Computerlinguistik II / Sprachtechnologie

Vorlesung im SS 2019
(M-GSW-10)

Prof. Dr. Udo Hahn

Lehrstuhl für Computerlinguistik
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

http://www.julielab.de

Computational Lexicons, Terminologies & Ontologies

- **Computational Lexicons**
  - Language-specific information (English, Spanish, German, etc.), cover common-sense knowledge
  - Cover, at best, all linguistic description levels for a lexical item but usually don't
  - Undetermined towards formalization, yet electronically available
- **Terminologies**
  - Language-independent (though verbally encoded!), cover domain-specific, expert-level knowledge
  - Cover lexico-semantic information only (semantic relations)
  - Informal, computational issues are (usually) of no concern
- **Ontologies**
  - Language-independent, cover domain-specific, expert-level knowledge
  - cover conceptual information (semantic relations, semantic integrity constraints, rules, etc.)
  - Formal specifications, computational issues are a major concern
  - Formal reasoning: inferences

Examples: Computational Lexicons, Terminologies & Ontologies

- **Computational Lexicons**
  - **WordNet (English) & EuroWordNet**
  - **GermaNet**
  - **FrameNet**

- **(Biomedical) Terminologies**
  - **Unified Medical Language System (UMLS)**
  - **BioPortal**
  - **Open Biological Ontologies (OBO)**
    - **Gene Ontology (GO)**

- **Ontologies**
  - **Formal reasoning (for text understanding)**

WordNet

- **English WordNet (V3.0)**
  - semantic (relation) lexicon of English (general language)
    - no morphology!, no syntax!, no etymology
  - groupings of words into sets of synonyms (synsets)
  - English definitions for lexical entries/synsets (glosses)
  - defines semantic relations between synsets
  - covers (base forms of) nouns, verbs, adjectives, adverbs
  - Size: more than 155,000 lexical entries

  http://wordnet.princeton.edu/

WordNet

- **EuroWordNet**
  - **Portuguese, Spanish, Spanish Catalan-Basque**
  - **French**
  - **Italian**
  - **German (licence required), Dutch**
  - **Russian, Czech, Hungarian, Slovene, …**
  - **...**
- **Global WordNet**   **http://globalwordnet.org**
  - **Arabic**
  - **Mandarin-Chinese**
  - **Hindi**
  - **….**