Logika prvního řádu a transparentní intenzionální logika (TIL)

Aleš Horák

E-mail: hales@fi.muni.cz
http://nlp.fi.muni.cz/uui/

Obsah:

- Predikátová logika prvního řádu
- Logická analýza přirozeného jazyka
- Transparentní intenzionální logika

Výhody a nevýhody výrokové logiky

- ☺ výroková logika je deklarativní: syntaxe přímo koresponduje s fakty
- ☺ výroková logika umožňuje zpracovávat částečné/disjunktivní/negované informace (což je víc, než umí většina datových struktur a databází)
- ☺ výroková logika je kompoziční:

  význam $P_1 \wedge P_2$ *je odvozen z významu* P_1 *a* P_2

- ☺ ve výrokové logice je význam kontextově nezávislý (narozdíl od přirozeného jazyka, kde význam závisí na kontextu)
- ☹ výroková logika má velice omezenou expresivitu (narozdíl od přirozeného jazyka)
  např. nemáme jak říct "*Jámy způsobují Vánek ve vedlejších místnostech*" jinak, než vyjmenovat odpovídající výrok pro každé pole

Predikátová logika prvního řádu

- *First-order predicate logic*, FOPL/PL1
- výroková logika → svět obsahuje fakty × PL1 předpokládá, že svět obsahuje:
  - objekty – lidi, domy, teorie, barvy, roky, ...
  - relace – červený, kulatý, prvočíselný, bratři, větší než, uvnitř, ...
  - funkce – otec někoho, nejlepší přítel, plus jedna, začátek čeho, ...

Syntaxe predikátové logiky

- základní prvky –

konstanty	KingJohn, 2, RichardTheLionheart, ...
funktory predikátů	Brother, $>$, ...
funkce	Sqrt, LeftLegOf, ...
proměnné	x, y, a, b, ...
spojky	$\wedge \ \vee \ \neg \ \Rightarrow \ \Leftrightarrow$
rovnost	$=$
kvantifikátory	$\forall \ \exists$

- atomické formule –

predikáty	Brother(KingJohn, RichardTheLionheart)
složené termy	$>\big(\mathrm{Length}(\mathrm{LeftLegOf}(\mathrm{Richard})), \mathrm{Length}(\mathrm{LeftLegOf}(\mathrm{KingJohn}))\big)$

- složené formule – tvoří se z atomických formulí pomocí spojek

$$\neg S, \quad S_1 \wedge S_2, \quad S_1 \vee S_2, \quad S_1 \Rightarrow S_2, \quad S_1 \Leftrightarrow S_2$$

např. $\mathrm{Sibling}(\mathrm{KingJohn}, \mathrm{Richard}) \Rightarrow \mathrm{Sibling}(\mathrm{Richard}, \mathrm{KingJohn})$

