

Jaké vlastnosti mají znalosti

(v sémantice možných světů)

VZ 2009



Platí

formule A je **platná** (platná v M) právě když formule $\neg A$ **není splnitelná** (není splnitelná v M).

Připomeňme, že stále předpokládáme, že relace K_i jsou ekvivalence.

Validních formulí je jistě velké množství (všechny výrokové tautologie, ...)

Hledáme nějaký způsob, jak je charakterizovat pomocí syntaktických prostředků.

*Existuje **FORMÁLMÍ SYSTÉM**, který to dokáže?*

VZ 2009



Jeden ze způsobů jak charakterizovat naši interpretaci znalostí je popsat formule, které jsou vždy platné (pravdivé).

Nechť je dána struktura $M = (S, \pi, K_1, \dots, K_n)$. Říkáme, že

(i) *formule A platí (validní) v M ($M \models A$)*, jestliže v každém stavu s je A pravdivá, tj. $(M, s) \models A$.

(ii) *formule A je splnitelná v M* , jestliže $(M, t) \models A$ v nějakém stavu t .

(iii) *formule A je platná (validní)* a píšeme $\models A$, je-li platná (validní) ve všech strukturách.

(iv) *formule A je splnitelná*, je-li splnitelná v nějaké struktuře.

(v) *formule B je logickým důsledkem formule A* , pokud pro každou strukturu M , ve které platí A , vždy platí i B - stručněji: pokud $M \models A$, pak také $M \models B$.

VZ 2009



Při zavedeném pojetí znalostí platí, že agent zná všechny logické důsledky svých znalostí.

Ví-li agent tvrzení A a také ví, že A implikuje B , potom obě formule A a $A \rightarrow B$ jsou pravdivé ve všech světech, které agent považuje za možné,

tedy také B musí být pravdivé ve všech světech, které agent považuje za možné, takže musí vědět (znát) i B .

Odtud dostáváme $\models (K_i A \wedge K_i (A \rightarrow B)) \rightarrow K_i B$

Této formuli se říká **axiom distribuce** nebo také Kripkův axiom a označuje se jako **K**, protože dovoluje distribuovat operátor K_i přes implikaci.

VZ 2009



Druhá vlastnost našeho pojetí znalosti zaručuje, že máme dostatečně silné a schopné agenty: předpokládáme totiž, že každý agent zná všechny formule platné (validní) v dané struktuře.

Je-li A pravdivá ve všech stavech (tj. možných světech) struktury M , pak je pravdivá ve všech těch světech, které agent v daném světě považuje za možné. Tedy:

Pro libovolnou strukturu M platí „je-li $M \models A$, pak $M \models K_i A$ “

Hledáme **formální systém** (soubor axiomů a pravidel) takový, že jej lze použít pro **dokazování** těch formulí, které jsou platné ve všech strukturách.

Příkladem **dokazovacího pravidla** je **pravidlo generalizace znalostí** „z A lze odvodit $K_i A$ “, což se někdy zapisuje jako

$$\frac{A}{K_i A}$$

VZ 2009



Agent nemusí znát všechna fakta, která jsou pravdivá! Ale když agent něco ví, pak to platí:

$$\models K_i A \rightarrow A$$

Tato vlastnost se obvykle nazývá **Axiom znalosti** nebo **Axiom pravdy** a často se označuje **T**.

Odůvodnění axiomu: svět, ve kterém se agent nachází, je vždy takový, který pokládá za možný. Platí-li $K_i A$ v nějakém světě (M, s) , pak A platí ve všech světech, které i považuje za možné, tedy i v (M, s) .

{Filosofové tento axiom považují za hlavní odlišení mezi **znalostí a přesvědčením** (vírou, *belief*). Mohu mít **nepravdivé přesvědčení**, ale nemohu vědět něco, co není pravda!}

VZ 2009



Pozor! Pravidlo generalizace je něco jiného než platnost implikace $A \rightarrow K_i A$

kteřá tvrdí „je-li A pravdivá, potom agent i to ví“: toto **NENÍ validní formule!** Agent nemusí vědět všechny věci, které jsou pravdivé.

