Consultas

Paula Severi

University of Leicester

Facultad de Ingeniería. Universidad de la República, Montevideo, Uruguay. Noviembre-Diciembre 2018.

Resumen de esta clase

1. Consultas a traves de ontologias
2. DL-Lite
3. Consultas conjuntivas
4. Reescritura de consultas

Ontology-mediated querying= Consultas a traves de ontologias

1. Mucha información en la web y otros lados esta incompleta y sin estructura

2. Ontologias resuelven este problema agregando conocimiento sobre el dominio

Servicio de razonamiento

Resolver consultas de una base de datos tomando en cuenta el conocimiento de la ontologia.

Reescritura de consultas

1. Nos restringimos a consultas conjuntivas = el fragmento select-from de SQL

2. dados q consulta conjuntiva y $\mathcal{T}$ Tbox

3. para todo Abox $\mathcal{A}$, q se transforma a $q_{\mathcal{T}}$ tal que las respuestas de q en $\mathcal{A}$ son idénticas a las respuestas dadas por $q_{\mathcal{T}}$ sobre una base de datos que guarda $\mathcal{A}$.

Consultas conjuntivas

Una consulta q tiene la forma

$$\exists x_1 \ldots x_k . \alpha_1 \wedge \ldots \wedge \alpha_n$$

α_i es $A(t)$ o $R(t, t')$

términos t, t' son individuos o variables

variables libres de q son las variables respuesta

q se escribe $q(y_1, \ldots, y_m)$ con $y_1, \ldots, y_m$ libres en q

Respuestas a una consulta

Respuesta sobre una interpretación

$a_1, \ldots, a_m$ es una respuesta a una consulta $q(y_1, \ldots, y_m)$ sobre $\mathcal{I}$ si

$$\mathcal{I} \models q(a_1, \ldots, a_m)$$

Respuesta cierta

$a_1, \ldots, a_m$ es una respuesta cierta a una consulta $q(y_1, \ldots, y_m)$ sobre $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ si

1. todos los individuos $a_1, \ldots, a_m$ ocurren en $\mathcal{A}$
2. $(a_1, \ldots, a_m)$ es una respuesta a q sobre todo modelo $\mathcal{I}$ de $\mathcal{K}$

Ejemplo

$$\mathcal{T} = \{\texttt{Estudiante} \sqsubseteq \exists \texttt{supervisa}^{-}.\texttt{Profesor}\}$$

$$\begin{aligned}\mathcal{A} = \{ \quad & \texttt{sara} : \texttt{Profesor}, \\ & \texttt{marco} : \texttt{Estudiante}, \\ & \texttt{alejandra} : \texttt{Estudiante}, \\ & \texttt{luis} : \texttt{Estudiante}, \\ & (\texttt{sara}, \texttt{marco}) : \texttt{supervisa}, \\ & (\texttt{sara}, \texttt{alejandra}) : \texttt{supervisa}\}\end{aligned}$$

$$q(x) = \exists y.(\texttt{Profesor}(y) \wedge \texttt{supervisa}(y, x) \wedge \texttt{Estudiante}(x))$$

Las respuestas ciertas son `marco`, `alejandra`, `luis`

Ejemplo

$$\mathcal{T} = \{\mathtt{Estudiante} \sqsubseteq \exists\mathtt{supervisa}^{-}.\mathtt{Profesor}\}$$

$$\begin{aligned}\mathcal{A} = \{ \quad & \mathtt{sara} : \mathtt{Profesor}, \\ & \mathtt{marco} : \mathtt{Estudiante}, \\ & \mathtt{alejandra} : \mathtt{Estudiante}, \\ & \mathtt{luis} : \mathtt{Estudiante}, \\ & (\mathtt{sara}, \mathtt{marco}) : \mathtt{supervisa}, \\ & (\mathtt{sara}, \mathtt{alejandra}) : \mathtt{supervisa}\}\end{aligned}$$

$$q(x) = (\mathtt{Profesor}(y) \wedge \mathtt{supervisa}(y, x) \wedge \mathtt{Estudiante}(x))$$

Las respuestas ciertas son $(\mathtt{sara}, \mathtt{marco}) \quad (\mathtt{sara}, \mathtt{alejandra})$

$\mathtt{luis}$ no está en las respuestas porque el supervisor no se conoce explicitamente, tiene diferentes supervisores en distintas interpretaciones

Como acomodar la Tbox sobre el modelo relacional?

