

PLIN037 Sémantika a počítače

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

11. dubna 2019

Reprezentace znalostí

Automatické použití znalostníchází

Reprezentace znalostí

K sémantické analýze potřebujeme znalost jazyka a znalost světa.

Znalost světa je dostupná ve znalostních bázích:

- lexikální síť (wordnety)
- valenční slovníky
- doménové znalosti (doménové ontologie)
- encyklopedické znalosti (DBPedia)
- common-sense (ConceptNet, CyC)
- metadata (DublinCore)

Znalostní báze

Původně manuálně vytvářené zdroje → velice drahé

V současnosti převažuje snaha:

- vytvářet znalostní báze (polo)automaticky z dat
- vytvářet znovupoužitelné zdroje
- mít ve zdrojích konzistentní data s jednotnou granularitou
- mít data uložena ve standardizovaném formátu
- moci data propojit s ostatními znalostními bázemi

- DBPedia lookup – výstupem je XML
- ConceptNet API – výstupem je JSON

JSON lze snadno přečíst (deserializovat) v Pythonu jako slovník (`dict`)