$>(1,2) \vee \leq(1,2)$

$>(1,2) \wedge \neg{>}(1,2)$

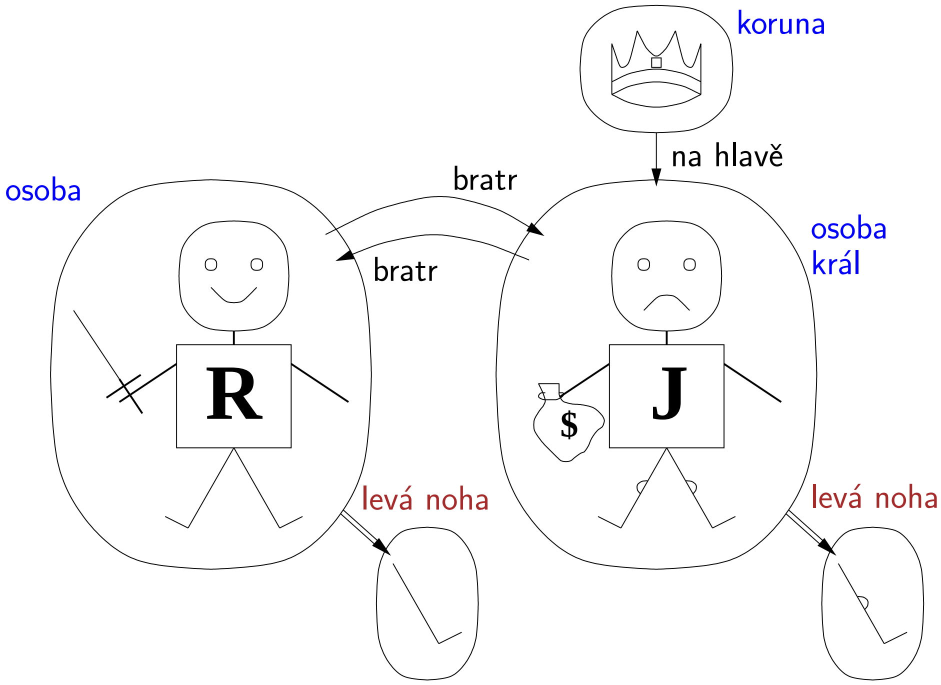
Pravdivost v predikátové logice

pravdivost formule (sémantika) se určuje vzhledem k *modelu* a *interpretaci*
model obsahuje ≥ 1 objektů a relace mezi nimi
interpretace definuje vztah mezi syntaxí a modelem – určuje referenty pro:

konstantní symboly $\rightarrow$ *objekty*
predikátové symboly $\rightarrow$ *relace*
funkční symboly $\rightarrow$ *funkce*

atomická formule **predikát(term$_1$, . . . , term$_n$)** je pravdivá $\Leftrightarrow$
$\Leftrightarrow$ *objekty* odkazované pomocí **term**$_1$, . . . , **term**$_n$ jsou v *relaci* pojmenované funktorem **predikát**.

Příklad modelu a interpretace ve FOPL



5 objektů, 2 binární relace, 3 unární relace (osoba, král, koruna) a 1 unární funkce (levá noha).

Univerzální kvantifikace

$\forall\langle \textit{proměnné}\rangle\ \langle \textit{formule}\rangle$

"Každý na FI MU je inteligentní:" $\forall x$ Na(x, FI MU) $\Rightarrow$ inteligentní(x)

$\forall x\ P$ je pravdivé v modelu m $\Leftrightarrow$ P je pravdivá pro $x =$ každý možný objekt z modelu m

zhruba odpovídá konjunkci instanciací P

$$\begin{array}{ll} & \text{Na(Petr, FI MU)} \Rightarrow \text{inteligentní(Petr)} \\ \wedge & \text{Na(Honza, FI MU)} \Rightarrow \text{inteligentní(Honza)} \\ \wedge & \text{Na(FI MU, FI MU)} \Rightarrow \text{inteligentní(FI MU)} \\ \wedge & \ldots \end{array}$$

Existenční kvantifikace

$\exists\langle \textit{proměnné}\rangle\ \langle \textit{formule}\rangle$

"Někdo na MFF UK je inteligentní:" $\exists x$ Na(x, MFF UK) $\wedge$ inteligentní(x)

$\exists x\ P$ je pravdivé v modelu m $\Leftrightarrow$ P je pravdivá pro $x =$ nějaký objekt z modelu m

zhruba odpovídá disjunkci instanciací P

$$\begin{array}{ll} & \text{Na(Petr, MFF UK)} \wedge \text{inteligentní(Petr)} \\ \vee & \text{Na(Honza, MFF UK)} \wedge \text{inteligentní(Honza)} \\ \vee & \text{Na(MFF UK, MFF UK)} \wedge \text{inteligentní(MFF UK)} \\ \vee & \ldots \end{array}$$

Vlastnosti kvantifikací

- pozor při použití kvantifikátorů na záměnu $\wedge$ a $\Rightarrow$:

	dobře	*špatně*	znamenalo by
"každý P je Q."	$\forall x\ P \Rightarrow Q$	$\forall x\ P \wedge Q$	"každý je P i Q."
"někdo P je Q."	$\exists x\ (P \wedge Q)$	$\exists x\ (P \Rightarrow Q)$	"někdo není P nebo je Q."

