

# PA153 Počítačové zpracování přirozeného jazyka

## 11 – Znalosti, parafráze, odvozování

Karel Pala, Zuzana Nevěřilová

Centrum ZPJ, FI MU, Brno

30. listopadu 2017

- 1 Znalosti
- 2 Parafráze
- 3 Přirozená logika
- 4 Belief–Desire–Intention

# Znalosti a odvozování

- znalosti o jazyce (lexikon, gramatické kategorie, syntax)
- znalosti o světě

Znalostní báze (knowledge base, KB): obsahuje fakta, která jsou premisami v deduktivním odvozování

# Znalosti a odvozování

- znalosti o jazyce (lexikon, gramatické kategorie, syntax)
- znalosti o světě

Znalostní báze (knowledge base, KB): obsahuje fakta, která jsou premisami v deduktivním odvozování

lidmi čitelné KB: how-to, FAQ, recepty, návody, diagramy

strojově čitelné KB: ontologie (SUMO-MILO), sémantické sítě (WordNet), dbPedia, ConceptNet

# Znalosti a odvozování

- znalosti o jazyce (lexikon, gramatické kategorie, syntax)
- znalosti o světě

Znalostní báze (knowledge base, KB): obsahuje fakta, která jsou premisami v deduktivním odvozování

lidmi čitelné KB: how-to, FAQ, recepty, návody, diagramy

strojově čitelné KB: ontologie (SUMO-MILO), sémantické sítě (WordNet), dbPedia, ConceptNet

Reprezentace znalostí (knowledge representation): znalostní báze + odvozovací pravidla

# Deklarativní vs. procedurální znalost

Deklarativní (formálně verifikovatelná, obecně platná) vs. procedurální (implicitní, méně obecná)

Příklad: robot, který se umí pohybovat po budově

procedurální znalost: „dojdi do místnosti“

deklarativní znalost: mapa objektu + základní kroky

# Deduktivní odvozování: monotónní a nemonotónní odvozování [Allen, 1995]

KB: Ptáci létají. Vrabec je pták.

---

Vrabec létá.

# Deduktivní odvozování: monotónní a nemonotónní odvozování [Allen, 1995]

KB: Ptáci létají. Vrabec je pták. Pštros je pták.

---

Vrabec létá. Pštros létá.



# Deduktivní odvozování: monotónní a nemonotónní odvozování [Allen, 1995]

KB: Ptáci létají. Vrabec je pták. Pštros je pták. Pštros nelétá.

---

Vrabec létá. ~~Pštros létá.~~

# Znalosti o světě

- encyklopedické (Jaké je hlavní město ČR?)
- common-sense (Jak je vhodné obléci se 30. listopadu 2017?)

neostrá hranice

počítačově zpracovatelné zdroje encyklopedických znalostí:

- encyklopedie
- znalostní hry
- dbPedia: strojově zpracovaná Wikipedie

# Common sense a odvozování

common sense: sdílená znalost, ne vždy v souladu s (vědeckými) fakty  
(V noci nesvítí slunce.)

## Common sense a odvozování

common sense: sdílená znalost, ne vždy v souladu s (vědeckými) fakty  
(V noci nesvítí slunce.)

Cheap apartments are rare.

Rare things are expensive.

---

Cheap apartments are expensive.

Deduktivní odvozování není možné použít vždy (ve skutečnosti skoro nikdy).

## Common sense: nejznámější projekty

- CyC: vývoj od r. 1985(!), reprezentace pomocí vlastního jazyka CyCL, mikroteorie
- ConceptNet: syntaktická analýza OpenMind, propojení s Wiktionary
- Never-ending Language Learning (NELL): prochází web a odvozuje, občas nutný lidský zásah (“I deleted my Internet cookies”, “I deleted my files” ⇒ soubor je stejná kategorie jako pečivo)

# Parafráze

Parafráze: promluva  $x$  je parafrází promluvy  $y$ , pokud  $x$  a  $y$  mají stejný nebo podobný význam.

Tento most postavila Nejlepší firma s.r.o.

Nejlepší firma s.r.o. postavila tento most.

Stavitelem tohoto mostu je Nejlepší firma s.r.o.

# Přesnější definice

## Logické vyplývání

Věta (závěr) logicky vyplývá z množiny vět (premis) právě tehdy když při všech pravdivostních hodnotách které splňují premisy splňují i závěr.

