Ontologies & Natural Language Processing

Paul Buitelaar

DERI – Unit for Natural Language Processing

National University of Ireland, Galway
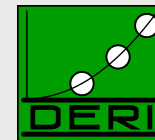
Some History – Ontologies & NLP

- Circumscription (McCarthy 1980)
- TACITUS Commonsense Physics (Hobbs et al 1987)
  - Natural Language Understanding
- Subworld Concept Lexicon (Nirenburg & Raskin 1987)
  - Sub-language, Knowledge-based Machine Translation
- Temporal Ontology (Moens & Steedman 1988)
  - Event Structure Analysis in Natural Language
- Naive Semantics (Dahlgren 1988)
  - Natural Language Understanding (PP attachment)
- PENMAN Upper Ontology (Bateman et al 1990)
  - Natural Language Generation
- MikroKosmos (Mahesh & Nirenburg 1995)
  - Knowledge-based Machine Translation

- Conclusions at end of 90s
  - Knowledge Representation effort/maintenance is to costly & less robust in applications
  - Cheaper, more robust but shallow (semantic) approaches needed
  - Turn towards empirical methods in NLP; KR loses central place in NLP

Recent Parallel Developments in KR & NLP

- **Knowledge Representation**
  - ☐ KR moves to the (Semantic) Web: RDF, DAML/OIL > OWL - standardization
  - ☐ Distributed, collaborative Ontology Development - less costly, more robust
  - ☐ Ontology sharing, merging, etc. – Ontology libraries/repositories

- **NLP**
  - ☐ Robust, statistical methods developed for syntactic analysis: part of speech tagging, chunking, dependency parsing
  - ☐ Renewed interest in semantic analysis: semantic role labelling, temporal analysis, entailment, taxonomy extraction
  - ☐ Applied work in ontology-based information extraction for specific domains, e.g. biomedical, business intelligence

- **KR & NLP moving slowly back together**
  - ☐ KR provides ontologies for use in NLP – information extraction
  - ☐ NLP provides input for ontology development – text mining

Current Trends (relevant for IAOA)

- **Ontology Learning**
  - ☐ Extracting domain ontology models from domain-specific text data

- **Ontology Population**
  - ☐ Semantic annotation, Ontology-based information extraction
  - ☐ Extracting instances from text for knowledge base generation

- **Lexicalized Ontologies, Lexical/linguistic ontology enrichment**
  - ☐ Ontologies often lack information on linguistic realization
  - ☐ Integration of linguistic information with domain semantics needed
  - ☐ LexInfo model: integrating lexical information with ontologies
    - – Paul Buitelaar, Philipp Cimiano, Peter Haase, Michael Sintek *Towards Linguistically Grounded Ontologies* In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009). Lecture Notes in Computer Science. Springer.