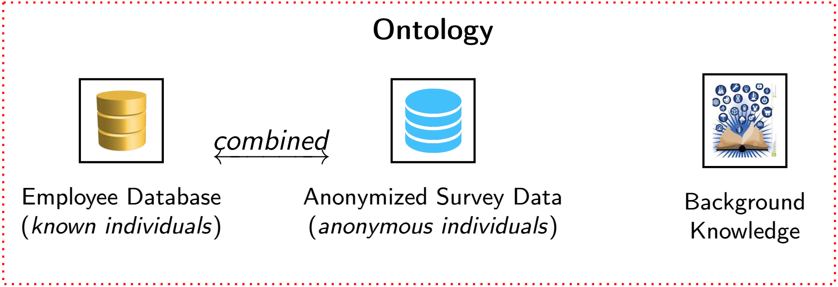
The Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding

Adrian Nuradiansyah

Franz Baader, Daniel Borchmann
Technische Universität Dresden

November 12, 2017

Identity Problem: Motivation



Identity Problem: Motivation

Ontology

combined ⟷

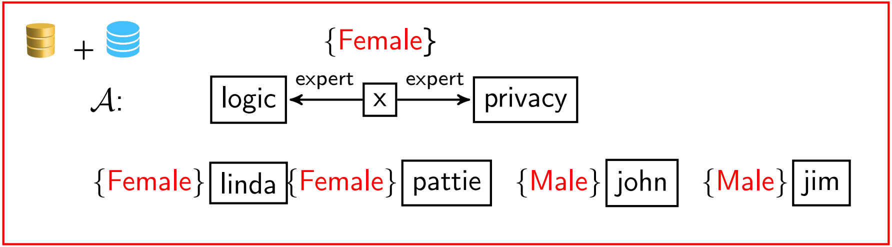
Employee Database (*known individuals*)

Anonymized Survey Data (*anonymous individuals*)

Background Knowledge

known a $\xleftarrow{\text{identical to}}$ **anonymous** x?

Identity Problem: Example



Identity Problem: Example

+

$\mathcal{A}$:

{Female}

logic ←expert— x —expert→ privacy

{Female} linda {Female} pattie {Male} john {Male} jim

$\mathcal{T}$:

$$Female \sqsubseteq \neg Male$$

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

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

Identity Problem: Example

$+$

{Female}

$\mathcal{A}$: logic ←expert— x —expert→ privacy

{Female} linda {Female} pattie {Male} john {Male} jim

$\mathcal{T}$: $Female \sqsubseteq \neg Male$

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

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

consequence: $x \doteq$ linda w.r.t. $\mathcal{A}$ and $\mathcal{T}$

Description Logics

- Decidable fragments of First Order Logics

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: *Male, Female, VerTeam, SecTeam*, ...
  - N_R: **role names** r: *expert, study*, ...
  - N_I: **individual names** a: *logic, privacy, linda, john, x*, ...

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: *Male, Female, VerTeam, SecTeam, ...*
  - N_R: **role names** r: *expert, study, ...*
  - N_I: **individual names** a: *logic, privacy, linda, john, x, ...*
- $\mathcal{ALC}$-concepts $C, D \rightarrow A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: *Male, Female, VerTeam, SecTeam, . . .*
  - N_R: **role names** r: *expert, study, . . .*
  - N_I: **individual names** a: *logic, privacy, linda, john, x, . . .*
- $\mathcal{ALC}$-concepts $C, D \to A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$
- The formal semantics is introduced by means of an **interpretation** $(\mathcal{I} = \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$
  - $\Delta^{\mathcal{I}}$: a non-empty set of domain elements
  - $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$; $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$; $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: *Male, Female, VerTeam, SecTeam, ...*
  - N_R: **role names** r: *expert, study, ...*
  - N_I: **individual names** a: *logic, privacy, linda, john, x, ...*
- $\mathcal{ALC}$-concepts $C, D \rightarrow A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$
- The formal semantics is introduced by means of an **interpretation** $(\mathcal{I} = \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$
  - $\Delta^{\mathcal{I}}$: a non-empty set of domain elements
  - $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$; $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$; $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- An **ontology** $\mathfrak{O}$ consists of **TBox** $\mathcal{T}$ and **ABox** $\mathcal{A}$.