## Textové vyplývání $\neq$ logické vyplývání

Z textu  $t$  textově vyplývá hypotéza  $h$  ( $t \Rightarrow h$ ), pokud lidé, kteří přečtou  $t$ , odvodí, že  $h$  je nejspíš pravda. [Dagan et al., 2007]

parafráze =  $h \Rightarrow t \wedge t \Rightarrow h$

# Rozpoznávání textových vyplývání/parafrází

hledání podobností:

- na řetězcích (např. Levenshteinova vzdálenost)
- na slovech
- na slovech s použitím znalostní báze (např. slovník synonym)
- na syntaktických stromech
- kombinace předchozích



# Rozpoznávání textových vyplívání/parafrází

využití:

- odpovídání na otázky
- chatbots
- detekce plagiátů
- výuka
- automatická sumarizace textu
- doplnění implicitní znalosti
  - ▶ logická analýza textu
  - ▶ znalostní modely v umělé inteligenci
- ...

# Korpusy parafrází

- Microsoft Research Paraphrase Corpus<sup>1</sup>
- The Boeing-Princeton-ISI (BPI) Textual Entailment Test Suite<sup>2</sup>
- Multiple Translation Chinese Corpus<sup>3</sup>
- The SEMILAR Corpus: The SEMantic SIMILARity Corpus<sup>4</sup>
- Paraphrase Discovery<sup>5</sup>

---

<sup>1</sup><http://research.microsoft.com/en-us/downloads/607d14d9-20cd-47e3-85bc-a2f65cd28042/>

<sup>2</sup><http://www.cs.utexas.edu/users/pclark/bpi-test-suite/>

<sup>3</sup><https://catalog.ldc.upenn.edu/LDC2002T01>

<sup>4</sup><http://deeptutor2.memphis.edu/Semilar-Web/public/semilar-api.html>

<sup>5</sup><http://nlp.cs.nyu.edu/paraphrase/>

# Paraphrase Discovery

vztahy mezi pojmenovanými entitami v korpusových datech:

```
[lemma="Hannibal"] []* [lemma="Hopkins"] within <s/>
```

ztvárnit	jako
hrát	odmítnout
s	na roli
si	hrající
/	se objevil
v podání	představoval
alias	působí v roli
se svým přítelem	
(	
po boku	

# Generování parafrází

Základní způsoby parafrázování:

- aktivní–pasivní větná konstrukce: Tento most byl postaven Nejlepší firmou s.r.o.
- synonyma: Tuto lávku postavila Nejlepší firma s.r.o.
- hyperonyma: Tuto stavbu postavila Nejlepší firma s.r.o.
- substantivizace, deverbalizace: Stavitelem tohoto mostu je Nejlepší firma s.r.o.
- kombinace: Tento most byl vytvořen Nejlepší firmou s.r.o.

Podrobněji v [Bhagat and Hovy, 2013].

# Přirozená logika [Lakoff, 1970]

nástrojem této logiky je přirozený jazyk

- **monotonicita (monotonicity):** víc než tisíc je hodně  
Mám víc než tisíc knih. Mám hodně knih.  
Nemám víc než tisíc knih. Nemám hodně knih.
- **obsažení/omezení (containment):** červené auto je auto  
Po ulici jelo červené auto. Po ulici jelo auto.  
Po ulici nejelo červené auto. Po ulici nejelo auto.
- **exkluze (exclusion):** pes není kočka  
Po ulici šel pes.  $\Rightarrow$  Po ulici nešla kočka.

odvození vs. presupozice:

Mark David Chapman zastřelil Johna Lennona.  $\Rightarrow$  John Lennon nežije.

Brazílie vyhrála mistrovství světa.  $\Rightarrow$  Brazílie hrála na mistrovství světa.

# Přirozená logika [Lakoff, 1970]

nástrojem této logiky je přirozený jazyk

- **monotonicita (monotonicity):** víc než tisíc je hodně  
Mám víc než tisíc knih. Mám hodně knih.  
Nemám víc než tisíc knih. Nemám hodně knih.
- **obsažení/omezení (containment):** červené auto je auto  
Po ulici jelo červené auto. Po ulici jelo auto.  
Po ulici nejelo červené auto. Po ulici nejelo auto.
- **exkluze (exclusion):** pes není kočka  
Po ulici šel pes.  $\Rightarrow$  Po ulici nešla kočka.

odvození vs. **presupozice:**

Mark David Chapman zastřelil Johna Lennona.  $\Rightarrow$  John Lennon nežije.

