Polynomial-time Reasoning Support for Design and Maintenance of Large-scale Biomedical Ontologies

Boontawee Suntisrivaraporn

TU Dresden

Supervisors: Franz Baader, Carsten Lutz

DFG

Aims and Objectives of the Research

- To understand tasks for ontology design and maintenance
  - the domain of medicine and biology
- To identify relevant ontology-based logical reasoning problems
- To design a *lightweight* but sufficient ontology language dialect
- To devise decent algorithms to solve those problems and implement as software tools

Justification for the Research Topic

- Ontologies are ubiquitous
  - Semantics-aware applications are increasingly important
- Life-science ontologies tend to be inclusive and thus large
  - Difficult to maintain and do reasoning on
  - Common tasks, e.g. classification, are highly complex
  - *Scalability* is an important issue

Research Questions

- What are the requirements/tasks for ontology design and maintenance?
  - In the domain of life sciences
- What (logical) reasoning supports are relevant?
- What kind of ontology language is best suited for use with (large-scale) biomedical ontologies?
  - Expressivity vs scalability
- Does automated reasoning support really help in the process?
  - How do usage scenarios look like?

Research Methodology

- Logic-driven methodology
  - Investigate into the computational complexity of several Description Logic dialects
  - Worst-case complexity analysis
  - Empirical experiments
- Biomedical ontology-driven methodology
  - Analyse biomedical ontologies
  - Extract common properties and essential requirements

Research Results to Date

- A DL dialect has been identified, EL++,
  - Expressive enough, e.g. SNOMED, GO, NCI, etc?
  - Polynomial-time classificaton
- The CEL reasoner
  - Robustly scalable, viz. classification of SNOMED takes less than 30 mins
- Reengineering of SNOMED CT by replacing SEP model with proper logical constructors
  - done with restructuring of the anatomical part