

Ontology Engineering via Thesaurus Re-Engineering

Udo Hahn

Computational Linguistics Lab, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

From an ontological point of view, knowledge types which are particularly relevant for biomedical reasoning include taxonomic, partonomic and topological knowledge. I report on a large-scale experiment how these rich forms of knowledge can be extracted automatically from a semantically weak, though high-coverage biomedical knowledge repository, the UMLS thesaurus. Upon extraction, these knowledge pieces are automatically transformed into a formally parsimonious description logics language (ALC). The terminological reasoning engine of the underlying knowledge representation system then allows for consistency and cycle checking of the emerging knowledge base which currently comprises 240,000 conceptual entities. Some empirical data will be discussed considering the work load required to curate such a knowledge base in terms of biomedical adequacy and quality.

Biographical notes:

Udo Hahn is a professor of natural language processing at Freiburg University, Germany (since 1990).

He was previously affiliated with departments of computer science (Passau University) and information science (Constance University).

He got his Ph.D. in information science from Constance University (1987) and a Master's degree in linguistics from Mainz University (1980).

His research interests cover the following areas: natural language processing, ontology engineering (applications in information technology and biomedicine), text understanding applications (text summarization, knowledge extraction and text mining), information retrieval