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: *Male, Female, VerTeam, SecTeam, ...*
  - N_R: **role names** r: *expert, study, ...*
  - N_I: **individual names** a: *logic, privacy, linda, john, x, ...*
- $\mathcal{ALC}$-concepts $C, D \rightarrow A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$
- The formal semantics is introduced by means of an **interpretation** $(\mathcal{I} = \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$
  - $\Delta^{\mathcal{I}}$: a non-empty set of domain elements
  - $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$; $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$; $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- An **ontology** $\mathfrak{O}$ consists of **TBox** $\mathcal{T}$ and **ABox** $\mathcal{A}$.
- A TBox $\mathcal{T}$ is a set of **General Concept Inclusions (GCIs)** $C \sqsubseteq D$

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: *Male, Female, VerTeam, SecTeam, ...*
  - N_R: **role names** r: *expert, study, ...*
  - N_I: **individual names** a: *logic, privacy, linda, john, x, ...*
- $\mathcal{ALC}$-concepts $C, D \rightarrow A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$
- The formal semantics is introduced by means of an **interpretation** $(\mathcal{I} = \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$
  - $\Delta^{\mathcal{I}}$: a non-empty set of domain elements
  - $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$; $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$; $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- An **ontology** $\mathfrak{O}$ consists of **TBox** $\mathcal{T}$ and **ABox** $\mathcal{A}$.
- A TBox $\mathcal{T}$ is a set of **General Concept Inclusions (GCIs)** $C \sqsubseteq D$
- An ABox $\mathcal{A}$ is a set of **concept assertions** $C(a)$ and **role assertions** $r(a, b)$

Description Logics

- Decidable fragments of First Order Logics
- The **basic signatures** are:
  - N_C: **concept names** A: $Male, Female, VerTeam, SecTeam, \ldots$
  - N_R: **role names** r: $expert, study, \ldots$
  - N_I: **individual names** a: $logic, privacy, linda, john, x, \ldots$
- $\mathcal{ALC}$-concepts $C, D \rightarrow A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$
- The formal semantics is introduced by means of an **interpretation** $(\mathcal{I} = \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$
  - $\Delta^{\mathcal{I}}$: a non-empty set of domain elements
  - $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$; $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$; $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- An **ontology** $\mathfrak{O}$ consists of **TBox** $\mathcal{T}$ and **ABox** $\mathcal{A}$.
- A TBox $\mathcal{T}$ is a set of **General Concept Inclusions (GCIs)** $C \sqsubseteq D$
- An ABox $\mathcal{A}$ is a set of **concept assertions** $C(a)$ and **role assertions** $r(a, b)$
- An interpretation $\mathcal{I}$ is a **model** of $\mathfrak{O}$ iff
  - For all GCIs in $\mathcal{T}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
  - For all assertions in $\mathcal{A}$, $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$

Identity Problem: Formal Definition

Identity Problem

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

Identity Problem: Formal Definition

Identity Problem

Given $a, b \in N_I$ and a consistent 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 is no consistent ontology $\mathfrak{O}$ 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 is no consistent ontology $\mathfrak{O}$ formulated in $\mathcal{L}$ and two distinct individuals $a, b, \in N_I$ s.t. $\mathfrak{O} \models a \doteq b$.

Theorem

If a DL can be translated to ***first-order logic without equality predicate****, then it is a DL without equality power.*

DLs without Equality Power

Definition

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

Theorem

If a DL can be translated to ***first-order logic without equality predicate****, then it is a DL without equality power.*

Examples:

- $\mathcal{ALC}$ 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

- $\mathcal{ALCO}$: individual as a singleton concept

DLs with Equality Power

- $\mathcal{ALCO}$: individual as a singleton concept

Example:

$\mathcal{A}$: logic $\xleftarrow{\text{expert}}$ x $\xrightarrow{\text{expert}}$ privacy, with {Female} above x

{Female} linda {Female} pattie {Male} john {Male} jim

$\mathcal{T}$:

$$Female \sqsubseteq \neg Male$$

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

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

consequence: $x \doteq$ linda w.r.t. $\mathcal{A}$ and $\mathcal{T}$

DLs with Equality Power

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

DLs with Equality Power

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

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

How to solve the identity problem?

Rely on the existing **instance checking** algorithm

Problem Reduction 1 (*Upper Bound*)

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

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

How to solve the identity problem?

