PLIN021 Sémantická analýza v praxi

OP VK Mezi bohemistikou a informatikou
www.projekt-inova.cz

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

12. listopadu 2013

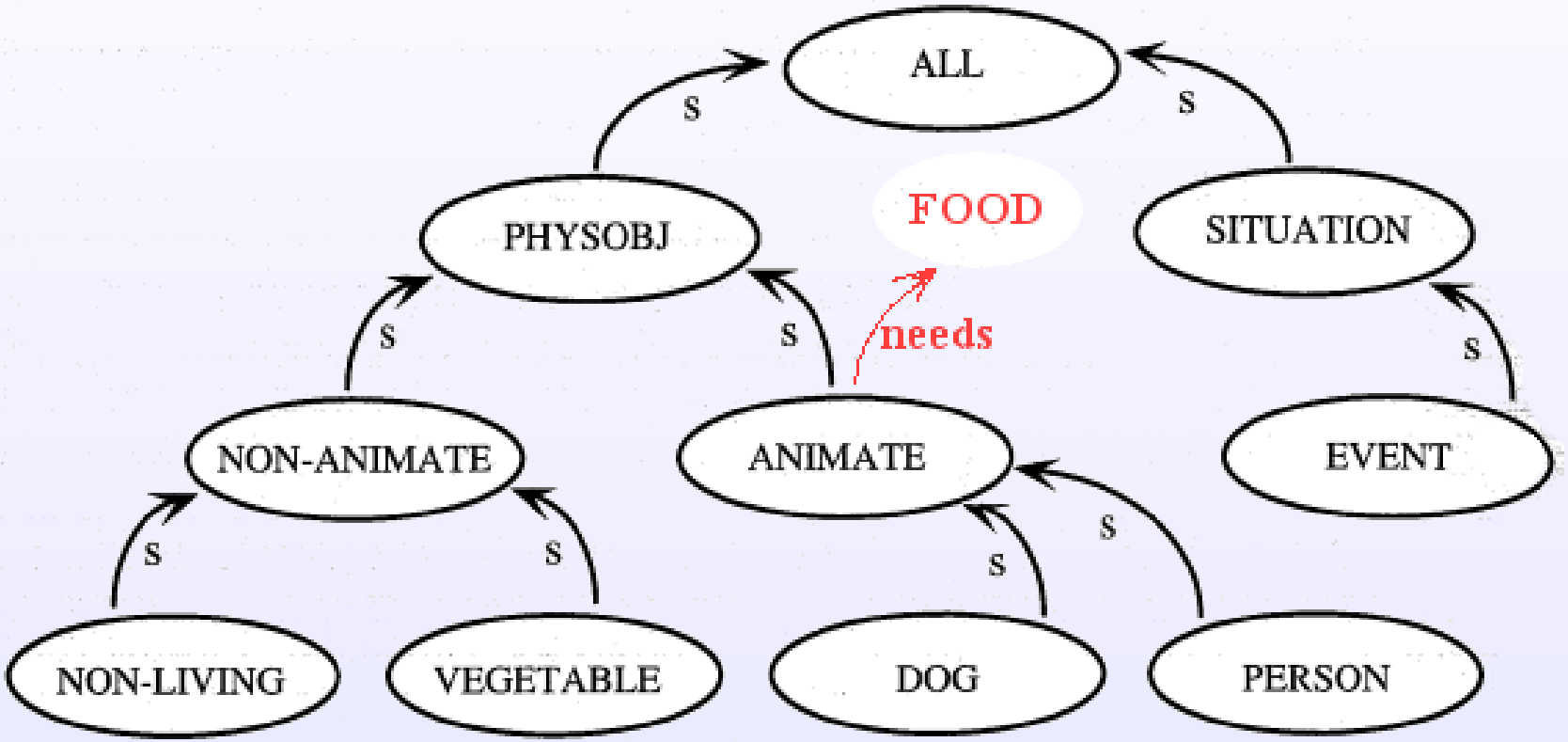
Sémantické rámce

SUMO/MILO

dbPedia

Skripty

Sémantické sítě – dědičnost



odvozování je monotónní

Praktická ukázka dědičnosti a odvozování:

1. silniční vozidlo má (has part) volant
2. dodávka je (isa) silniční vozidlo
3. dodávka má (has part) volant
4. Mercedes Sprinter je (member of) dodávka
5. Mercedes Sprinter má (has part) volant

Asociativní sítě

Někdy totožné se sémantickými sítěmi, jindy u asociativních sítí neplatí předpoklad, že jeden koncept odpovídá jednomu uzlu.

