

Centrum NLP FI MU 2008

- Potřeba kvalitních (lexikálních) zdrojů – lexikální databáze VerbaLex
- Dokončování lexikální databáze VerbaLex
- Kontroly konzistence val.rámců, pokrytí – 68 %
- 6360 synsetů, 10 482 lemmat (infinitivů)
- 21 193 slovesných významů
- 19 556 komplexních valenčních rámců (VLXall)
- VerbaLex existuje jako txt, XML, HTML a PDF
- Linky na PWN – nástroj WNAssistant
- Sémantické třídy – 82, domény

II

- globální sítě národních sémantických sítí typu WordNet, tzv. Global WordNet Grid (GWG)
- Nástroje pro práci s nimi – platforma DEBII
- Modifikace nástroje DebVisDic (editor, browser)
- Kooperace s PWN (Ch. Fellbaum)
- Kooperace ve vazbě na EU projekt KYOTO (P. Vossen)
- Provedena reimplementace nástroje DebVisDic pro Firefox 3
- GWG nyní obsahuje 5 národních WN (50)

III

- Práce na jazykové ontologii pro analýzu kontextových vzorců (angličtina)
- 600 ang. sloves – něco přes 3 000 vzorců
- Mělká ontologie pro CPA
- Nový přístup k analýze metaforických významů
- Možná je to příliš, odvážné
- Formálně popis významů anglických sloves pro počítačové zpracování
- Vztahy ke korpusovým datům

IV

- Automatické budování ontologií – oblast biomediciny
- Extrakce taxonomií z volného textu
- Shlukování konceptů
- Dynamická integrace z volných textů
- Úložiště článků z oblasti mediciny obsahující 18 databází

V

- Ve volné vazbě na projekt byla organizována konference TSD 2008
- Sborník konference obsahuje publikace relevantní k projektu sém. webu
- Konference byla finančně soběstačná.

VI

- Výsledky:
- Databáze: Verbalex, Cz. Wordnet – vztahy k Princetonskému WordNetu (pracnost)
- sw. nástroje – DebVisdic, DEB II, editor Verbalexu, Wnassistant, výstupy z Word Sketch Engine
- Uplatnitelnost: syntaktická a sémantické analýza PJ, vyhledávání, extrakce informací
- Ontologie, jejich automatické budování, rámce jako metadata
- Pokusů o aplikace – mohlo být víc

VII

- Od začátku projektu: 107 publikací – spolu s prof. Zezulou
- Publikace za r. 2008: 19 a 16