Intenzionální sémantika a reprezentace znalostí

Aleš Horák

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

Obsah:

- Intenzionální sémantika
- Reprezentace znalostí

Logická analýza přirozeného jazyka

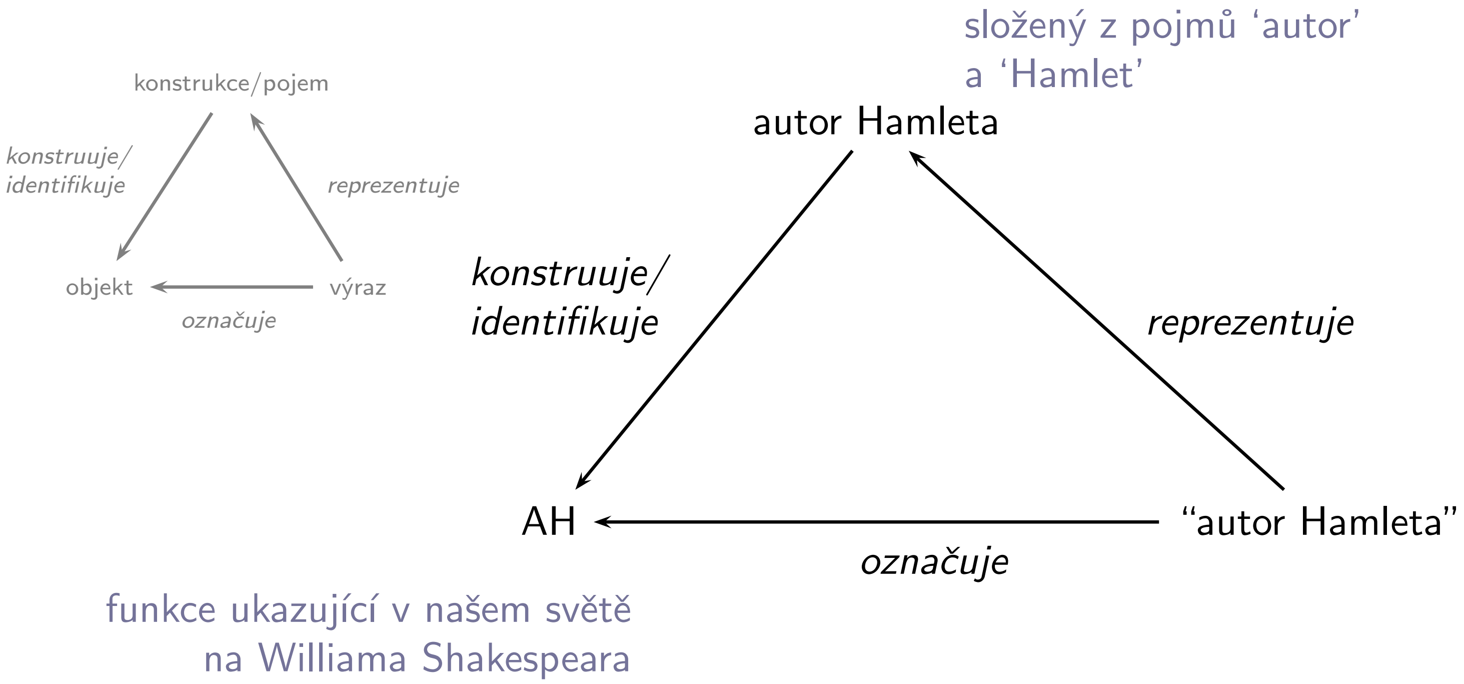
logická analýza PJ – analýza **významu** výrazů (vět) PJ
přirozený **jazyk** = nástroj *pojmového* uchopení reality
pojem – kritéria/procedury umožňující identifikovat různé konkrétní a abstraktní objekty

např. "planeta" – třída nebeských těles s určitými charakteristikami – obíhá po oběžné dráze kolem stálice, není zdrojem světla, ...

- **pojem** $\neq$ **výraz** – např. výrazy v různých jazycích často reprezentují stejný pojem (pojem("prvočíslo") $\equiv$ pojem("prime number"))
- **pojem** $\neq$ **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 – pokrač.

