Aplikace: Znalostní báze

laboratory Gerstner

Znalostní báze

- je systém, který *dostává fakta o prostředí a dotazy* o něm.
- ***Znalostní báze*** *je agentem ve větším systému*, který obsahuje prostředí (také agent), ***správce*** (agent), popřípadě další agenty.
    - *Správce* ukládá do znalostní báze fakta o prostředí.
    - *Prostředí* používáme jako model vnějšího světa.
    - Chceme, aby *stav prostředí* poskytoval úplný popis (relevantních rysů) vnějšího světa.
- *Lokální stav znalostní báze* popisuje informace, které báze má o světě (posloupnost informací, které se báze dosud dozvěděla)
- *Lokální stav správce* obsahuje to, co „on ví o vnějším světě“ + „informace, které předal ***ZB***“ ...

Tento neformální popis poskytuje ještě hodně volnosti, jak modelovat globální stavy.

V nejjednodušším případě přijímáme tato omezení

• *Vnější svět lze popsat výrokově* pomocí výroků z konečné množiny Φ.

• Vnější svět *je stabilní*, tj. pravdivostní hodnoty výroků popisujících svět se s časem nemění.

• *Správce* má úplnou informaci o vnějším světě.

• Vše, *co je uloženo* ve znalostní bázi *je pravdivé*.

• Nejsou *žádné předběžné (a priori) znalosti* o vnějším světě nebo o tom, co bude do znalostní báze uloženo.

Pro jednoduchost reprezentace systému předpokládejme, že

• **Vnější svět** lze popsat pravdivostním ohodnocením **α** prvotních formulí z množiny Φ (**α** se v průběhu práce s databází **nemění**!).

• **Stav správce** obsahuje ohodnocení **α** a posloupnost faktů, která do báze až dosud uložil.

• **Lokální stav** znalostní báze obsahuje posloupnost $A_1, \ldots, A_n$ údajů, které dosud byly uloženy (může jít *o **výrokové** nebo o **modální formule***).

• **Globální stav** označujeme $(\alpha, \langle A_1, \ldots, A_n \rangle, .)$

1. Databáze uchovávající **jen výroková tvrzení**

- *Do znalostní báze* se ukládají a jsou dotazovány *jen **informace o vnějším světě vyjádřitelná ve výrokových formulích***, nikoliv fakta o bázi samé.
- Vše, *co je uloženo* ve znalostní bázi, *je pravdivé*.
- Nejsou *žádné předběžné (a priori) znalosti* o vnějším světě nebo o tom, co bude do znalostní báze uloženo.

Tyto předpoklady představují omezující podmínky pro konstrukci odpovídajících Kripkeho struktur.

Jak má báze odpovědět na dotaz *B* ? Možnost a)

Předpokládejme, že v nějakém okamžiku (r,m) je bázi položen dotaz ***B***, kde ***B*** je výroková formule. Protože báze nemá přímý přístup ke všem informacím o stavu prostředí, ***B*** nemůže být interpretováno jako dotaz na stav vnějšího světa, ale **jen na to, co o něm báze ví K_{KB} (*B*)**.

Znalostní báze by měla odpovědět

$$\begin{cases} \text{ANO} & \text{jestliže } (\mathcal{I}^{kb}, r, m) \models K_{KB} B \\ \text{NE} & \text{jestliže } (\mathcal{I}^{kb}, r, m) \models K_{KB} \neg B \\ \text{NEVÍM} & \text{jinak} \end{cases}$$

Možnost b)

Většinou si znalostní báze pamatuje konjunkci toho, co do ní bylo uloženo.

Předpokládejme, že báze je v lokálním stavu $\langle A_1, \ldots, A_k \rangle$, že $\kappa = A_1 \wedge \ldots \wedge A_k$ a znalostní báze ví jenom to, co plyne z κ.

Potom můžeme na dotaz ***B*** odpovědět dvojím způsobem:

$$\text{ANO právě když} \begin{cases} B \text{ je důsledkem } \kappa \\ \text{nebo} \\ K_{KB}B \text{ je důsledkem } K_{KB}\kappa \end{cases}$$

Když se do znalostní báze se ukládají *jen fakta o vnějším světě* (vyjádřitelná ve výrokových formulích) a nikoliv fakta o bázi samé, pak odpovědi na dotazy formulované jako výrokové formule jsou totožné pro obě uvažované možnosti a) i b):

Věta KB1.

Předpokládejme, že do databáze jsou ukládány jen výrokové formule

$$r_{KB}(m) = \langle A_1, \ldots, A_k \rangle, \quad \kappa = A_1 \wedge \ldots \wedge A_k$$

a B je výroková formule.

Potom následující tvrzení jsou **ekvivalentní**

$(i)\ (\mathcal{I}^{kb}, r, m) \models K_{KB}B$

$(ii)\ \kappa \rightarrow B$ je výroková tautologie

$(iii)\ M_n^{rst} \models K_{KB}\kappa \rightarrow K_{KB}B$

Dotazy formulované **NEJEN pomocí výrokových tvrzení**

Jak ma odpovídat na dotazy, které nejsou výrokové?