Mark David Chapman nezastřelil Johna Lennona.  $\nRightarrow$  John Lennon nežije.

Brazílie vyhrála mistrovství světa.  $\Rightarrow$  Brazílie hrála na mistrovství světa.

Brazílie nevyhrála mistrovství světa.  $\Rightarrow$  Brazílie hrála na mistrovství světa.

## BDI: Znalost nebo domněnka?

KB: Ptáci létají. Vrabec je pták. Pštros je pták. Pštros nelétá. Mrtvý vrabec nelétá.

## BDI: Znalost nebo domněnka?

KB: Ptáci létají. Vrabec je pták. Pštros je pták. Pštros nelétá. Mrtvý vrabec nelétá.

Znalostní báze se mění. Některé znalosti mají poměrně krátké trvání (Nejsem unavená. Je půl čtvrté.)

V umělé inteligenci se používá termín domněnka (belief) [Mařík et al., 2001].



# Umělá inteligence: modely uvažování inteligentních agentů

**Intencionální systém:** agent umí „uvažovat“ o svých znalostech. Je schopen přemýšlet o svých přáních a jak jich lze dosáhnout [Mařík et al., 2001].

**Mentální postoje:**

- informační postoje – znalosti, fakta získaná senzory
- proaktivní postoje – cíle, plány, závazky

Psychologické modely lidského uvažování [Bratman, 1987]: kognitivní stavy, afektivní stavy, konnativní stavy.

Domněnka–přání–záměr: softwarový model pro aktivní inteligentní agenty

# Umělá inteligence: belief–desire–intention

## Záměr, Intention

Aby bylo možné vytvořit aktivního agenta, je třeba, aby „věděl, co chce“ (intention). Pokud ví, co chce (tj. má **záměr**), vytvoří si agent nějaký **plán** (lokální cíl).

Příklad: najdi cestu z domu  $X$  na FI

$Int\ a\ \phi$  agent si vybírá vždy cesty tak, aby na nich někdy platila  $\phi$

## Přání, Desire

**Přání** vyjadřuje agentovu motivaci. Motivovaný agent má **cíle** (cílové stavy). Cíle by neměly být v rozporu.

Příklad: najdi nejkratší cestu z domu  $X$  na FI

$Des\ a\ \phi$  pravdivost formule  $\phi$  je cílem agenta  $a$





## Belief, Domněnka

**Domněnka** představuje agentovu **bázi znalostí**. Informace mohou být pravdivé, agent v ně v daný okamžik věří a chápe je jako nedokonalé přiblížení obrazu okolního světa [Mařík et al., 2001].

Příklad: najdi nejkratší cestu z domu na FI. Mostecká je neprůjezdná.

$Bel\ a\ \phi$  agent  $a$  věří v pravdivost formule  $\phi$

# Odkazy I

-  Allen, J. (1995).  
*Natural Language Understanding (2nd ed.)*.  
Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA.
-  Bhagat, R. and Hovy, E. (2013).  
What is a paraphrase?  
*Computational Linguistics*, 39(3):463–472.
-  Bratman, M. (1987).  
*Intention, plans, and practical reason*.  
Harvard University Press.
-  Dagan, I., Roth, D., and Zanzotto, F. M. (2007).  
Tutorial notes.  
In *45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic. The Association of Computational Linguistics.

## Odkazy II



Lakoff, G. (1970).  
Linguistics and natural logic.  
*Synthese*, 22(1-2):151–271.



Mařík, V., Štěpánková, O., and Lažanský, J. (2001).  
*Umělá inteligence*.  
Number svazek 3 in *Umělá inteligence*. Academia.



Pease, A. (2011).  
*Ontology: A Practical Guide*.  
Articulate Software Press.



Schank, R. C. and Abelson, R. P. (1977).  
*Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures (Artificial Intelligence)*.  
Lawrence Erlbaum Associates, 1 edition.  
Published: Hardcover.

## Odkazy III



Smith, B. (1995).

Formal ontology, common sense and cognitive science.

*International Journal of Human-Computer Studies*, pages 641–667.



Wasserman, K. (1985).

Physical object representation and generalization: A survey of programs for semantic-based natural language processing.

*AI Magazine*, 5(4):28–42.