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

analýza v **PL1**:

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

$\check{C}_1$ 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**

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)

Extenzionalismus PL1 – pokrač.

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

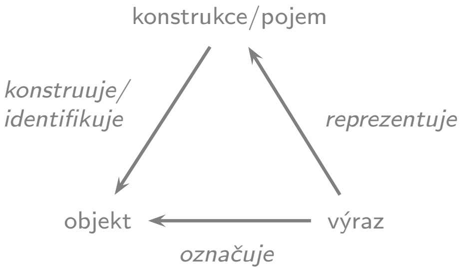
Definujeme:

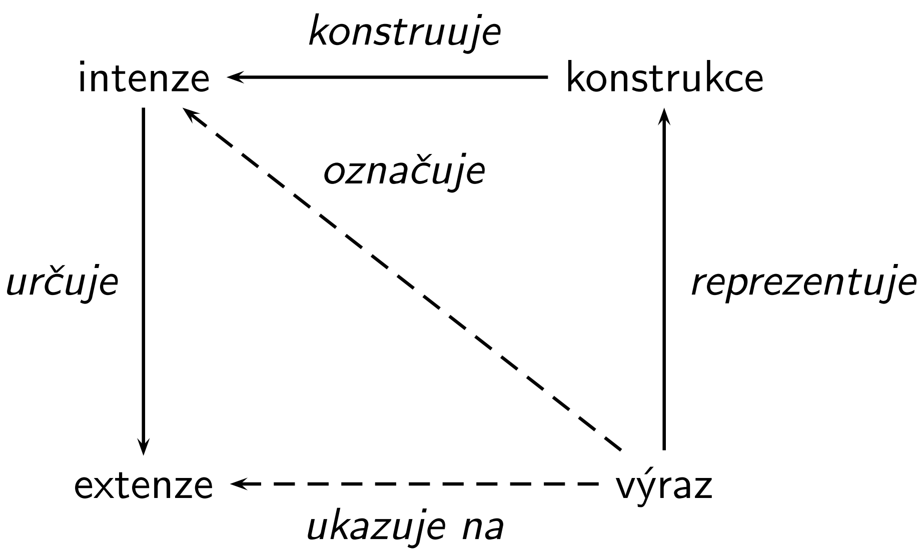
- intenze – objekty typu funkcí, jejichž hodnoty závisí na světě a čase
- extenze – ostatní objekty (na světě a čase nezávislé)

časté extenze a intenze:

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

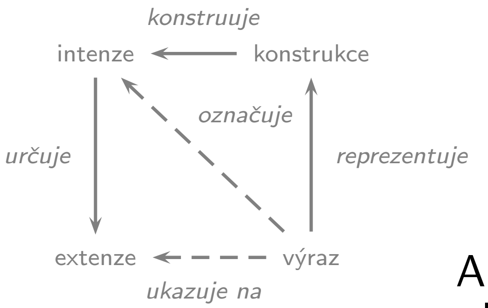
Rozšířený vztah výrazu a významu u intenzí

