

Dotazování do deskripčních logik

Petr Křemen

FEL ČVUT

Co nás čeká

1 Od jazyka deskripční logiky \mathcal{ALC} k OWL(2)-DL

2 Další rozšíření

Konjunktivní dotazy

Typy dotazů

Konjunktivní (ABox) dotazy – dotazy parametrizující individuály.

Metadotazy – dotazy parametrizující navíc i koncepty a role. Existuje několik jazyků pro metadotazování, např. SPARQL-DL, OWL-SAIQL, apod.

Příklad

V jazyce SPARQL-DL lze formulovat dotaz “Nalezni všechny osoby spolu s jejich typem.” :

Type(?x, ?c), SubClassOf(?c, Person)

Konjunktivní dotazy

Konjunktivní ABox dotazy jsou analogií databázových SELECT-PROJECT-JOIN dotazů. Konjunktivní dotaz uvažujeme ve tvaru

$$Q(?x_1, \dots, ?x_D) \leftarrow t_1, \dots, t_T,$$

kde každý t_i je buď tvaru $C(y_k)$, nebo $R(y_k, y_l)$. Každé y_i jsou buď (i) individuály z dotazované ontologie, nebo (ii) proměnné z nové množiny V (my odlišíme proměnné od individuálů prefixem “?”) a C označuje koncept a R roli dotazované ontologie. Dále požadujeme, aby všechny $?x_i$ byly přítomny i v těle dotazu

Příklad

Uvažujme dotaz “Nalezni všechny matky a jejich dcery, mají-li tyto dcery bratra ?” :

$$Q(?x, ?z) \leftarrow \text{Woman}(?x), \text{hasChild}(?x, ?y), \text{hasChild}(?x, ?z), \\ \text{Man}(?y), \text{Woman}(?z)$$

Konjunktivní dotazy – sémantika

- Konjunktivní dotazy tvaru $Q()$ nazýváme *booleovský* – ten ověřuje existenci dotazem popsané relační struktury v každém modelu \mathcal{I} ontologie \mathcal{K} .
- Uvažujme tedy libovolnou interpretaci $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. *Ohodnocení* η je funkce z množiny individuálů a proměnných do $\Delta^{\mathcal{I}}$, která se pro individuály shoduje s \mathcal{I} .
- Potom $\mathcal{I} \models_{\eta} Q()$, pokud
 - $\eta(y_k) \in C^{\mathcal{I}}$ pro každý atom $C(y_k)$ z $Q()$ a
 - $\langle \eta(y_k), \eta(y_l) \rangle \in R^{\mathcal{I}}$ pro každý atom $R(y_k, y_l)$ z $Q()$
- Řekneme, že interpretace \mathcal{I} je modelem $Q()$, pokud $\mathcal{I} \models_{\eta} Q()$ pro nějaké η .
- Dále $\mathcal{K} \models Q()$ ($Q()$ je splnitelný v \mathcal{K}) iff $\mathcal{I} \models Q()$ kdykoliv $\mathcal{I} \models \mathcal{K}$

Konjunktivní dotazy – proměnné

- Dotazy bez proměnných nejsou příliš prakticky zajímavé. Pro dotazy s proměnnými definujeme sémantiku již snadno. N -tice $\langle i_1, \dots, i_n \rangle$ je řešením dotazu $Q(?x_1, \dots, ?x_n)$ v ontologii \mathcal{K} , je-li $\mathcal{K} \models Q'()$, pro booleovský dotaz Q' vzniklý z Q nahrazením všech výskytů $?x_1$ v těle dotazu individuálem i_1 , atd.
- V konjunktivním dotazu vystupují proměnné ve dvou situacích:
 - rozlišené (distinguished) se vyskytují v těle i hlavě dotazu, např. $?x, ?z$ v předchozím příkladě. Ohodnocením těchto proměnných jsou doménové prvky, které jsou nutně interpretací některého individuálu z \mathcal{K} . Právě tento individuál je pak očekávaným vázáním rozlišené proměnné ve výsledku dotazu.
 - nerozlišené (undistinguished) se vyskytují pouze v těle dotazu, např. $?y$ v předchozím příkladě. Jejich ohodnocením jsou libovolné doménové prvky.

Konjunktivní dotazy – příklady

Příklad

Mějme ontologii $\mathcal{K}_4 = (\emptyset, \{(\exists R_1 \cdot C_1)(i), R_2(i, j), C_2(j)\})$.

- Platí $\mathcal{K} \models Q_1()$ pro $Q_1() \leftarrow R_1(?x_1, ?x_2)$?
- Jaká jsou řešení dotazu $Q_2(?x_1) \leftarrow R_1(?x_1, ?x_2)$ pro \mathcal{K} ?
- Jaká jsou řešení dotazu $Q_3(?x_1, ?x_2) \leftarrow R_1(?x_1, ?x_2)$ pro \mathcal{K} ?