Rely on the existing **instance checking** algorithm

Problem Reduction 1 (*Upper Bound*)

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

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

Problem Reduction 2 (*Lower Bound*)

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

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

How to solve the identity problem?

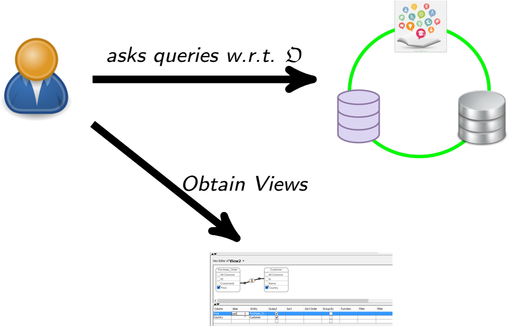
Complexity Results

The identity problem is

- ExpTime-complete in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$
- coNExpTime-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**

View-based Identity Problem



View and Queries

- A view V is a finite collection of queries together with their answers
- Only consider subsumption, instance, and role relationship queries

View-based Identity Problem

Given a partially visible ontology $\mathfrak{O}_I$

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. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$?

Hidden Identity

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

Hidden Identity

- 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 $\mathfrak{O}$ is $idn(x, \mathfrak{O}) = \{a \in N_{KI} \mid \mathfrak{O} \models x \doteq a\}$.

Hidden Identity

- 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 $\mathfrak{O}$ is $idn(x, \mathfrak{O}) = \{a \in N_{KI} \mid \mathfrak{O} \models x \doteq a\}$.

- The identity of $x \in N_{AI}$ is **hidden** w.r.t. $\mathfrak{O}$ iff $idn(x, \mathfrak{O}) = \emptyset$.

Hidden Identity

- 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 $\mathfrak{O}$ is $idn(x, \mathfrak{O}) = \{a \in N_{KI} \mid \mathfrak{O} \models x \doteq a\}$.
- The identity of $x \in N_{AI}$ is **hidden** w.r.t. $\mathfrak{O}$ iff $idn(x, \mathfrak{O}) = \emptyset$.
- The identity of $x \in N_{AI}$ is **hidden** w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ iff

$$\bigcap_{\mathfrak{P} \in \mathtt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})} idn(x, \mathfrak{P}) = \emptyset$$

How to solve the View-based Identity Problem?

Canonical Ontology

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

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

How to solve the View-based Identity Problem?

Canonical Ontology

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

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

Theorem

The identity of $x \in N_{AI}$ is **hidden** w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ iff $idn(x, \mathcal{C}_V) = \emptyset$.

How to solve the View-based Identity Problem?

Canonical Ontology

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

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

Theorem

The identity of $x \in N_{AI}$ is **hidden** w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ iff $idn(x, \mathcal{C}_V) = \emptyset$.

Complexity

- For $\mathcal{L} \in \{\mathcal{ALCO}, \mathcal{ALCQ}\}$, we can check in **exponential time** whether an anonymous individual x is hidden w.r.t. views $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$.
- For $\mathcal{L} \in \{\mathcal{ALCOIQ}\}$, this problem can be solved in **NExpTime**.

Future Work

- Considering the problem of **k-anonymity**
  The identity of anonymous individuals cannot be distinguished from at least k known individuals.

Future Work

- Considering the problem of **k-anonymity**
  The identity of anonymous individuals cannot be distinguished from at least k known individuals.

- **Anonymizing Description Logic Ontologies**
  Idea from Ontology Repair: Weakening the axioms while keeping as much information as possible

Future Work

- Considering the problem of **k-anonymity**
  The identity of anonymous individuals cannot be distinguished from at least k known individuals.
- **Anonymizing Description Logic Ontologies**
  Idea from Ontology Repair: Weakening the axioms while keeping as much information as possible
- **Rôle-based Pseudonymity**
  Implementation on Smarthome Ontologies

Future Work

- Considering the problem of **k-anonymity**
  The identity of anonymous individuals cannot be distinguished from at least k known individuals.
- **Anonymizing Description Logic Ontologies**
  Idea from Ontology Repair: Weakening the axioms while keeping as much information as possible
- **Rôle-based Pseudonymity**
  Implementation on Smarthome Ontologies
- **Probabilistic-based Reasoning**
  Two individuals are equal with certain probability.
  Subjective probabilistic in DLs with equality power is more suitable

Thank You

ROSI