{Například při hře *zablácených dětí* může mít jedno z nich *zablácené čelo*, ale nemusí to vědět.}

Agenti znají všechny validní formule, ale nic víc! Jinými slovy znají JEN formule, které jsou *nutně* pravdivé.

Neznají však formule, které jsou (*řízením osudu*) pravdivé jen v nějakém z možných světů.

VZ 2009



V případě, že chceme popisovat to, čemu agent věří, nikoliv to, co ví, nahradíme axiom pravdy

$$\models K_i A \rightarrow A$$

slabším požadavkem:

Axiom konzistence požaduje $\neg K_i \text{false}$, který se často se označuje jako **D**.

VZ 2009



Další dvě vlastnosti se týkají přirozeného požadavku, aby agenti věděli o svých znalostech (pomocí introspekce).

Agenti vědí, co vědí a vědí, co nevědí.

$$\models K_i A \rightarrow K_i K_i A$$

$$\models \neg K_i A \rightarrow K_i \neg K_i A$$

První z těchto vlastností se jmenuje **Axiom pozitivní introspekce** (často označovaný **4**),

druhá **Axiom negativní introspekce** (často označovaný (často označovaný **5**).

VZ 2009



Formální (axiomatický) systém K_n

Axiomy: **A1.** Všechny tautologie výrokové logiky

$$\mathbf{A2.} (K_i \alpha \wedge K_i (\alpha \rightarrow \beta)) \rightarrow K_i (\beta)$$

Odvozovací pravidla:

R1. Z α formulí $\alpha \rightarrow \beta$ odvoďte β (modus ponens)

R2. Z formule α odvoďte $K_i \alpha$ (pravidlo zobecnění znalosti)

V K_n lze dokázat například tvrzení $K_i (\alpha \wedge \beta) \rightarrow K_i \alpha$ (viz dále), což se značí: $K_n \vdash K_i (\alpha \wedge \beta) \rightarrow K_i \alpha$

Jaké vlastnosti má systém K_n ? Jaký je vztah mezi formulemi, které jsou dokazatelné a těmi, které jsou platné v množině všech Kr. struktur s n agenty?

Formální systém je **korektní**, pokud každá dokazatelná formule je i platná (jinými slovy: „pro libovolnou formuli A z $\vdash A$ plyne $\models A$ “).

Form.systém je **úplný**, pokud „pro lib.formuli A z $\models A$ plyne $\vdash A$ “

VZ 2009



Formální (axiomatický) systém K_n

Axiomy: **A1.** Všechny tautologie výrokové logiky

$$\mathbf{A2.} (K_i \alpha \wedge K_i (\alpha \rightarrow \beta)) \rightarrow K_i (\beta)$$

Odvozovací pravidla:

R1. Z α formulí $\alpha \rightarrow \beta$ odvoďte β (modus ponens)

R2. Z formule α odvoďte $K_i \alpha$ (pravidlo zobecnění znalosti)

Důkaz formule φ ve formálním systému je posloupnost formulí $\delta_1, \delta_2, \dots, \delta_n$ taková, že δ_n je dokazované φ a pro každé δ_i ($i < n+1$) platí

- δ_i je buď axiomem formálního systému
- nebo existují čísla j, k menší než i taková, že δ_i vznikne použitím odvozovacího pravidla na δ_j a δ_k .

Formule φ je dokazatelná ve formálním systému (značí se $\vdash \varphi$), pokud existuje její důkaz.

VZ 2009



Označme $\mathcal{M}_n(\Phi)$ množinu všech Kripkeho struktur nad množinou prvotních formulí Φ pro n agentů takových, že na relace K_i nejsou kladeny žádné požadavky.

$\mathcal{M}_n^{rst}(\Phi)$ necht' je pak podmnožina $\mathcal{M}_n(\Phi)$, ve které každá relace přípustnosti splňuje vlastnosti *rst*:

reflexivní

symetrická

transitivní.