Vyhodnocování konjunktivních dotazů v jazyce \mathcal{ALC}

Splnitelnost booleovských dotazů v jazyce \mathcal{ALC}

- Splnitelnost booleovského dotazu $Q()$ ve tvaru kořenového stromu v \mathcal{ALC} lze ověřit pomocí tzv. *zabalovací procedury* (rolling-up technique).
 - Každý atom tvaru $R(y_k, y_l)$ lze nahradit termem $(\exists R \cdot X)(y_k)$, pokud se y_k nevyskytuje v jiném atomu dotazu. X je rovno
 - (i) \top , je-li y_l proměnná,
 - (ii) jinak Y_l , neboli tzv. *reprezentativní koncept* individuálu y_l . Jedná se o nový (v \mathcal{K} ani Q se dosud nevyskytující atomický koncept). Pro každý takový y_l vložíme navíc do ABoxu ontologie \mathcal{K} tvrzení $Y_l(y_l)$.
 - Každý atom tvaru $R(y_k, y_l)$ lze nahradit termem $(\exists R \cdot C)(y_k)$, pokud se y_k vyskytuje v dotazu v pouze jediném dalším atomu tvaru $C(y_l)$.
 - Každé dva atomy $C_1(y_k)$ a $C_2(y_k)$ atomem jediným tvaru $(C_1 \sqcap C_2)(y_k)$.
- Kontrolní otázka: *Napadá vás, pro jaké deskripční logiky nebudeme potřebovat reprezentativní koncept ?*

Splnitelnost booleovských dotazů v jazyce \mathcal{ALC} (2)

... po zabalení dotazu získáme dotaz $Q' \leftarrow C(y)$, který je splněn v \mathcal{K} , iff $Q()$ je splněn v \mathcal{K} :

- Je-li y individuál, potom je $Q'()$ splněn, platí-li $\mathcal{K} \models C(y)$, neboli je-li $\mathcal{K} \cup \{(\neg C)(y)\}$ nekonzistentní
- Je-li y proměnná, potom je $Q'()$ splněn, je-li $\mathcal{K} \cup \{C \sqsubseteq \perp\}$ je nekonzistentní. Proč ?

Příklad

Uvažujme dotaz $Q_4() \leftarrow R_1(?x_1, ?x_2), R_2(?x_1, ?x_3), C_2(?x_3)$. Tento dotaz lze zabalit do dotazu $Q'_4 \leftarrow (\exists R_1 \cdot X_2 \sqcap \exists R_2 \cdot C_2)(?x_1)$. Tento dotaz je splnitelný v ontologii \mathcal{K}_4 , neboť $\mathcal{K}_4 \cup \{(\exists R_1 \cdot X_2 \sqcap \exists R_2 \cdot C_2 \sqsubseteq \perp)\}$ je nekonzistentní.

Splnitelnost booleovských dotazů v jazyce \mathcal{ALC} (2)

... a co můžeme dělat s obecnými dotazy ?

- Uvažujme pouze dotazy, které tvoří “komponentu souvislosti” a obsahují pro nějakou proměnnou y_k , alespoň dva atomy tvaru $R_1(y_1, y_k)$ a $R_2(y_2, y_k)$.
- Kontrolní otázka: *Je to dostatečně obecný případ ? Proč nám stačí uvažování jedné komponenty souvislosti ? Proč nemusíme uvažovat stejný případ pro individuály ?*
- Využijeme TMP logiky \mathcal{ALC} . Každá dvojice atomů $R_1(y_1, y_k)$ a $R_2(y_2, y_k)$ může být splněna pouze tehdy, je-li y_k ohodnocen doménovým prvkem, který je interpretací některého individuálu. Proč ? Stačí nám tedy, abychom y_k postupně nahradili v našem dotazu každým individuálem vyskytujícím se v \mathcal{K} .
- A jak je to v případě \mathcal{SHOIN} a \mathcal{SROIQ} ? Bohužel není známa rozhodovací procedura pro obecné booleovské dotazy ani pro \mathcal{SHOIN} .

Dotazy s rozlišenými proměnnými

Mějme libovolný dotaz $Q(?x_1, \dots, ?x_D)$. Jak jej vyhodnotit ?

- Naivně podle definice: Každou rozlišenou proměnnou x_i se pokusíme postupně nahradit každým individuálem z \mathcal{K} . Řešením budou ty D-tice $\langle i_1, \dots, i_D \rangle$, pro které je booleovský dotaz vzniklý nahrazením každé x_k příslušným i_k splnitelný.
- Trochu sofistikovaněji: Zkusíme nahradit nejprve první proměnnou x_1 každým individuálem z \mathcal{K} . Pokud potom alespoň jeden atom bez proměnných ve vzniklém dotazu není logickým důsledkem \mathcal{K} , potom nemusíme testovat vázání ostatních proměnných.
- ... aj. V této oblasti je známo mnoho dalších heuristik a optimalizací.