Rámce – použití

Rámce můžeme použít pro desambiguaci slov i celých vět [Laparra and Rigau, 2009].

[Bernard Lansky]$_{STUDENT}$ ***studied*** [the piano]$_{SUBJECT}$ [with Peter Wallfisch]$_{TEACHER}$.

Rámce – použití

Rámce můžeme použít pro doplnění implicitní (nezmiňované) znalosti.

Koupila jsem ojetou felicii. Byly to vyhozené peníze.

koupit:

- má_činitele člověk/instituce/skupina
- má_benefaktora člověk/instituce/skupina
- má_předmět výrobek/nemovitost/zvíře/rostlina/přírodnina
- má_část činitel dá peníze
- má_část benefaktor dá předmět

Suggested Upper Merged Ontology (SUMO)

- „největší volná (formální) ontologie“ (cca 25 000 konceptů a 80 000 axiomů)
- SUMO – „horní“ vrstva
- MId-Level Ontology (MILO) – „střední“ vrstva
- specializované domény (vojenství, ekonomie, geografie apod.) jsou pokryté jinými ontologiemi

Suggested Upper Merged Ontology (SUMO)

ukázka z www.ontologyportal.org

dbPedia

http://cs.dbpedia.org

- sémantická síť automaticky vytvořená z Wikipedie
- od března 2012 i česká verze (zatím ve zrodu)
- jak získat vlastnosti (Property) a jejich hodnoty (Value)?
- co je uzel?
- co je hrana?
- synonymie?

Skripty, scénáře (Abelson)

skript: v restauraci, prvky skriptu mohou být rámce

- host (člověk, není v zaměstnání, má u sebe *peníze*, sedí na *židli*, jí *jídlo*)
- číšník (člověk, je v zaměstnání)
- kuchař (člověk, je v zaměstnání)
- místnost (obsahuje *židle*, stoly, příjemnou teplotu)
- jídlo (uvařil *kuchař*, donesl *číšník hostovi*)
- peníze (zaplatil *host číšníkovi* za *jídlo*)

„Pepovi u večeře zazvonil telefon. Chvíli poslouchal, pak položil telefon a opustil restauraci.“

Skripty, scénáře (Abelson)

„Pepovi u večeře zazvonil telefon. Chvíli poslouchal, pak položil telefon a opustil restauraci.“

Předpokládáme, že mezi „položil telefon“ a „opustil restauraci“ se stalo:

- Číšník donesl účet.
- Pepa zaplatil.
- Pepa vstal ze židle.
- Pepa se oblékl.

Skripty, scénáře (Abelson)

skript: v restauraci

Usuzování v rámcích může být implicitní (podobné jako v sém. sítích) i speciální pro daný rámec.

Usuzování v rámcích může být nemonotónní.

Příklad: host zaplatil $\Rightarrow$ číšník má u sebe peníze

Příklad: každý host musí zaplatit svoji útratu.
host zaplatil útratu za jiného hosta $\Rightarrow$ jiný host nemusí platit svoji útratu

Skripty, scénáře (Abelson)

skript: v restauraci

Skripty popisují typické situace. Stereotypická je i informace o zaplněnosti slotů, např. restaurace musí mít číšníka.

Pořadí ve scénáři je chronologické: host přijde do restaurace, objedná si jídlo, kuchař jídlo uvaří, host sní jídlo, host zaplatí číšníkovi. . .

Můžeme nějak měřit vybočení ze stereotypu?

Laparra, E. and Rigau, G. (2009).
Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm.
In *RANLP*, Borovets, Bulgaria.