Word Sense Disambiguation

Natural Language Understanding Lab

Evgeny A. Stepanov,
Mahed Mousavi, Gabriel Roccabruna

SISL, DISI, UniTN & VUI, Inc.
evgeny.stepanov@unitn.it

Objectives

- Understanding:
  - Lexical Relations
  - Word senses in WordNet
  - Semantic Similarity (in WordNet)
- Learning how to:
  - Dictionary-based Word Sense Disambiguation with WordNet
    - Lesk Algorithm
    - Graph-based Methods
  - Supervised Word Sense Disambiguation
    - Feature Extractions for Word Sense Classification
    - Training and Evaluation
  - evaluate ngram model

Signals & Interactive Systems

Outline

1. Introduction to WSD
   - Task Variants
   - Evaluation
   - Lexical Relations
2. Introduction to WordNet (in NLTK)
   - Synsets
   - Lemmatization
   - Lexical Relations between Synsets

Outline: Approaches to WSD

3. Lesk Algorithm
   - *Exercises*: 20 min
4. Graph-Based Similarity
   - *Exercises*: 15 min
5. Evaluation on Corpus
   - *Exercises*: 15 min
6. Supervised WSD
   - Feature Representation
7. ***Lab Exercise***: Supervised WSD: 30 min

Recommended Reading

- Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft)
  - Chapter 18: Word Senses and WordNet
- Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python
  - Chapter 2: Accessing Text Corpora and Lexical Resources. Section 5: WordNet