Jedná se tedy o relace **ekvivalence**.

VZ 2009



Věta (viz cvičení).

Pro všechny struktury M pro n agentů s relacemi K_i , které jsou ekvivalencemi, a pro libovolné formule A, B platí

- (i) $M \models (K_i A \wedge K_i (A \rightarrow B)) \rightarrow K_i B$
- (ii) je-li $M \models A$ potom $M \models K_i A$
- (iii) $M \models K_i A \rightarrow A$
- (iv) $M \models K_i A \rightarrow K_i K_i A$
- (v) $M \models \neg K_i A \rightarrow K_i \neg K_i A$

Důkaz formule φ ve formálním systému z předpokladu α je posloupnost formulí $\delta_1, \delta_2, \dots, \delta_n$ taková, že δ_n je dokazované φ a pro každé δ_i ($i < n+1$) platí

- δ_i je buď axiomem formálního systému nebo se jedná o předpoklad α
- nebo existují čísla j, k menší než i taková, že δ_i vznikne použitím odvozovacího pravidla na δ_j a δ_k .

Formule φ je dokazatelná ve formálním systému z předpokladu α (značí se $\alpha \vdash \varphi$), pokud existuje důkaz φ z předpokladu α .

Axiomy modální logiky

1. Výrokové tautologie
2. Distribuční axiom (označovaný někdy jako **K**)
 $(K_i A \wedge K_i (A \rightarrow B)) \rightarrow K_i B$
3. Axiom znalostí (Axiom pravdy) (ozn. jako **T**) $K_i A \rightarrow A$
4. Axiom pozitivní introspekce (ozn. jako **4**) $K_i A \rightarrow K_i K_i A$
5. Axiom negativní introspekce (ozn. jako **5**) $\neg K_i A \rightarrow K_i \neg K_i A$
6. Axiom konzistence (ozn. jako **D**) $\neg K_i \text{false}$

Odvozovací pravidla:

- R1.** Z α formulí $\alpha \rightarrow \beta$ odvodte β (modus ponens)
- R2.** Z formule α odvod $K_i \alpha$ (pravidlo zobecnění znalosti)

Trzení 1 (příklad důkazu formule $K_i \varphi \rightarrow K_i \psi$ z předpokladu $(\varphi \rightarrow \psi)$ ve formálním systému \mathbf{K}_n): $\mathbf{K}_n, (\varphi \rightarrow \psi) \vdash K_i \varphi \rightarrow K_i \psi$

1. $(\varphi \rightarrow \psi)$ [předpoklad]
2. $K_i(\varphi \rightarrow \psi)$ [R2]
3. $K_i \varphi \rightarrow (K_i(\varphi \rightarrow \psi) \rightarrow K_i \psi)$ [K2]
4. $(K_i \varphi \rightarrow (K_i(\varphi \rightarrow \psi) \rightarrow K_i \psi)) \rightarrow (K_i(\varphi \rightarrow \psi) \rightarrow (K_i \varphi \rightarrow K_i \psi))$
[Prop-T1: $(\varphi \rightarrow (\psi \rightarrow \tau)) \rightarrow (\psi \rightarrow (\varphi \rightarrow \tau))$]
5. $K_i(\varphi \rightarrow \psi) \rightarrow (K_i \varphi \rightarrow K_i \psi)$ [MP 4 a 3]
6. $(K_i \varphi \rightarrow K_i \psi)$ [MP 5 a 2]

Trzení 2: Jsou-li dvě výrokové φ, ψ formule ekvivalentní (tj. $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$ je tautologie, značeno $\varphi \equiv \psi$), pak $\mathbf{K}_n \vdash K_i \varphi \equiv K_i \psi$.

Důsledek právě dokazaného tvrzení.

Věta 1: Systém K_n představuje korektní a úplný syntaktický popis všech formulí validních v množině $\mathcal{M}_n(\Phi)$ Kripkeho struktur (K_n je axiomatizací vzhledem k $\mathcal{M}_n(\Phi)$).