Uvažujme dotaz $B \equiv (p \to K_{KB}\, p)$ "Je pravda, že pokud platí p, ví to znalostní báze?"

Zde bychom také rádi měli odpovědi na dotaz *B:*

$$\begin{cases} \text{ANO} & \text{jestliže } (\mathcal{I}^{kb}, r, m) \models K_{KB} B \\ \text{NE} & \text{jestliže } (\mathcal{I}^{kb}, r, m) \models K_{KB} \neg B \\ \text{NEVÍM} & \text{jinak} \end{cases}$$

Kdy platí formule $K_{KB}(p \to K_{KB}\, p)$? (1)

Platí (viz **Mod_T8 a, Mod_T8b**), že $K_{KB}\,(p \to K_{KB}\, p)$ je dokazatelné právě když je dokazatelná formule

$K_{KB}\, p \lor K_{KB} \neg p$ (2)

Důkazy následujících vztahů se opírají o výrokovou tautologii:

P-T2 : $(\alpha 1 \to \beta) \to ((\alpha 2 \to \beta) \to ((\alpha 1 \lor \alpha 2) \to \beta))$

Mod_T8a: $\mathbf{K_n}$, $K_i\,(\boldsymbol{p} \to K_i\,\boldsymbol{p}) \mid\!- K_i\, p \lor K_i \neg p$

1. $K_i\, p \lor \neg K_i\, p$ — výroková tautologie $\alpha \lor \neg \alpha$
2. $K_i\, p \to \boldsymbol{K_i\, p \lor K_i \neg p}$ — výroková tautologie $\alpha \to (\alpha \lor \beta)$
3. $\boldsymbol{p} \to K_i\,\boldsymbol{p}$ — **Ax 3** a předpoklad $K_i\,(\boldsymbol{p} \to K_i\,\boldsymbol{p})$
4. $(\boldsymbol{p} \to K_i\,\boldsymbol{p}) \to (\neg K_i\, p \to \neg p)$ — výroková tautologie
5. $\neg K_i\, p \to \neg p$ — Modus Ponens na řádky 3 a 4
6. $\neg K_i\, p \to K_i \neg K_i\, p$ — **Ax 5**
7. $K_i \neg K_i\, p \to K_i \neg p$ — **Mod_T1** pro řádku 5
8. $\neg K_i\, p \to K_i \neg p$ — transitivita implikace pro řádky 5 a 6
9. $(\alpha \to \beta) \to (\gamma \lor \alpha \to \gamma \lor \beta)$ — výroková tautologie
10. $\neg K_i\, p \to \boldsymbol{K_i\, p \lor K_i \neg p}$ — Modus Ponens na řádky 8 a 9 pro γ rovno $K_i\, p$
11. P-T2
12. $(K_i\, p \lor \neg K_i\, p) \to (\boldsymbol{K_i\, p \lor K_i \neg p})$ — Modus Ponens na řádky 11, 2 a 10
13. $(K_i\, p \lor K_i \neg p)$ — Modus Ponens na řádky 12 a 1

Mod_T8b: **K_n** ,($K_i p \vee K_i \neg p$) |- $K_i(p \to K_i p)$

1. $K_i p \to (p \to K_i p)$ výrokový axiom
2. $K_i K_i p \to K_i(p \to K_i p)$ **Mod_T1** pro řádku 1
3. $K_i p \to K_i K_i p$ **Ax3**
4. $K_i p \to K_i(p \to K_i p)$ transitivita implikace pro řádky 3 a 2
5. $\neg p \to (p \to K_i p)$ **výroková tautologie**
6. $K_i \neg p \to K_i(p \to K_i p)$ **Mod_T1** pro řádku 5
7. P-T2
8. $(K_i p \vee K_i \neg p) \to K_i(p \to K_i p)$ Modus ponens 2x pro řádky 7, 1 a 6
9. $K_i(p \to K_i p)$ Modus ponens pro ř.8 a předpoklad

Odpověď na dotaz ***B*= (*p* →K_{KB}*p*)** bude díky ekvivalentnímu vyjádření **K_{KB} (*p* →K_{KB}*p*)** = $K_{KB}p \vee K_{KB} \neg p$

ANO, pokud p nebo $\neg p$ plyne z toho, co bylo do báze uloženo a

NEVÍM jinak.

Pozor! Odpověď NE není možná, neboť pro dotaz ***B***

$$B \equiv (p \to K_{KB}\, p)$$

platí

$$K_{KB} \neg B \quad \leftrightarrow \quad K_{KB}(p \wedge \neg K_{KB}\, p)$$

Ovšem tato formule podle **Mod_T7** (viz nasl.str.) nemůže být dokazatelná!