Enfoques:

1. materialización
   Abox original = datos en la base de datos
   se extienden con las consecuencias semánticas

2. reescritura de consultas
   anticipamos las consecuencias semanticas en la consulta

Reescritura de consultas

A partir de $\mathcal{T}$ y q construimos una consulta $q_{\mathcal{T}}$ tal que para todo $\mathcal{A}$ las respuestas ciertas a q en $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ son exactamente las respuestas que devolveria un sistema de base de datos al ejecutar $q_{\mathcal{T}}$ tomando $\mathcal{A}$ como datos.

Conjunto de datos en base de datos es una interpretación en DL

Lenguaje al que reescribimos tiene que ser SQL o algo equivalente

Asumimos que el sistema de base de datos puede procesar las consultas como fórmulas de la lógica de primer orden y que además las fórmulas solo contienen predicados unarios o binarios.

Lo que esta en la base de datos se considera como una Abox donde la interpretación $\mathcal{I}_{\mathcal{A}}$ se define

$\Delta^{\mathcal{I}_{\mathcal{A}}}$ son los individuos de la Abox $\mathcal{A}$

$A^{\mathcal{I}_{\mathcal{A}}} = \{a \mid A(a) \in \mathcal{A}\}$

$R^{\mathcal{I}_{\mathcal{A}}} = \{(a, b) \mid R(a, b) \in \mathcal{A}\}$

Asumimos Aboxes simples, i.e. si $a : C \in \mathcal{A}$ entonces C es un nombre de concepto

Reescritura de primer orden

$q_{\mathcal{T}}$ es una reescritura de primer orden de q con respecto a $\mathcal{T}$ si para todo $\mathcal{A}$, el conjunto de las respuestas ciertas de q sobre $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ es el mismo que el conjunto de respuestas de $q_{\mathcal{T}}$ sobre $\mathcal{I}_{\mathcal{A}}$ siempre que $\mathcal{K}$ sea consistente.

Ejemplo

$\mathcal{T} = \{B_1 \sqsubseteq A, B_2 \sqsubseteq A\}$
$q(x) = A(x)$ se reescribe a $q_{\mathcal{T}}(x) = A(x) \vee B_1(x) \vee B_2(x)$

$\mathcal{T} = \{A \sqsubseteq \exists R.A\}$
$q(x) = A(x)$ se reescribe a $q_{\mathcal{T}}(x) = A(x)$

La reescritura no siempre está garantizada

Propiedad

La consulta $q(x) = A(x)$ con respecto a $\exists R.A \sqsubseteq A$ no tiene una reescritura de primer orden.

Idea de la demostración.
$q_{\mathcal{T}}(x)$ tiene que expresar que existe una cadena de elementos de x a uno en A. Esto no es expresable en primer orden.

La Tbox esta en $\mathcal{EL}$ que es bastante inexpresiva...

Esta familia de DL's se introdujo para garantizar la reescritura de consultas.

$$C := \top \mid A \mid \exists R \mid \exists R^{-}$$

Tboxes tienen inclusiones

$C_1 \sqsubseteq C_2 \qquad C_1 \sqsubseteq \neg C_2 \qquad R \sqsubseteq S$

Ejemplo

$$
\begin{array}{ll}
\texttt{Profesor} \sqsubseteq \texttt{Docente} & \texttt{Docente} \sqsubseteq \texttt{Persona} \\
\texttt{Docente} \sqsubseteq \exists \texttt{ensenia} & \texttt{Curso} \sqsubseteq \neg \texttt{Persona} \\
\exists \texttt{enseniacurso}^{-1} \sqsubseteq \texttt{Curso} & \texttt{Curso} \sqsubseteq \exists \texttt{enseniacurso}^{-1} \\
\texttt{enseniacurso} \sqsubseteq \texttt{ensenia} &
\end{array}
$$

$$q(x) = \exists y.\texttt{Persona}(x) \wedge \texttt{ensenia}(x, y) \wedge \texttt{Curso}(y)$$

$$q_{\mathcal{T}}(x) = (\texttt{Docente}(x) \vee \texttt{Profesor}(x) \vee \texttt{Persona}(x)) \wedge$$
$$(\texttt{ensenia}(x, y) \wedge \texttt{Curso}(y) \vee \texttt{enseniacurso}(x, y))$$