Věta 2:

Axiom T má tvar $K_i A \rightarrow A$. Systém $T_n = (K_n + \text{axiom T})$ je axiomatizací vzhledem k $\mathcal{M}_n^r(\Phi)$.

Axiom 4 má tvar $K_i A \rightarrow K_i K_i A$. Systém $S4_n = (T_n + \text{axiom 4})$ je axiomatizací vzhledem k $\mathcal{M}_n^{rs}(\Phi)$.

Axiom 5 má tvar $\neg K_i A \rightarrow K_i \neg K_i A$. Systém $S5_n = (S4_n + \text{axiom 5})$ je axiomatizací vzhledem k $\mathcal{M}_n^{rs}(\Phi)$.

$K_n \vdash K_i (a \wedge \beta) \rightarrow K_i a :$

Důkaz: [Jako zdůvodnění musíme uvést buď odvolání na axiom nebo informaci o použití odvoz. pravidla na předchozí řádky důkazu]

- $(a \wedge \beta) \rightarrow a$ [Výroková tautologie]
- $K_i ((a \wedge \beta) \rightarrow a)$ [R2: 1, čili R2 použito na formuli z řádky 1]
- $(K_i (a \wedge \beta) \wedge K_i ((a \wedge \beta) \rightarrow a)) \rightarrow K_i a$ [K]
- $((K_i (a \wedge \beta) \wedge K_i ((a \wedge \beta) \rightarrow a)) \rightarrow K_i a) \rightarrow (K_i ((a \wedge \beta) \rightarrow a) \rightarrow (K_i (a \wedge \beta) \rightarrow K_i a))$
[Výr. tautologie $((p \wedge q) \rightarrow r) \rightarrow (q \rightarrow (p \rightarrow r))$]
- $K_i ((a \wedge \beta) \rightarrow a) \rightarrow (K_i (a \wedge \beta) \rightarrow K_i a)$ [R1: 3,4]
- $K_i (a \wedge \beta) \rightarrow K_i a$ [R1: 2,5]

Další příklady důkazu některých platných vztahů:

e1) $K2, T(Axiom 3) \vdash \neg K_i false$

- $K_i false \rightarrow false$ [A3]
- $\neg false \rightarrow (\neg K_i false)$ [ekv. úprava 1]
- $\neg K_i false$ [použití výrokové tautologie na 2]

e2) $K2, T \vdash \neg K_i a \vee \neg K_i \neg K_i a$

e3) $K2, T \vdash \neg K_i (a \wedge \neg K_i a)$

- $K_i \neg K_i a \rightarrow \neg K_i a$ (A3, axiom pravdy)
- $\neg K_i \neg K_i a \vee \neg K_i a$ (výrok. Tautologie – přepis 1), viz a1
- $\neg (K_i \neg K_i a \wedge K_i a)$ (výrok. Tautologie – přepis 2)
- $\neg K_i (\neg K_i a \wedge a)$ (tvrzení o ekvivalenci a) pro formuli 3), viz a2

$K_n \vdash K_i (a \wedge \beta) \equiv K_i a \wedge K_i \beta$

- $K_n \vdash K_i (a \wedge \beta) \rightarrow K_i a$ [viz předchozí strana]
- $K_n \vdash K_i (a \wedge \beta) \rightarrow K_i \beta$ [důkaz obdobný tomu na předchozí straně]
- $(K_i (a \wedge \beta) \rightarrow K_i a) \rightarrow ((K_i (a \wedge \beta) \rightarrow K_i \beta) \rightarrow (K_i (a \wedge \beta) \rightarrow (K_i a \wedge K_i \beta)))$
[$(\phi \rightarrow \psi) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow (\psi \wedge \chi)))$ výroková tautologie]
- $(K_i (a \wedge \beta) \rightarrow K_i \beta) \rightarrow (K_i (a \wedge \beta) \rightarrow (K_i a \wedge K_i \beta))$ [MP: 3,1]
- $K_i (a \wedge \beta) \rightarrow (K_i a \wedge K_i \beta)$ [MP: 4,2]

Tvrzení 1: $K_n \vdash K_i (a \wedge \beta) \equiv K_i a \wedge K_i \beta$

Důkaz: kombinace důkazů na této a následující straně.

