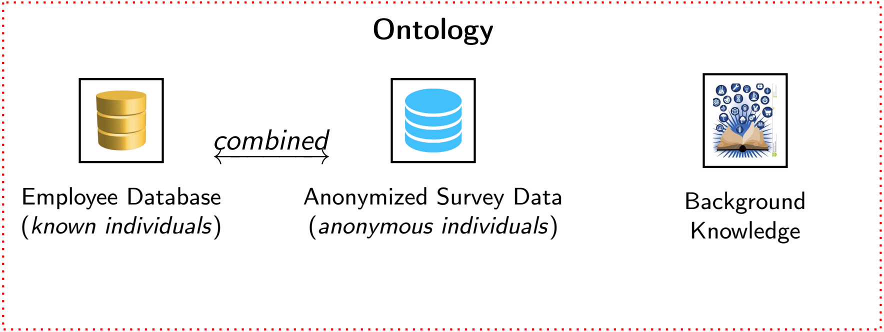
Preliminary Results on The Identity Problem in Description Logic Ontologies

Adrian Nuradiansyah

Franz Baader, Daniel Borchmann
Technische Universität Dresden

July 21, 2017

Identification Problem



Identification Problem

Ontology

combined ⟷

Employee Database
(*known individuals*)

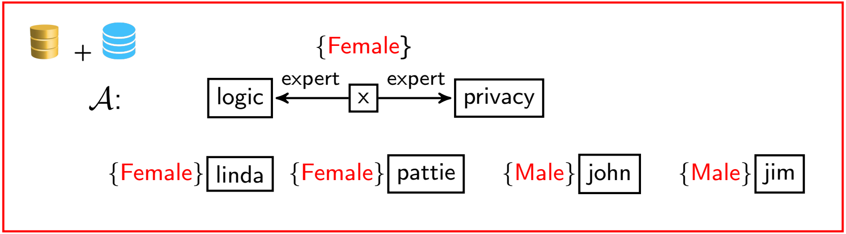
Anonymized Survey Data
(*anonymous individuals*)

Background
Knowledge

known a $\xleftarrow{\text{identity of}}$ **anonymous** x

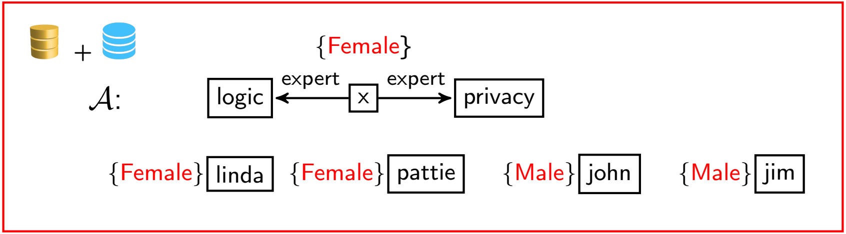
Identification Problem

An attacker still can access some axioms in the ontology s.t. he knows:



Identification Problem

An attacker still can access some axioms in the ontology s.t. he knows:



$\mathcal{T}$:

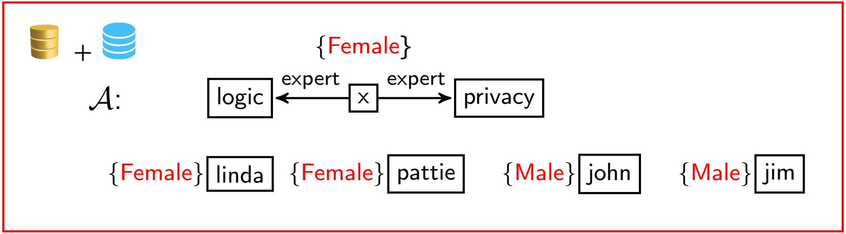
$$Female \sqsubseteq \neg Male$$

$$\exists expert.\{logic\} \sqsubseteq VerTeam \qquad \exists expert.\{privacy\} \sqsubseteq SecTeam$$

$$VerTeam \equiv \{linda, john, pattie\} \qquad SecTeam \equiv \{linda, john, jim\}$$

Identification Problem

An attacker still can access some axioms in the ontology s.t. he knows:



