Abstract by author:
Valuable information in the biomedical domain is represented in textual form only (e.g. scientific publications, patents). Natural language processing for this domain (BioNLP) aims at making this knowledge accessible by means of Text Retrieval (TR) and Information Extraction (IE). Names of biochemical compounds occur frequently in these texts. These names need to be mapped to unique references, i.e. the corresponding molecular structures, for TR and IE tasks as well as for database curation. Additionally, the names need to be assigned chemical classes. Classification enables mapping to existing taxonomies and ontologies and building new taxonomies. In this talk I will first describe the particularities of the biochemical terminology and point out the limitations of current approaches for processing biochemical compound names. Then I will introduce a new system for syntactic and semantic processing of these names. Our system aims at automatically modelling a constraint satisfaction problem (CSP) for each given name-to-structure task. This way, identification and classification can be achieved. I will conclude this talk with indicating directions for future research.