- $\forall x \forall y$ je stejné jako $\forall y \forall x$
  $\exists x \exists y$ je stejné jako $\exists y \exists x$
  $\exists x \forall y$ není stejné jako $\forall y \exists x$
  $\exists x \forall y$ má_rád(x, y) – "Existuje osoba, která má ráda všechny lidi na světě."
  $\forall y \exists x$ má_rád(x, y) – "Každého na světě má alespoň jedna osoba ráda." (potenciálně každého jiná)
- dualita kvantifikátorů
  oba mohou být vyjádřeny pomocí druhého

$$\forall x\ \text{má\_rád}(x, \text{zmrzlina}) \equiv \neg\exists x\ \neg\text{má\_rád}(x, \text{zmrzlina})$$
$$\exists x\ \text{má\_rád}(x, \text{mrkev}) \equiv \neg\forall x\ \neg\text{má\_rád}(x, \text{mrkev})$$

Inference ve FOPL

teoreticky můžeme určit všechny modely výčtem ze slovníku *KB*:
  pro počet objektů $n = 1, \ldots, (\infty)$
    pro každý k-ární predikát P_k ze slovníku
      pro každou možnou k-ární relaci na n objektech
        pro každý konstantní symbol C ze slovníku
          pro každou volbu referenta pro C z n objektů ...

prakticky je *kontrola modelů* nepoužitelná

inference je možná pouze podle inferenčních pravidel (dopředné/zpětné řetězení, rezoluce, ...)

základní inferenční pravidlo – zobecněné Modus Ponens (*Generalized Modus Ponens, GMP*)

$$\frac{p_1', \ p_2', \ldots, p_n', \ (p_1 \wedge p_2 \wedge \ldots \wedge p_n \Rightarrow q)}{\textsc{Subst}(\theta, q)}$$

kde $\forall i\ \textsc{Subst}(\theta, p_i') = \textsc{Subst}(\theta, p_i)$
pro atomické formule p_i, p_i' a q

– používá navíc unifikaci
– vzniká z MP pomocí liftingu
– využívá upravené verze inferenčních algoritmů – dopředné/zpětné řetězení, rezoluce

Báze znalostí ve FOPL

předpokládejme, že agent ve Wumpusově jeskyni cítí Zápach a Vánek, ale nevidí Třpyt, nenarazil do zdi a nezabil Wumpuse v čase $t = 5$:

```
tell(KB, percept([zápach, vánek, nic, nic, nic], 5)).
?- ask(KB,action(A,5)). %    ∃A action(A,5)  ?
```

tj. dotaz "Vyplývá nějaká akce z KB v čase $t = 5$?"
odpověď: $true, \{a/\text{Výstřel}\}$ ← substituce (hodnot proměnným)
pro větu S a substituci σ → $S\sigma$ označuje výsledek aplikace σ na S:

$$
\begin{aligned}
S &= \text{chytřejší}(x, y) \\
\sigma &= \{x/\text{Petr}, y/\text{Honza}\} \\
S\sigma &= \text{chytřejší}(\text{Petr}, \text{Honza})
\end{aligned}
$$

$\textsc{Ask}(KB, S)$ vrací některá/všechna σ takové, že $KB \models S\sigma$

Báze znalostí pro Wumpusovu jeskyni

Vnímání:

$\forall v, tr, n, w, t\ \ Percept([Zápach, v, tr, n, w], t) \Rightarrow Je\_zápach(t)$
$\forall z, v, n, w, t\ \ Percept([z, v, Třpyt, n, w], t) \Rightarrow Máme\_zlato(t)$

Reflex:

$\forall t\ \ Máme\_zlato(t) \Rightarrow Action(Zvednutí, t)$

Reflex s vnitřním stavem: neměli jsme už zlato?

$\forall t\ \ Máme\_zlato(t) \wedge \neg Držím(Zlato, t) \Rightarrow Action(Zvednutí, t)$

Držím(Zlato, t) není pozorovatelné ⇒ je důležité držet si informace o vnitřních stavech

Báze znalostí pro Wumpusovu jeskyni pokrač.