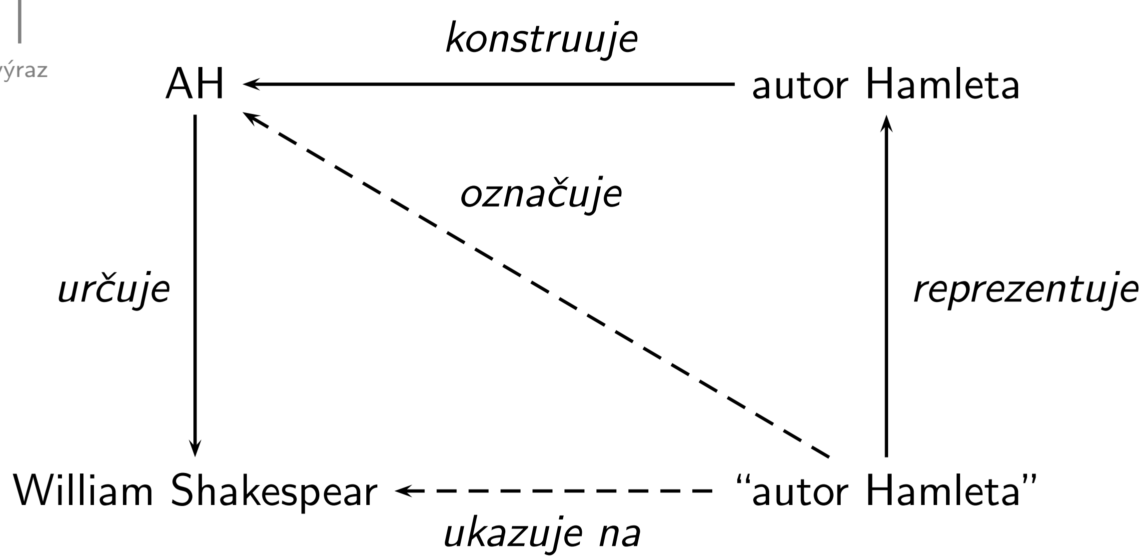




Intenzionální sémantika   Extenze a intenze

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)
požadavky na definici "možného světa:"

- soubor **myslitelných faktů**
- je **konzistentní** a **maximální** ze všech takových souborů
- je **objektivní** (nezávislý na individuálním názoru)

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

Možné světy v TILu

možný svět v TILu = *rozhodovací systém*, pro ∀ prvek intenzionální báze obsahuje **konzistentní přiřazení** hodnot
příklad – realita s 2 objekty a 2 vlastnostmi (9 možných světů):

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

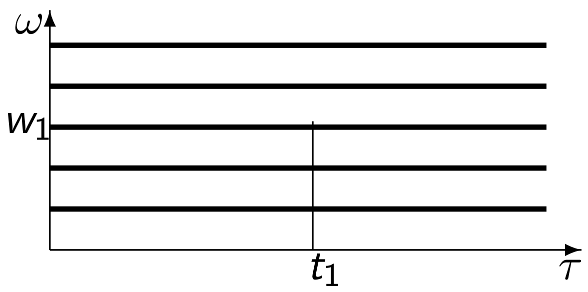
Princip intenzí v TILu

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
  $x \ldots \iota$
- trivializace objektu **A** typu α, konstruuje právě objekt **A**
  $^0\mathrm{A} \ldots \alpha$ $\qquad$ $\mathbf{A} \ldots \alpha$
- aplikace konstrukce $X \ldots (\alpha\beta_1 \ldots \beta_n)$ na konstrukce $Y_1, \ldots, Y_n$ typů $\beta_1, \ldots, \beta_n$, konstruuje objekt typu α
  $[XY_1 \ldots Y_n] \ldots \alpha$
- abstrakce konstrukce $Y \ldots \alpha$ na proměnných $x_1, \ldots, x_n$ typů $\beta_1, \ldots, \beta_n$, konstruuje objekt/funkci typu $(\alpha\beta_1 \ldots \beta_n)$
  $\lambda x_1 \ldots x_n[Y] \ldots (\alpha\beta_1 \ldots \beta_n)$

Příklady analýzy podstatných jmen

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

  $$(*) \quad \lambda w\lambda t\left[{}^0\neg[Ex_{wt}\,\mathbf{jednorožec}]\right], \quad Ex \stackrel{df}{=} \lambda w\lambda t\lambda p\left[{}^0\textstyle\sum_{\iota}\left[\lambda x[p_{wt}\,x]\right]\right]$$

  $$Ex \ldots (o(o\iota)_{\tau\omega})_{\tau\omega}$$

  $(*)$ … "třída všech individuí s vlastností 'být jednorožcem' je v daném světě a čase prázdná."

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

otázka:

Jak zapíšeme znalosti o problému/doméně?
Když je zapíšeme, můžeme z nich mechanicky odvodit nová fakta?

- **reprezentace znalostí** (*knowledge representation*) – hledá způsob vyjádření znalostí počítačově zpracovatelnou formou (za účelem odvozování)
- **vyvozování znalostí** (*reasoning*) – zpracovává znalosti uložené v **bázi znalostí** (*knowledge base, KB*) a provádí **odvození** (inference) nových závěrů:
  - odpovědi na dotazy
  - zjištění faktů, které vyplývají z faktů a pravidel v KB
  - odvodit akci, která vyplývá z dodaných znalostí, ...

