UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth
http://www.ukp.tu-darmstadt.de/uby/

UBY in a Nutshell
Developing a large-scale unified lexical-semantic resource (LSR) in a standardized format for NLP

Motivation:
1. Problem: Limited resource coverage in NLP
   → Solution: Integration of LSRs
2. Problem: Incompatible formats of (integrated) LSRs
   → Solution: Standards for modeling LSRs

Vision: One-stop resource

Contributions:
1. UBY-LMF: Lexicon Model
   • ISO Standard Lexical Markup Framework (LMF)
   • Expert-built and collaboratively created LSRs
   • Fine-grained modeling of information types
2. UBY
   • 9 LSRs in English and German
   • 9 pairwise sense alignments between LSRs
3. UBY-API
   • Uniform access to LSRs in UBY via JAVA API

Applications:
• Word Sense Disambiguation, Information Extraction, Machine Translation, Question Answering...

UBY-LMF

Enable semantic interoperability between resources in UBY:
- Senses linked by SenseAxis class
- Existing alignments
  • Created by experts, users, or automatically
  • Created by Alignement Framework:
    - Cross-lingual alignment (OmegaWiki–DE → WordNet)
    - Planned alignments
      • OmegaWiki–EN → Wiktionary...

UBY-API
• JAVA with Hibernate
• NLP application:
  • Easy swapping of resources
  • Easy combination of resources

Open licenses for data and software:
- UBY-API, LSR to UBY conversion routines: http://code.google.com/p/uby/
- UBY-LMF DTD, UBY release 1.0 (MySQL database dump): www.ukp.tu-darmstadt.de/uby/

See also

UBY-LMF

Subcat-LMF