Vyvozování skrytých skutečností:

- vlastnosti pozice:

$$\forall x, t\ Na\_poli(Agent, x, t) \land Je\_zápach(t) \Rightarrow Zapáchá(x)$$
$$\forall x, t\ Na\_poli(Agent, x, t) \land Je\_vánek(t) \Rightarrow S\_vánkem(x)$$

- "V poli vedle Jámy je Vánek:"
  - diagnostické pravidlo – odvodí příčiny z následku

$$\forall y\ S\_vánkem(y) \Rightarrow \exists x\ Jáma(x) \land Vedle(x, y)$$

  - příčinné pravidlo – odvodí výsledek z premisy

$$\forall x, y\ Jáma(x) \land Vedle(x, y) \Rightarrow S\_vánkem(y)$$

  - ani jedno z nich není úplné
    např. příčinné pravidlo neříká, jestli v poli daleko od Jámy nemůže být Vánek
  - definice vztahu Vánku a Jámy:

$$\forall y\ S\_vánkem(y) \Leftrightarrow [\exists x\ Jáma(x) \land Vedle(x, y)]$$

Báze znalostí pro Wumpusovu jeskyni – rozhodování

- počáteční podmínka v KB:

$$Na\_poli(Agent, [1, 1], S_0)$$

- dotaz

$$\textsc{Ask}(KB, \exists s\ Držím(Zlato, s))$$

  tj., "V jaké situaci budu držet Zlato?"

- situace jsou propojeny pomocí funkce *Result*:

$$Result(a, s)\ \ldots\ \textit{situace, která je výsledkem činnosti } a \textit{ v } s$$

- odpověď (např. v situaci, kdy hned na vedlejším poli je Zlato)

$$\{s/Result(Zvednutí, Result(Krok\ dopředu, S_0))\}$$

  tj., jdi dopředu a zvedni Zlato

Shrnutí

logický agent aplikuje inferenci na bázi znalostí pro vyvození nových znalostí a tvorbu rozhodnutí
základní koncepty logiky:

syntaxe: formální struktura vět
sémantika: pravdivost vět podle modelů
vyplývání: nutná pravdivost věty v závislosti na jiné větě

inference: vyvození věty z jiných vět
bezespornost: inference produkuje jen vyplývající věty
úplnost: inference vyprodukuje ∀ vyplývající věty

výroková logika nemá dostatečnou expresivitu

predikátová logika prvního řádu:

– syntaxe: konstanty, funkce, predikáty, rovnost, kvantifikátory
– větší expresivita – dostatečná pro Wumpusovu jeskyni
– "poslední" logika, pro kterou existuje bezesporná a úplná inference (Gödelovy věty o neúplnosti)

jiné možné logiky:

jazyk	ontologie	pravdivostní hodnoty
výroková logika	fakty	true/false/$\perp$
predikátová logika 1. řádu	fakty, objekty, relace	true/false/$\perp$
temporální logika	fakty, objekty, relace, čas	true/false/$\perp$
teorie pravděpodobnosti	fakty	míra pravděpodobnosti $\in [0,1]$
fuzzy logika	míra pravdivosti $\in [0,1]$	intervaly hodnot

