Visualizing sloWNet

Darja Fišer
University of Ljubljana

Jernej Novak
University of Maribor
• What is sloWNet
• Motivation & Background for this research
• Presentation of sloWTool
• Conclusions
• Future plans

Presentation outline
sloWNet: WordNet for Slovene

- Based on Princeton WordNet (Fellbaum 1998)
  - words grouped into sets of synonyms, each expressing a concept
  - synsets are interlinked with semantic and lexical relations
  - 118,000 synsets, 157,000 literals, 207,000 lit-syn pairs
- Developed automatically from heterogeneous resources
  - bilingual dictionary
  - parallel corpus
  - Wikipedia
  - partially manually revised
  - used for semantic annotation of jos100k corpus
- sloWNet 3.0:
  - 43,000 synsets, 39,700 literals, 82,700 syn-lit pairs
  - average synset length 1.92, average polysemy 2.07
Motivation & Background

**Dictionaries**
- word-based
- semantic info implicit
- for human users

**Semantic Lexicons**
- concept-based
- explicit encoding of semantic info
- still language-oriented
- intended for human as well as machine use

**Ontologies**
- highly formalized
- compiled for reasoning in AI tasks

**Dictionary-writing systems**
(e.g. TshwaneLex)

**Wordnet browsers, editors, visualizers**
(which tool to use?)

**Ontology-management tools**
(e.g. Protégé)
available tools:
- Princeton WordNet: Princeton WordNet Browser
- EuroWordNet: Polaris & Periscope
- BalkaNet: DEBVisDic

ideal tool:
- freely available, platform-independent, on-line
- all-in-one tool (browsing, editing & visualization)
- support for standardized formats (e.g. LMF)
- support for multilingual scenarios
- easy integration of third-party resources (e.g. domains, coarse-grained sense clusters, images)

Analysis of existing wn tools
Advanced search

- **Language**: Slovenian
- **Literals**: prs*
- **Part of speech**: Noun
- **Synset ID**: 
- **Definitions**: 
- **Examples**: 
- **Stamps**: Empty
- **BCS**: 1, 2

Wildcards * ?
POS: Noun  ID: eng-30-05297523-n  BCS: 1  DOMAIN: anatomy

SYNONYM (SLV): organ, orgle

SYNONYM (ENG): organ

DEFINITION: a fully differentiated structural and functional unit in an animal that is specialized for some particular function

→ [ENG_DERIVATIVE]: organski, organic
→ [HYPERNYM]: del telesa, body part

STAMP: darja 2008-01-01 00:00:00
Visualization
• Advantages of sloWTool:
  • all-in-one browser, editor & visualisation tool for wordnets
  • can be used for wordnets in other languages as well as for other similar databases
  • allows simple or advanced search & view settings
• External resources integrated in sloWTool:
  • semantically annotated corpus josSENSE (Fišer&Erjavec 2010)
  • WordNet Domains Hierarchy (Bentivogli et al. 2004)
  • Coarse-grained Sense Clusters (Navigli 2006)
• availability of sloWTool:
  • for browsing: http://kista.asuscomm.com/search
  • for download: https://launchpad.net/slowtool

Conclusions
Future plans

• sloWTool enhancements:
  • enhance the graph-based visualization module
  • extend sloWTool to allow semantic annotation of corpora
  • extend sloWTool for crowd-sourcing sloWNet refinement

• integration of external resources:
  • add wordnets for other languages (e.g. WOLF)
  • add Princeton semantically annotated glosses
  • integrate with ImageNet database
  • add links to GeoWordNet
Thank you for your attention!

general questions: darja.fiser@ff.uni-lj.si
technical questions: jernej.novak1@uni-mb.si