \sqcap D_{d'}) \tag{3}$$

$$\bigsqcap_{d \in \mathcal{D}} \left(D_d \Rightarrow \exists x. \bigsqcup_{(d,d') \in H} D_{d'} \right)$$

$$\bigsqcap_{d \in \mathcal{D}} \left(D_d \Rightarrow \exists y. \bigsqcup_{(d,d') \in V} D_{d'} \right)$$

Lemma The following statements are equivalent.

1. $\mathcal{T}_1 = \{(1), (3)\}$ is satisfiable.

2. $\mathcal{T}_2 = \{(2), (3)\}$ is satisfiable.

3. $\mathcal{D}$ has a solution τ.

Decidability results

Theorem Let $\mathcal{T}$ be an $\mathcal{X}$-TBox and F, G be $\mathcal{SDL}(\mathcal{X})$ service expressions with arity (k, m). Let $\vec{s}$ and $\vec{t}$ be two sequences of k and m distinct nominals (respectively), not occurring in F, G, $\mathcal{T}$. Then $F \sqsubseteq_{\mathcal{T}} G$ iff $\mathcal{T} \cup \{t_1 \sqsubseteq \Psi_1^{\vec{s};\vec{t}}(F),\ \Psi_1^{\vec{s};\vec{t}}(G) \sqsubseteq \bot\}$ is not satisfiable.

Lemma Let F be a service expression in $\mathcal{SDL}(\mathcal{X})$ with arity (k, m) and let $1 \leq n \leq m$. Let $\mathcal{I}$ be an interpretation. Let $\vec{C}$ and $\vec{D}$ be two sequences of $\mathcal{X}$-concept expressions with k and m elements, respectively. Then the following statements are equivalent:

1. There exist two sequences of individuals $\vec{d} \in \vec{C}^{\mathcal{I}}$ and $\vec{d'} \in \vec{D}^{\mathcal{I}}$, such that $\vec{d} \mapsto \vec{d'} \in F^{\mathcal{I}}$.

2. There exists an individual d such that $d \in \Psi_n^{\vec{C};\vec{D}}(F)^{\mathcal{I}}$.

■ Translation

$$
\begin{aligned}
\Psi_n^{\vec{C};\vec{D}}(\{\,\vec{p}\to\vec{q}\mid E\,\}) &= D_n \sqcap \exists q_n^-.\left[E \sqcap \Big(\bigsqcap_i \exists p_i.C_i\Big) \sqcap \Big(\bigsqcap_{j\neq n} \exists q_j.D_j\Big)\right]\\
\Psi_1^{;D}(C_\to) &= D \sqcap C\\
\Psi_1^{C;D}(R_\to) &= D \sqcap \exists R^-.C\\
\Psi_n^{\vec{C};\vec{D}}(F \sqcup G) &= \Psi_n^{\vec{C};\vec{D}}(F) \sqcup \Psi_n^{\vec{C};\vec{D}}(G)\\
\Psi_n^{\vec{C};\vec{D}}(F \circ_i G) &= \Psi_n^{\vec{C}';\vec{D}}(F)\,,\ \text{where}
\end{aligned}
$$

$$
C'_j = \begin{cases} C_j & \text{if } j < i\\ \Psi_1^{\vec{C};\top}(G) & \text{if } i = j\\ C_{j+k-1} & \text{if } j > i,\ in(G) = k \end{cases}
$$

$$
\Psi_n^{\vec{C};\vec{D}}(F[i_1,\ldots,i_m]) = \Psi_{i_n}^{\vec{C};\vec{D}'}(F)\,,\quad \text{where}
$$

$$
D'_l = \begin{cases} D_j & \text{if } l = i_j\\ \top & \text{otherwise} \end{cases}
$$

$$
\Psi_n^{\vec{C};\vec{D}}(F(i_1,\ldots,i_k/n)) = \Psi_n^{\vec{C}';\vec{D}}(F)\,,\ \text{where } \vec{C}' = C_{i_1},\ldots,C_{i_k}
$$

$$
\Psi_n^{\vec{C};\vec{D}}(F \Downarrow (E_1 * \ldots * E_k \to E_{k+1} * \ldots * E_m)) = \Psi_n^{\vec{C}';\vec{D}'}(F)\,,
$$

$$
\begin{aligned}
& \text{where } C_i' = C_i \sqcap E_i \text{ and } D_i' = D_i \sqcap E_{i+k-1} \\
\Psi_1^{C;D}(F^*) &= \Psi_1^{C;D}(\mathsf{Id}) \sqcup \Psi_1^{C';D}(F) \\
& \text{where } C' = \mu X. \left(\Psi_1^{C;\top}(\mathsf{Id}) \sqcup \Psi_1^{X;\top}(F)\right) \\
\Psi_1^{C;D}(\mathsf{Id}) &= D \sqcap C
\end{aligned}
$$

■ Main decidability result

Corollary If $\mathcal{X}$ is an extension of $\mathcal{ALCIO}_\mu$ and TBox satisfiability is decidable in $\mathcal{X}$, then entailment from $\mathcal{X}$-TBoxes is decidable in $\mathcal{SDL}(\mathcal{X})$.

In particular, entailment from $\mathcal{X}$-TBoxes in $\mathcal{SDL}(\mathcal{ALCIO}_\mu)$ is decidable.

■ Future work

- Complexity of the basic reasoning problems
- Specialized calculi / engines
- More operators (inverse, parallel composition, general fixpoints), more general TBoxes (acyclic definitions)
- Non-functional semantics
- Usage, cost, quality, ...