Logická analýza přirozeného jazyka

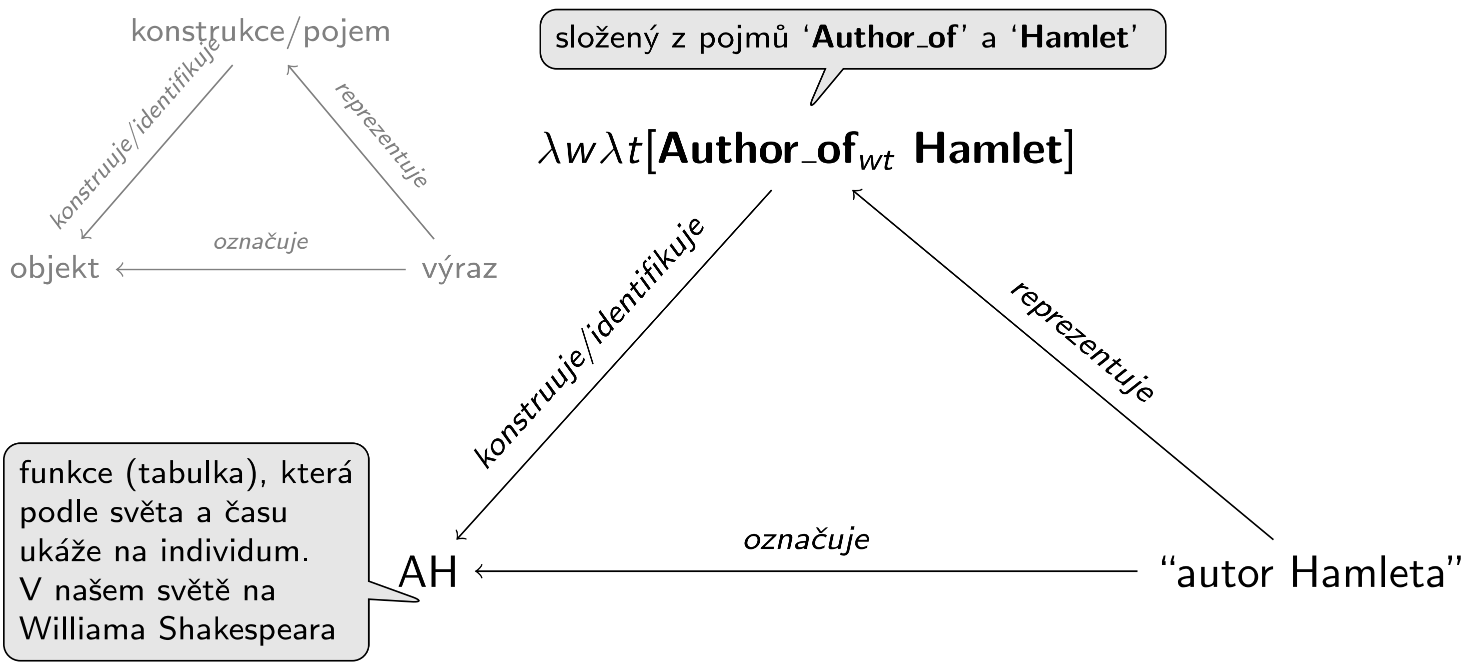
logická analýza PJ – analýza významu výrazů (vět) PJ
přirozený jazyk (čeština, angličtina, ...) = nástroj pojmového uchopení reality

pojem – kritéria/procedury umožňující identifikovat různé konkrétní a abstraktní objekty (např. "planeta" – třída nebeských těles s určitými charakteristikami – obíhá po oběžné dráze kolem stálice, není zdrojem světla, ...)

– pojem ≠ *výraz* – např. výrazy v různých jazycích často reprezentují stejný pojem (pojem("prvočíslo") ≡ pojem("prime number"))
– pojem ≠ *představa* – představa je *subjektivní*, pojem je objektivní
– pojmy mohou identifikovat různé objekty:
  - jedno individuum – individuální pojmy (např. Petr, Pegas, prezident ČR)
  - třídu objektů – vlastnost (např. červený, šelma, hora)
  - *n*-člennou relaci – vztah (např. otec (někoho), křivdit (někdo někomu))
  - pravdivostní hodnotu – propozice (např. v Brně prší)
  - funkcionální přiřazení – empirické funkce (např. rychlost)
  - číslo – (fyzikální) veličiny (např. rychlost světla)

Vztah pojmu a výrazu

ve zjednodušené podobě: pojem odpovídá logické konstrukci



Omezenost predikátové logiky 1. řádu

dva omezující rysy:

– nedostatečná expresivita

– extenzionalismus

Expresivita: vyjadřovací síla jazyka

“Je-li barva stropu pokoje č. 3 uklidňující, je pokoj č. 3 vhodný pro pacienta X a není vhodný pro pacienta Y.”

analýza ve výrokové logice:

$P \Rightarrow (Q \wedge \neg R)$

P	“Barva stropu pokoje č. 3 je uklidňující.”
Q	“Pokoj č. 3 je vhodný pro pacienta X.”
R	“Pokoj č. 3 je vhodný pro pacienta Y.”

analýza v PL1:

$U(B) \Rightarrow (V(P, X) \wedge \neg V(P, Y))$

U	třída uklidňujících objektů
B	individuum ‘barva stropu pokoje č. 3’
V	relace mezi individuy ‘být vhodný pro’
P	individuum ‘pokoj č. 3’
X, Y	individua ‘pacient X’ a ‘pacient Y’

Nedostatečná expresivita PL1

Červená barva je krásnější než hnědá barva. *Kostka je červená.*

analýza v PL1:

$Kr(\check{C}_1, H)$ $\check{C}_2(Ko)$

$\check{C}_1$ individuum 'červená barva'
$\check{C}_2$ vlastnost individuí 'být červený' (třída červených objektů)
nelze vyjádřit $\check{C}_1 \equiv \check{C}_2$

Extenzionalismus PL1

Varšava *hlavní město Polska*

Varšava	–	jméno individua, jasně identifikovatelné a odlišitelné
hlavní město Polska	–	individuová role, momentálně identifikuje Varšavu, ale dříve to byl i Krakov

'hlavní město Polska'

- závisí na světě a čase
- pochopení významu, ale není vázané na znalost obsahu – tj. význam na světě a čase nezávisí

Extenzionalismus PL1 pokrač.

číslo X je větší než číslo Y *budova X je větší než budova Y*

matematické větší než – relace dvojic čísel, pevně daná
empirické větší než – vztah dvou individuí, který se může měnit v čase (otec a syn)

ano *V Brně prší*

ano – pravdivostní hodnota *true*
V Brně prší – propozice – označuje pravdivostní hodnotu, která se mění (alespoň) v čase

i když hodnota někdy závisí na světě a čase, samotný význam na nich nezávisí

Extenze a intenze

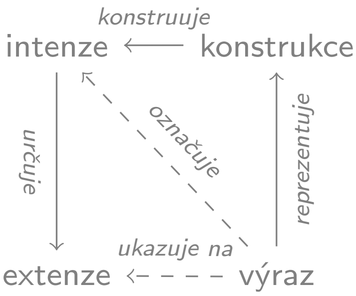
Definujeme:

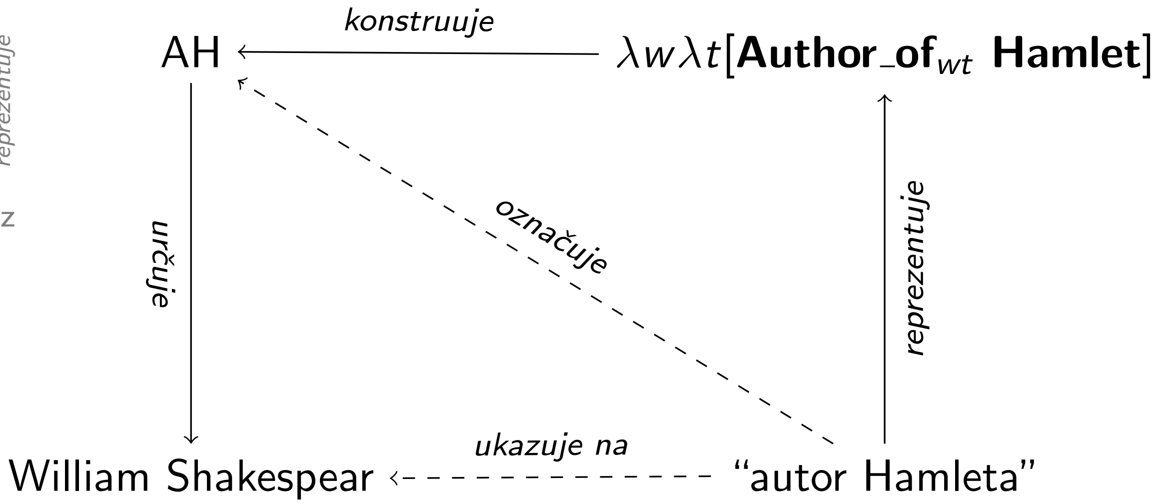
- intenze – objekty typu funkcí, jejichž hodnoty závisí na světě a čase
- extenze – ostatní objekty (na světě a čase nezávislé)

časté extenze a intenze:

extenze	*intenze*
individua	individuové role
třídy	vlastnosti
relace	vztahy
pravdivostní hodnoty	propozice
funkce	empirické funkce
čísla	veličiny