$K_n \vdash K_i \alpha \wedge K_i \beta \rightarrow K_i(\alpha \wedge \beta)$

6. $\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))$ [výroková tautologie]
7. $K_i(\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta)))$ [R2:6]
8. $K_i \alpha \rightarrow (K_i(\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))) \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$ [A2]
9. $K_i(\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))) \rightarrow (K_i \alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$ [výroková úprava 8]
10. $(K_i \alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$ [MP: 9,7]
11. $(K_i \alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow (K_i \beta \rightarrow (K_i \alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta))))$ [Výr.Ax1]
12. $K_i \beta \rightarrow (K_i \alpha \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta)))$ [MP: 11,10]
13. $(K_i \alpha \wedge K_i \beta) \rightarrow K_i(\beta \rightarrow (\alpha \wedge \beta))$ [výroková úprava 12]
14. $(K_i \alpha \wedge K_i \beta) \rightarrow (K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)))$ [výroková úprava 13]
15. $K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta)) \rightarrow K_i(\alpha \wedge \beta)$ [A2]
16. $(K_i \alpha \wedge K_i \beta) \rightarrow (K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow K_i(\alpha \wedge \beta)$
 $\rightarrow ((K_i \alpha \wedge K_i \beta) \rightarrow K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow ((K_i \alpha \wedge K_i \beta) \rightarrow K_i(\alpha \wedge \beta))$ [Výr. Ax3]
17. $(K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow K_i(\alpha \wedge \beta) \rightarrow ((K_i \alpha \wedge K_i \beta) \rightarrow (K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow K_i(\alpha \wedge \beta))$ [V.Ax1]
18. $((K_i \alpha \wedge K_i \beta) \rightarrow (K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow K_i(\alpha \wedge \beta))$ [MP: 17,15]
19. $((K_i \alpha \wedge K_i \beta) \rightarrow K_i \beta \wedge K_i(\beta \rightarrow (\alpha \wedge \beta))) \rightarrow ((K_i \alpha \wedge K_i \beta) \rightarrow K_i(\alpha \wedge \beta))$ [MP: 16,18]
20. $(K_i \alpha \wedge K_i \beta) \rightarrow K_i(\alpha \wedge \beta)$ [MP: 19,14]

VZ 2009

$E_G \quad C_G \quad D_G$

G je podmnožinou $\{1, 2, \dots, n\}$, $E_G A$ je **pravdivé**, právě když **každý agent z G ví A** . Tedy

Axiom C1. $E_G A \Leftrightarrow \bigcap_{i \in G} K_i A$

Intuitivně, **společná** (všeobecná, *common*) **znalost** je to „co je každému zcela jasně a ví to“, proto nepřekvapí, že společná znalost má všechny vlastnosti znalostí podobné **Distribučnímu axiomu, Axiomu znalostí, Axiomům pozitivní a negativní introspekce**.

Pro všeobecnou znalost mezi skupinami agentů platí

$Je-li \quad Q \subseteq G \quad potom \quad C_G A \rightarrow C_Q A$

VZ 2009

Další vztahy, které lze dokázat:

- a) $K_n \vdash K_i(\alpha \wedge \beta) \equiv K_i \alpha \wedge K_i \beta$
- b) $(K_n + A6) \vdash \neg(K_i \alpha \wedge K_i \neg \alpha)$
- c) $(K_n + A3) \vdash A6$
- d) $K_n \vdash K_i \neg(p \rightarrow K_i p) \equiv K_i(p \wedge \neg K_i p) \equiv (K_i p \wedge K_i(\neg K_i p))$
- e) $\forall (K_n + A3)$ nemůže být dokazatelné $K_i \neg(p \rightarrow K_i p)$

VZ 2009

Cvičení. Dokažte, že následující formule jsou validní (platné ve všech Kripkeho strukturách):

- (i) $(C_G A \wedge C_G(A \rightarrow B)) \rightarrow C_G B$
- (ii) $C_G A \rightarrow A$
- (iii) $C_G A \rightarrow C_G C_G A$
- (iv) $\neg C_G A \rightarrow C_G \neg C_G A$

Předpoklady o relacích K_i jsou stejné, jako při odvození odpovídajících axiomů pro znalosti.