$Ax3: K_i A \rightarrow A$

Mod_T7: V (K_n + Ax3) nemůže být dokazatelné $K_i \neg (\boldsymbol{p} \rightarrow K_i \boldsymbol{p})$

Kdyby $K_i \neg (\boldsymbol{p} \rightarrow K_i \boldsymbol{p})$ bylo dokazatelné, bylo by podle **Mod_T6 b** dokazatelné i $K_i \boldsymbol{p} \wedge K_i (\neg K_i \boldsymbol{p})$. Předpokládejme, že tomu tak je:

1. $K_i \boldsymbol{p} \wedge K_i (\neg K_i \boldsymbol{p})$ [předpoklad]
2. $K_i \boldsymbol{p}$ [řádek 1 a vlastnost konjunkce $(\alpha \wedge \beta) \rightarrow \alpha$]
3. $K_i (\neg K_i \boldsymbol{p})$ [řádek 1 a vlastnost konjunkce $(\alpha \wedge \beta) \rightarrow \beta$]
4. $K_i (\neg K_i \boldsymbol{p}) \rightarrow \neg K_i \boldsymbol{p}$ **[Ax3]**
5. $\neg K_i \boldsymbol{p}$ [řádky 4, 3 a Modus ponens]
6. ***false*** [definice ***false*** a řádky 2 a 5]

Vzhledem k tomu, že systém axiomů (K_n + Ax3) není sporný, nemůže v něm být dokazatelná formule ***false***.

Tedy nemůže platit předpoklad „formule $K_i \neg (\boldsymbol{p} \rightarrow K_i \boldsymbol{p})$ je dokazatelná“**.**

Má smysl vkládat do znalostní báze informace, které nejsou reprezentovány výrokovými formulemi?

Například předpokládejme, že do znalostní báze, jejíž veškeré informace reprezentuje posloupnost formulí $<F_1, \ldots, F_i>$ je vložena další informace $\boldsymbol{F_{i+1}} = (\boldsymbol{p} \rightarrow \boldsymbol{K_{KB} p})$, která říká „je-li pravdivé *p*, potom báze o tom ví“.

Taková informace může být velmi užitečná, pokud báze umí ověřit, co ví a co neví. Pokud například $<F_1, \ldots, F_i> \vdash \neg K_{KB} p$, pak z informace $\boldsymbol{F_{i+1}}$ může *DB* odvodit, že *p* není pravdivé.

Tento příklad ukazuje, že když báze získává informace o své znalosti vnějšího světa, pak ona sama **může** prostřednictvím introspekce **odvodit důsledky o vnějším světě**.

2. Databáze uchovávající **nejen výroková tvrzení**

Jsou-li bázi dávána **tvrzení, která nejsou výroková**, pak už znalosti databáze nemůžeme reprezentovat jako ***konjunkcí tvrzení***, která do ní byla vložena (jak jsme činili ve výrokovém případě).

Příklad (viz další strana): Můžeme totiž do báze vložit fakt, který

- byl pravdivý v okamžiku, kdy byl vložen,
- ale nezůstává platný v žádném dalším časovém bodě.

Za těchto okolností totiž znalost nemůže být popsána ***konjunkcí tvrzení***, která do ní byla vložena. ***Proč?*** Tato konjunkce totiž může být ekvivalentní spornému tvrzení a tudíž by ze znalostí databáze plynulo cokoliv!

Příklad: Předpokládejme, že do báze vložíme fakt, který

- je pravdivý v okamžiku, kdy byl vložen,
- ale nezůstává platný v žádném dalším časovém bodu.

Předpokládejme, že primitivní výrok $\boldsymbol{p}$ je ve vnějším světě pravdivý, ale báze zatím o tom (až do taktu $\boldsymbol{j}$) nedostala žádnou informaci. V takové situaci je jistě pravdivá formule

$$\boldsymbol{p} \,\&\, \neg K_{KB}\boldsymbol{p} \qquad (3)$$

Tuto pravdivou formuli tedy může správce v taktu $\boldsymbol{j+1}$ poskytnout znalostní bázi jako novou informaci. Ovšem i když báze získala informaci (3), jistě neplatí, že báze ví (3), tj. (3) platí v každém dalším časovém bodu. Kdyby tomu tak bylo, muselo by totiž platit $K_{KB}(p \,\&\, \neg K_{KB}\, p)$, což není možné, neboť tato formule je ve sporu s S5.

$\boldsymbol{p}$ & $\neg\, K_{KB}\boldsymbol{p}$ (3)

Na příkladu formule (3) jsme se už přesvědčili, že dostane-li báze informaci φ, pak **nemusí nutně platit**, že $K_{KB}\,\varphi$ to ví.

Nicméně, znalostní báze by přece jen měla něco získat vložením informace (3) : měla by vědět, že $\boldsymbol{p}$ je pravdivé, tj. po té, co do báze vložíme (3), by mělo platit $K_{KB}\boldsymbol{p}$.

Jak zajistit to, aby s informacemi tohoto typu znalostní báze zacházela tak, jak odpovídá naší intuici? Určitě znalostní báze *nemůže pracovat stejným způsobem* jako v případě výrokovém. Pokud hodláme připustit, aby do báze byla vkládána fakta obsahující tvrzení o jejích znalostech, je třeba popsat jiný postup, který se bude opírat právě o modální logiku!