Rozšířený vztah výrazu a významu u intenzí





Transparentní intenzionální logika

- Transparent Intensional Logic, TIL
- logický systém speciálně navržený pro zachycení významu výrazů PJ
- autor Pavel Tichý: *The Foundations of Frege's Logic*, de Gruyter, Berlin, New York, 1988.
- obdobná teorie – *Montagueho intenzionální logika* – Tichý ukazuje její nedostatky
- Tichý vychází z myšlenek – *Gottlob Frege* (1848 – 1925, logik) a *Alonzo Church* (1903 – 1995, teorie typů)
- vlastnosti:
  - rozvětvená typová hierarchie (s typy vyšších řádů)
  - temporální
  - intenzionální (intenze × extenze)
- transparentost:
  1. nositel významu (konstrukce) není prvek formálního aparátu, tento aparát pouze *studuje* konstrukce
  2. zachycení intenzionality je přesně popsáno z matematického hlediska

Typy v TILu

typ objektu:

– základní typy – typová báze $= \{o, \iota, \tau, \omega\}$

– funcionální typy – funkce nad typovou bází

  např. ι, $((\iota\tau)\omega)$, $(o\iota)$, $(((o\iota)\tau)\omega)$, $((o\tau)\omega), \ldots$

  $((\alpha\tau)\omega)$ . . . závislost na světě a čase, vyjadřuje intenze – zápis $\alpha_{\tau\omega}$

– typy vyšších řádů – obsahují i třídy konstrukcí řádu n – $*_n$

Základní typy TILu

umožňují přiřadit typ objektům z intenzionální báze jazyka – třída základních vlastností (barvy, rozměry, postoje, . . .) popisujících stav světa

- o (omikron, o) . . . pravdivostní hodnoty Pravda (*true*, T) a Nepravda (*false*, F)
  přesně odpovídají běžným logikám, typy logických operátorů – $(oo), (ooo)$
- ι (jota) . . . třída individuí
  individua ovšem ne jako kompletní objekty, ale jako numerická identifikace nestrukturované entity
- τ (tau) . . . třída časových okamžiků (jako časového kontinua)
  zachycení závislosti na čase; současně třída reálných čísel
- ω (omega) . . . třída možných světů
  zachycení empirické závislosti na stavu světa

Možné světy

termín možný svět – Gottfried Wilhelm von Leibniz (1646–1716, filozof a matematik)

$\forall$ možný svět je:
- soubor myslitelných faktů
- je konzistentní a maximální ze všech takových souborů
- je objektivní (nezávislý na individuálním názoru)

mezi možnými světy $\exists$ právě jeden aktuální svět – jeho znalost $\equiv$ vševědoucnost

možný svět v TILu = rozhodovací systém, pro $\forall$ prvek intenzionální báze obsahuje konzistentní přiřazení hodnot

příklad – realita s 2 objekty a 2 vlastnostmi (9 možných světů $w_1, \ldots, w_9$):

	být tlustý			
být hubený	{Laurel, Hardy}	{Laurel}	{Hardy}	$\emptyset$
{Laurel, Hardy}	$\times$	$\times$	$\times$	w_1
{Laurel}	$\times$	$\times$	$\boxed{w_2}$	w_3
{Hardy}	$\times$	w_4	$\times$	w_5
$\emptyset$	w_6	w_7	w_8	w_9

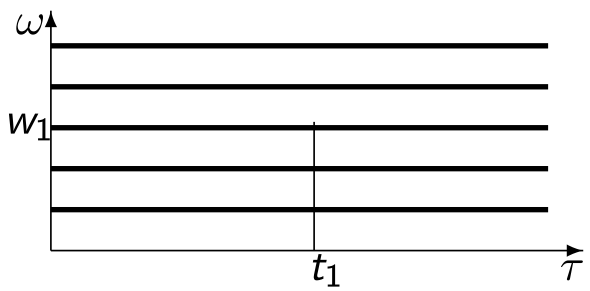
Princip intenzí v TILu

být hubený	...	objekt typu $(o\iota)_{\tau\omega}$, funkce z možných světů a času do tříd individuí
w	...	proměnná typu ω, možný svět
t	...	proměnná typu τ, časový okamžik
[**být hubený** w t]	...	konstruuje $(o\iota)$-objekt, třídu individuí, kteří mají ve světě w a čase t vlastnost být hubený (značíme **být hubený**$_{wt}$)

pokud aplikujeme jen w – získáme chronologii

Americký prezident$_{w_{act}}$ (zkr. $\mathbf{P}_{w_{act}}$) ... ι_τ $\mathbf{P}_{w_{act}t_0}$ ... ι:

intenzionální sestup – identifikace extenze pomocí intenze, světa w_1 a času t_1



Nejčastější typy

extenze			*intenze*		
individua	...	ι	individuové role	...	$\iota_{\tau\omega}$
třídy	...	$(o\iota)$	vlastnosti	...	$(o\iota)_{\tau\omega}$
relace	...	$(o\alpha\beta)$	vztahy	...	$(o\alpha\beta)_{\tau\omega}$
pravdivostní hodnoty	...	o	propozice	...	$o_{\tau\omega}$, π
funkce	...	$(\alpha\beta)$	empirické funkce	...	$(\alpha\beta)_{\tau\omega}$
čísla	...	τ	veličiny	...	$\tau_{\tau\omega}$

Konstrukce

konstrukce v TILu:

- proměnná typu α, v závislosti na valuaci konstruuje α-objekt
  $x \dots \iota$
- trivializace objektu A typu α, konstruuje právě objekt A
  $^0\mathrm{A} \dots \alpha$, často také $\mathbf{A} \dots \alpha$
  trivializace složené konstrukce – přechod k vyšším řádům
- aplikace konstrukce $X \dots (\alpha\beta_1 \dots \beta_n)$ na konstrukce $Y_1, \dots, Y_n$ typů $\beta_1, \dots, \beta_n$, konstruuje objekt typu α
  $[XY_1 \dots Y_n] \dots \alpha$
- abstrakce konstrukce $Y \dots \alpha$ na proměnných $x_1, \dots, x_n$ typů $\beta_1, \dots, \beta_n$, konstruuje objekt/funkci typu $(\alpha\beta_1 \dots \beta_n)$
  $\lambda x_1 \dots x_n[Y] \dots (\alpha\beta_1 \dots \beta_n)$

U aplikace i abstrakce se tady jedná o zápis *funkcí více proměnných*, ne o částečné aplikace

Příklady analýzy podstatných jmen

pes, člověk	$x \ldots \iota$: $\mathbf{pes}_{wt}x$, $pes/(o\iota)_{\tau\omega}$	individuum z dané třídy individuí
prezident	$prezident/\iota_{\tau\omega}$	individuová role
volitelnost	$volitelnost/(o\iota_{\tau\omega})_{\tau\omega}$	vlastnost individuové role
výška	$výška/(\tau\iota)_{\tau\omega}$	empirická funkce
výrok, tvrzení	$p \ldots *_n$: $\mathbf{výrok}_{wt}p$, $výrok/(o*_n)_{\tau\omega}$	konstrukce propozice z dané třídy konstrukcí propozic
válka, smích, zvonění	$válka/(o(o\pi))_{\omega}$	třída epizod – aktivita, která koresponduje se slovesem
leden, podzim	$leden/(o(o\tau))$	třída časových okamžiků – časové intervaly

Příklady přínosu TILu

- propoziční postoje

  Petr říká, že Tom věří, že Země je kulatá.

$$\lambda w \lambda t \left[\mathbf{říká}_{wt} Petr^0 \left[\lambda w \lambda t \left[\mathbf{věří}_{wt} Tom^0 \left[\lambda w \lambda t [\mathbf{kulatá}_{wt} \mathbf{Země}]\right]\right]\right]\right]$$

- existence neexistujícího

  Pes existuje.   *Jednorožec neexistuje.*

v PL1: $\cancel{\exists x (x = \text{pes})}$ $\cancel{\neg\exists x (x = \text{jednorožec})}$

$$(\text{jednorožec} = \text{jednorožec}) \Rightarrow (\exists x (x = \text{jednorožec}))$$

v TILu: (*) $\lambda w \lambda t \left[{}^0\neg [Ex_{wt}\, \mathbf{jednorožec}]\right]$

$$Ex \overset{df}{=} \lambda w \lambda t \lambda p \left[{}^0\textstyle\sum_\iota \left[\lambda x [p_{wt}\, x]\right]\right], \quad Ex \ldots (o(o\iota)_{\tau\omega})_{\tau\omega}$$

(*) … “třída všech individuí s vlastností ‘být jednorožcem’ je v daném světě a čase prázdná.”

- intenzionalita, vlastnosti vlastností, analýza epizod, analýza gramatického času, …