$\mathcal{T}$:

$$Female \sqsubseteq \neg Male$$

$$\exists expert.\{logic\} \sqsubseteq VerTeam \qquad \exists expert.\{privacy\} \sqsubseteq SecTeam$$

$$VerTeam \equiv \{linda, john, pattie\} \qquad SecTeam \equiv \{linda, john, jim\}$$

consequence: $x \doteq$ linda w.r.t. $\mathfrak{O}$

The Identity Problem

Identity Problem

Given $a, b \in N_I$ and an ontology $\mathfrak{O}$. Check whether $a^{\mathcal{I}} = b^{\mathcal{I}}$ for **all models** $\mathcal{I}$ of $\mathfrak{O}$. It is denoted by $(\mathfrak{O} \models a \doteq b)$.

The Identity Problem

Identity Problem

Given $a, b \in N_I$ and an ontology $\mathfrak{O}$. Check whether $a^{\mathcal{I}} = b^{\mathcal{I}}$ for **all models** $\mathcal{I}$ of $\mathfrak{O}$. It is denoted by $(\mathfrak{O} \models a \doteq b)$.

Not all DLs are able to derive equalities between two individuals :(

DLs without Equality Power

Definition

$\mathcal{L}$ is a **DL without equality power** if there are no ontologies $\mathfrak{O}$ formulated in $\mathcal{L}$ and two distinct individuals $a, b, \in N_I$ s.t. $\mathfrak{O} \models a \doteq b$.

DLs without Equality Power

Definition

$\mathcal{L}$ is a **DL without equality power** if there are no ontologies $\mathfrak{O}$ formulated in $\mathcal{L}$ and two distinct individuals $a, b, \in N_I$ s.t. $\mathfrak{O} \models a \doteq b$.

Theorem

Every DL translated to a ***first-order logic without equality predicate*** *is a DL without equality power*

DLs without Equality Power

Definition

$\mathcal{L}$ is a **DL without equality power** if there are no ontologies $\mathfrak{O}$ formulated in $\mathcal{L}$ and two distinct individuals $a, b, \in N_I$ s.t. $\mathfrak{O} \models a \doteq b$.

Theorem

*Every DL translated to a **first-order logic without equality predicate** is a DL without equality power*

They are:

- $\mathcal{ALC}$ and its **fragments**: $\mathcal{EL}, \mathcal{FL}_0, \mathcal{FLE}, \ldots$
- $\mathcal{SRI}$: extending $\mathcal{ALC}$ with **inverse roles**, **role axioms**, **role compositions**, and **transitive roles**.

DLs with Equality Power

- $\mathcal{ALCO}$: lifting up an individual into a concept
  Example: Case of Employee.

DLs with Equality Power

- $\mathcal{ALCO}$: lifting up an individual into a concept
  Example: Case of Employee.

- $\mathcal{ALCQ}$: restricting the number of successors of a domain element
  Example: $\mathfrak{O} = (\{\mathsf{PhDstudent} \sqsubseteq \leq 1 \mathit{supervised}.\top\},$
  $\{\mathit{supervised}(\mathsf{adrian}, \mathsf{y}), \mathit{supervised}(\mathsf{adrian}, \mathsf{franz})\})$

DLs with Equality Power

- $\mathcal{ALCO}$: lifting up an individual into a concept
  Example: Case of Employee.

- $\mathcal{ALCQ}$: restricting the number of successors of a domain element
  Example: $\mathfrak{O} = (\{\text{PhDstudent} \sqsubseteq \leq 1 \textit{supervised}.\top\},$
  $\{\textit{supervised}(\text{adrian}, \text{y}), \textit{supervised}(\text{adrian}, \text{franz})\})$

- $\mathcal{CFD}_{nc}$: featuring functional dependencies
  Functional Dependencies: if two individuals agree on **some attributes**, then they are unique.
  Example: $\mathfrak{O} = (\{A \sqsubseteq A : f \rightarrow id\},$
  $\{\mathbf{A(a)}, \mathbf{A(x)}, f(a) = b, f(x) = b\})$

How to solve the identity problem?

Problem Reduction 1 (*Upper Bound*)

Identity $\underrightarrow{\textit{reduced}}$ Instance **for all DLs with equality power**.

$$\mathfrak{O}_1 \models a \doteq b \text{ iff } (\mathfrak{O}_1 \cup A(a)) \models A(b), \text{ where } A \in N_C \text{ is new}$$

How to solve the identity problem?