VZ 2009

Axiom o pevném bodu.

Nechť p je výrok „alespoň jeden z vás má zablácené čelo“.

Připomeňme, že ve hře zablácených dětí, se z p stala všeobecná znalost v okamžiku, kdy otec řekl p tak, aby všechny děti věděly,

- že platí p a
- že všechny děti to vědí.

Axiom C2.

$$\models C_G A \leftrightarrow E_G(A \wedge C_G A)$$

{ $C_G A$ je pevným bodem funkce $f(x) = E_G(A \wedge x)$ }

VZ 2009



Následující tvrzení dává sémantické odůvodnění charakteristiky operátoru E_G , Axiomu o pevném bodu a Indukčního pravidla.

Věta - shrnutí

Pro všechny struktury M , všechny neprázdné podmnožiny G množiny $\{1, 2, \dots, n\}$ a libovolné formule A, B platí

$$(i) \quad M \models E_G A \leftrightarrow \bigcap_{i \in G} K_i A$$

kde $\bigcap_{i \in G} K_i A$ označuje konjunkci všech $K_i A$ pro vš. $i \in G$

$$(ii) \quad M \models C_G A \leftrightarrow E_G(A \wedge C_G A)$$

$$(iii) \quad \text{je-li } M \models A \rightarrow E_G(B \wedge A), \text{ potom } M \models A \rightarrow C_G B$$

VZ 2009



Následující odvozovací pravidlo dává způsob, jak ukázat, že v nějaké struktuře platí společná znalost.

Indukční pravidlo RC1

Pro všechny struktury M platí

$$\text{je-li } M \models A \rightarrow E_G(B \wedge A), \text{ potom } M \models A \rightarrow C_G B$$

Idea. Předpoklad indukčního pravidla dává možnost indukci podle k dokázat, že formule

$$A \rightarrow E_G^k(B \wedge A) \text{ je validní pro každé } k.$$

VZ 2009



Distribuované znalosti

charakterizují znalosti, které odpovídají tomu, když „všichni agenti dají hlavy dohromady“. Proto splňují stejné vlastnosti jako znalosti. Navíc upozorníme na dvě drobnosti.

$$\models D_{\{i\}} A \leftrightarrow K_i A$$

pro jednočlennou skupinu se distribuované znalosti shodují s tím, co agent ví.

Naopak, čím je skupina větší, tím větší jsou její distribuované znalosti.

$$\text{Je-li } G \subseteq Q \text{ potom } \models D_G A \rightarrow D_Q A$$

VZ 2009



DÚ-MOL2a Anna a Bob (2 body)

- Anna a Bob vědí, že organizátor vybere z osudí nějaké přirozené číslo $n < 10$, které napíše na čelo jednomu z nich a druhému napíše číslo „sousední“, tj. buď $n+1$ nebo $n-1$. Ani Anna ani Bob neznají své číslo - vidí jen to partnerovo. Anna vidí číslo 4.
- Nakreslete odpovídající Kripkeho strukturu a popište podrobně postup (výměnu informací typu „Neznám své číslo!“), při němž se oba snaží své číslo zjistit.

VZ 2009



DÚ-MOL2b Dokazování (3 body)

Předvěčte se, že v $(K_n + A3)$ nemůže být dokazatelné $K_i \neg (p \rightarrow K_i p)$.

Návod: Předpokládejte, že tato formule dokazatelná je a za tohoto předpokladu lze dokázat jak $\neg K_i p$, tak $K_i p$.

Podle věty o úplnosti by obě tyto formule by pak musely být platné v libovolné vhodné Kripkeho struktuře. Taková situace však nemůže nikdy nastat!

VZ 2009