Reprezentace znalostí

proč je potřeba speciální **reprezentace znalostí**?

vnímání lidí × *vnímání počítačů*

- člověk
  - když dostane novou věc (třeba pomeranč) – **prozkoumá** a **zapamatuje** si ho (a třeba sni)
  - během tohoto procesu člověk zjistí a uloží všechny základní vlastnosti
  - později, když se **zmíní** daná věc, vyhledají se a připomenou uložené informace
- počítač
  - musí se spolehnout na informace od lidí
  - jednodušší informace – přímé *programování*
  - složité informace – zadané v **symbolickém jazyce**

Volba reprezentace znalostí

která **reprezentace znalostí** je **nejlepší**?

> *Pro řešení skutečně obtížných problémů musíme použít několik různých reprezentací. Důvodem pro to je to, že každý typ datových struktur má své přínosy i nedostatky a žádná z nich není adekvátní pro všechny různé funkce používané v tom, čemu říkáme "zdravý rozum" (common sense).*
>
> *– Marvin Minsky*

Reprezentace znalostí pomocí logiky nebo datových struktur

Logika:

- znalosti uloženy ve formě **logických formulí**
- vyvozování nových znalostí = hledání **důkazu**

Specializované datové struktury:

- sémantické sítě
- rámce
- pravidlové systémy
- struktury pro práci s nejistotou a pravděpodobností

Sémantické sítě

- reprezentace faktových znalostí (pojmy + vztahy)
- znalosti jsou uloženy ve formě grafu
- nejdůležitější vztahy:
  - **podtřída** (*subclass*) – vztah mezi třídami
  - **instance** – vztah mezi konkrétním objektem a jeho rodičovskou třídou

  jiné vztahy – část (has-part), barva, ...

Rámce

Rámce (*frames*):

- varianta sémantických sítí
- velice populární pro reprezentaci znalostí v expertních systémech
- všechny informace relevantní pro daný pojem se ukládají do univerzálních struktur – **rámců**
- stejně jako sémantické sítě, rámce podporují dědičnost
- OO programovací jazyky vycházejí z teorie rámců

Rámce – příklad

rámec obsahuje **objekty**, *sloty* a hodnoty slotů
příklady rámců:

savec:		
	podtřída:	zvíře
	část:	hlava
	**má_kožich*:	ano
slon:		
	podtřída:	savec
	**barva*:	šedá
	**velikost*:	velký
Nellie:		
	instance:	slon
	mít_rád:	jablka

'*' označuje **vzorové hodnoty**, které mohou měnit hodnoty u podtříd a instancí

Sémantické sítě × rámce

sémantické sítě	**rámce**
uzly	objekty
spoje	sloty
uzel na druhém konci spoje	hodnota slotu

deskripční logika – logický systém, který manipuluje přímo s rámci

Pravidlové systémy

- snaha zachytit **produkčními pravidly** znalosti, které má expert
- obecná forma pravidel

  IF *podmínka*
  THEN *akce*

  - podmínky – booleovské výrazy, dotazy na hodnoty **proměnných**
  - akce – nastavení hodnot proměnných, příznaků, ...
- důležité vlastnosti:
  - znalosti mohou být strukturovány do modulů
  - systém může být snadno rozšířen přidáním nových pravidel beze změny zbytku systému

Metody pro práci s nejistotou

definujme akci A_t jako "Vyrazit na letiště t hodin před odletem letadla."
jak najít odpověď na otázku "*Dostanu se akcí A_t na letiště včas?*"

- defaultní/nemonotónní logika

  Předpokládejme, že nepíchnu cestou kolo.
  Předpokládejme, že A_5 bude OK, pokud se nenajde protipříklad.
- pravidla s faktory nejistoty

  $A_5 \mapsto_{0.3}$ dostat se na letiště včas.
  zalévání $\mapsto_{0.99}$ mokrý trávník
  mokrý trávník $\mapsto_{0.7}$ déšť
- pravděpodobnost

  Vzhledem k dostupným informacím, A_3 mě tam dostane včas s pravděpodobností 0.05.
  Použití **náhodných proměnných** a pravidel pro výpočet pravděpodobnosti logicky souvisejících událostí (podmíněná pravděpodobnost, bayesovské pravidlo, ...)