Problem Reduction 1 (*Upper Bound*)

Identity $\xrightarrow{\textit{reduced}}$ Instance **for all DLs with equality power**.

$$\mathfrak{O}_1 \models a \doteq b \text{ iff } (\mathfrak{O}_1 \cup A(a)) \models A(b), \text{ where } A \in N_C \text{ is new}$$

Problem Reduction 2 (*Lower Bound*)

Instance $\xrightarrow{\textit{reduced}}$ Identity in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$

HornSAT $\xrightarrow{\textit{reduced}}$ Identity in $\mathcal{CFD}_{nc}$

How to solve the identity problem?

Problem Reduction 1 (*Upper Bound*)

Identity $\xrightarrow{\textit{reduced}}$ Instance **for all DLs with equality power**.

$$\mathfrak{O}_1 \models a \doteq b \text{ iff } (\mathfrak{O}_1 \cup A(a)) \models A(b), \text{ where } A \in N_C \text{ is new}$$

Problem Reduction 2 (*Lower Bound*)

Instance $\xrightarrow{\textit{reduced}}$ Identity in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$

HornSAT $\xrightarrow{\textit{reduced}}$ Identity in $\mathcal{CFD}_{nc}$

Complexity Results

- ExpTime-complete in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$
- NExpTime-complete in $\mathcal{ALCOIQ}$
- PTime-complete in $\mathcal{CFD}_{nc}$

Complexities of identity and **instance** problem are **not the same** in $\mathcal{ALC}^{=}$ allowing $\{a \doteq b \mid a, b \in N_I\} \subseteq \mathcal{A} \rightarrow$ PTime vs **ExpTime-hard**

The View-based Identity Problem

A rôle-based access control scenario:

A partially visible ontology $\mathfrak{O}_I$

combined ⟷

Employee Databases
(*known individuals*)

Anonymized
Survey Data
(*anonymous individuals*)

Background
knowledge

A rôle-based access control scenario:

A partially visible ontology $\mathfrak{O}_I$

Employee Databases
(*known individuals*)

$\xleftrightarrow{\text{combined}}$

Anonymized
Survey Data
(*anonymous individuals*)

Background
knowledge

At **rôle** $\hat{r}_1$
- **queries** through $\mathfrak{O}_{\hat{r}_1} \subseteq \mathfrak{O}_I$
- obtains **View** $V_{\hat{r}_1}$

$\xrightarrow{\text{switch}} \ldots \xrightarrow{\text{switch}}$

At **rôle** $\hat{r}_k$
- **queries** through $\mathfrak{O}_{\hat{r}_k} \subseteq \mathfrak{O}_I$
- obtains **View** $V_{\hat{r}_k}$

At rôle $\hat{r}_{k+1}$, is the identity of an anonymous x hidden w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$?

Query Answering and View

- Let $N_I = N_{KI} \cup N_{AI}$, where N_{KI} and N_{AI} are the sets of **known** and **anonymous** individuals, respectively.

Query Answering and View

- Let $N_I = N_{KI} \cup N_{AI}$, where N_{KI} and N_{AI} are the sets of **known** and **anonymous** individuals, respectively.
- Let $x \in N_{AI}$. The **identity** of x w.r.t. an ontology $\mathfrak{O}_I$ is $idn(x, \mathfrak{O}_I) = \{a \in N_{KI} \mid \mathfrak{O}_I \models x \doteq a\}$

