

# Information Extraction - Summary

---

Günter Neumann, LT lab, DFKI

# Topics that we covered - session 1 & 2

---

- Introduction of term
  - Information Extraction
  - Relation of IE to other NL-related tasks
  - Standard Definition of Information (SDI)
- IE - Architecture and Tasks Definition
  - IE core functionality; template, template instance
  - Major IE tasks
  - Evaluation Metrics (Recall, Precision, F-measure)
  - Knowledge Engineering versus Machine Learning IE (advantages and disadvantages)

# Topics that we covered - session 3 - 6

---

- Named Entity Extraction:
  - what is it, task, problems
  - Machine Learning of NE, IOB scheme
  - classification of different learning strategies
- Bootstrapping NE-lists
  - idea, generic algorithm
  - co-training algorithm of Singer Collins
  - Algorithm Nomen by R. Yangarger
  - Differences between SingerCollins and Yangarber

# Topics that we covered - session 7 & 8

---

- NE set expansion: what is it/general idea, difference to classical NER
- Vector Space Model approach by Saramento et al.
  - how are NE candidates determined and how is VSM used to select good candidates
- Semantic labeling approach of Van Durme and Pasca: core idea, main algorithm and TFIDF approach

# Topics that we covered - session 9-12

---

- Machine reading and Open-Domain IE
  - basic terms, core idea of OpenIE
- OpenIE system TextRunner: binary relation (triple) extraction, filtering, self-training, relation synonymy, distributional hypothesis and application for OpenIE
- OpenIE system WOE: Self-Supervision with Wikipedia; shallow versus deep patterns
- OpenIE system ReVerb: core algorithm, incoherent/uninformative extractions, relation phrase, relation extraction phrase
- Comparison of the three approaches

# Topics that we covered - session 12

---

- Hybrid Information Extraction: definition, possible strategies, multi-strategy, simple voting strategies for Hybrid NER,