

Sémantická analýza

CJBB85

Mgr. Dana Hlaváčková, Ph.D.

Sémantická analýza

- sémantika
- formální popis významů
- **ontologie**, koncepty
 - mělká (*shallow*)
 - strukturovaná, hierarchická
 - **vrcholová** (*top ontology, upper ontology*)
 - SUMO (*Suggested Upper Merged Ontology*)
 - **doménová**
 - FOAF (*Friend of a Friend*)

WordNet – motivace

- **G. A. Miller** (*Princeton University*) – psycholog a psycholinguista, psycholexikologie
- uspořádání významů v lidské lexikální paměti
- hierarchie
- experimenty – asociační testy, schopnost zpracovávat anaforické výrazy
- <http://wordnet.princeton.edu>

WordNet – struktura

- model lexikální paměti, sémantická síť
- **synset** – synonymická řada (blízká synonyma), **literál** + číslo významu
- substantiva, adjektiva, verba, adverbia
- **hierarchická struktura** – hyperonyma, hyponyma, kohyponyma
- substantiva – tematické hierarchie
- verba – vztah vyplývání
- další **sémantické vztahy** – antonyma, holonyma, meronyma, domény ontologií SUMO, MILO
- derivační vztahy

WordNet – projekty

- **Princeton WordNet** (1990–1995) – americká angličtina, G. A. Miller, Christiane Fellbaum
- **EuroWordNet** – Piek Vossen, *University of Amsterdam*
 - **EWN I** (1997–1998) – *angličtina, holandština, italština, španělština*
 - **EWN II** (1998–1999) – *francouzština, němčina, čeština, estonština*
- **Balkanet** (2001–2004), D. Christodoulakis, **University of Patras** – *turečtina, rumunština, řečtina, bulharština, srbština, čeština*

EuroWordNet

- **Base Concepts** – jádro slovní zásoby (cca 1000 synsetů)
- **Top-Ontology** – 63 konceptů
 - entity 1. řádu = objekty
 - entity 2. řádu = stavy a procesy
 - entity 3. řádu = abstraktní pojmy (množina)
- **Interlingual Index** – číslo, které propojuje významy v jednotlivých wordnetech

- Global WordNet Association,
<http://www.globalwordnet.org>
- nástroje, prohlížeče
- VisDic, DEBVisDic