Query Answering and View

- Let $N_I = N_{KI} \cup N_{AI}$, where N_{KI} and N_{AI} are the sets of **known** and **anonymous** individuals, respectively.
- Let $x \in N_{AI}$. The **identity** of x w.r.t. an ontology $\mathfrak{O}_I$ is $idn(x, \mathfrak{O}_I) = \{a \in N_{KI} \mid \mathfrak{O}_I \models x \doteq a\}$
- Given $\mathfrak{O}_I$, $\mathfrak{O}_{\hat{r}} \subseteq \mathfrak{O}_I$ accessed by a user with a rôle $\hat{r}$, and a (*subsumption* or *retrieval*) query q, the **answer** to q w.r.t. $\hat{r}$ is:
  - $ans(q, \hat{r}) := \{\mathtt{true}\}$, if $q = C \sqsubseteq D$ and $\mathfrak{O}_{\hat{r}} \models C \sqsubseteq D$,
  - $ans(q, \hat{r}) := \emptyset$, if $q = C \sqsubseteq D$ and $\mathfrak{O}_{\hat{r}} \not\models C \sqsubseteq D$,
  - $ans(q, \hat{r}) := \{a \in N_I \mid \mathfrak{O}_{\hat{r}} \models C(a)\}$, if $q = C$,
  - $ans(q, \hat{r}) := \{(a, b) \in N_I \times N_I \mid \mathfrak{O}_{\hat{r}} \models r(a, b)\}$, if $q = r \in N_R$.

Query Answering and View

- Let $N_I = N_{KI} \cup N_{AI}$, where N_{KI} and N_{AI} are the sets of **known** and **anonymous** individuals, respectively.
- Let $x \in N_{AI}$. The **identity** of x w.r.t. an ontology $\mathfrak{O}_I$ is $idn(x, \mathfrak{O}_I) = \{a \in N_{KI} \mid \mathfrak{O}_I \models x \doteq a\}$
- Given $\mathfrak{O}_I$, $\mathfrak{O}_{\hat{r}} \subseteq \mathfrak{O}_I$ accessed by a user with a rôle $\hat{r}$, and a (*subsumption* or *retrieval*) query q, the **answer** to q w.r.t. $\hat{r}$ is:
  - $ans(q, \hat{r}) := \{\texttt{true}\}$, if $q = C \sqsubseteq D$ and $\mathfrak{O}_{\hat{r}} \models C \sqsubseteq D$,
  - $ans(q, \hat{r}) := \emptyset$, if $q = C \sqsubseteq D$ and $\mathfrak{O}_{\hat{r}} \not\models C \sqsubseteq D$,
  - $ans(q, \hat{r}) := \{a \in N_I \mid \mathfrak{O}_{\hat{r}} \models C(a)\}$, if $q = C$,
  - $ans(q, \hat{r}) := \{(a, b) \in N_I \times N_I \mid \mathfrak{O}_{\hat{r}} \models r(a, b)\}$, if $q = r \in N_R$.
- Given a rôle $\hat{r}$, a **view** is a total function $V_{\hat{r}} : dom(V_{\hat{r}}) \to 2^{N_I} \cup 2^{N_I \times N_I} \cup \{\{\texttt{true}\}\}$, where
  - **View definition** $dom(V_{\hat{r}})$ is a finite set of queries.
  - $V_{\hat{r}}(q)$ is a finite set of answers for all $q \in dom(V_{\hat{r}})$.

How to solve the View-based Identity Problem?

Canonical Ontology

The **canonical ontology** $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ of $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ is defined as $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) := (\mathcal{T}, \mathcal{A})$ where

$$\begin{aligned}
\mathcal{T} &:= \{C \sqsubseteq D \mid V_{\hat{r}_i}(C \sqsubseteq D) = \{\mathtt{true}\} \text{ for some } i, 1 \leq i \leq k\} \\
\mathcal{A} &:= \{C(a) \mid a \in V_{\hat{r}_i}(C) \text{ for some } i, 1 \leq i \leq k\} \cup \\
& \{r(a,b) \mid (a,b) \in V_{\hat{r}_i}(r) \text{ for some } i, 1 \leq i \leq k\}.
\end{aligned}$$

Hidden Identity

The identity of $x \in N_{AI}$ is **hidden** w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ iff

$$idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})) = \emptyset.$$

Future Work

- **Probabilistic-based Reasoning**
  Two individuals are equal with certain probability.
  Subjective probabilistic in DLs with equality power is more suitable

Future Work

- **Probabilistic-based Reasoning**
  Two individuals are equal with certain probability.
  Subjective probabilistic in DLs with equality power is more suitable

Future Work

- **Probabilistic-based Reasoning**
  Two individuals are equal with certain probability.
  Subjective probabilistic in DLs with equality power is more suitable
- **Anonymizing Description Logic Ontologies**
  Generalizing concepts/nominals on the right hand side of GCIs as specific as possible.

Thank You