

# PLIN021 Sémantická analýza v praxi

OP VK Mezi bohemistikou a informatikou  
[www.projekt-inova.cz](http://www.projekt-inova.cz)

Zuzana Nevěřilová  
[xpopelk@fi.muni.cz](mailto:xpopelk@fi.muni.cz)

Centrum zpracování přirozeného jazyka, B203  
Fakulta informatiky, Masarykova univerzita

11. dubna 2012

dbPedia

WordNety

VerbaLex

SUMO/MILO

Skripty

# 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?

# Princeton WordNet (PWN)

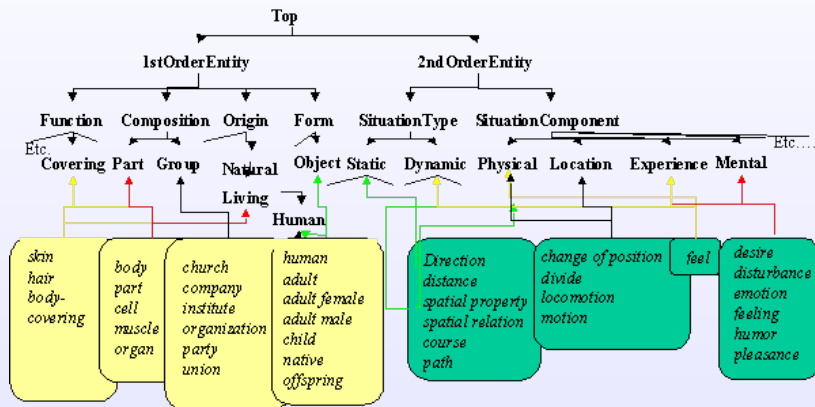
- verze 1.5, 1.6, 1.7
- verze 2.0, 2.1
- verze 3.0

## EuroWordNet (EWN): rozdíly oproti PWN

- každý národní WordNet budovaný nezávisle na ostatních jazycích
- každý synset v národních WordNetech je propojený s PWN pomocí InterLingual Index (ILI)
- jednotlivé stromy jsou spojeny pomocí vrcholové ontologie, která je společná všem WordNetům

# EuroWordNet (EWN)

## Vrcholová ontologie (Top ontology)



[Association, 2012]

# EuroWordNet (EWN): vrcholová ontologie

- entity 1. řádu – konkrétní objekty a substance
- entity 2. řádu – stavy, situace, události
- entity 3. řádu – myšlenky, znalosti

# EuroWordNet (EWN): základní koncepty

## Základní koncepty (Base Concepts)

- 66 konkrétních synsetů (podstatná jména)
- 98 abstraktních synsetů (63 podstatných jmen a 35 sloves)



## EuroWordNet (EWN), BalkaNet a další

WordNety pro jiné jazyky než angličtina (ukázka)

<http://www.globalwordnet.org>

# VerbaLex

Pokrytí VerbaLexu a propojení s WordNetem:

- V.: 19 158 rámců
- V. odkazuje na 811 různých synsetů ve WN
- není to zbytečně moc?
- jak ověřit, že rámec odkazuje na ten správný synset?

## 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](http://www.ontologyportal.org)

## 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. Chvilí poslouchal, pak položil telefon a opustil restauraci.“

## Skripty, scénáře (Abelson)

„Pepovi u večere 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 se oblékl.

## Skripty, scénáře (Abelson)

skript: v restauraci

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

Usuzování v rámci 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?





Association, T. G. W. (2012).

Eurowordnet top ontology.

[online; accessed 2012-04-11 <[http://www.globalwordnet.org/gwa/ewn\\_to\\_bc/ewnTopOntology.htm](http://www.globalwordnet.org/gwa/ewn_to_bc/ewnTopOntology.htm)>].