Automation of the Unification Problem in Description Logics

The great advancements of the knowledge about our world, society, health, biology and the applications thereof in the various services facilitating human endeavors, require automation of processes that has up to now been left for humans to do and that has always been time consuming, error-prone and strenuous. The gathered knowledge is otherwise unmanageable and not used to the whole of its potential. Therefore, for the purpose of providing a formal and rigorous representation of knowledge Description Logic was introduced in 1990s. Description Logic's approach is to capture knowledge in the formally defined concepts. Starting from the undefined, simple names, it constructs the complicated concepts using formal constructors. Since the constructors have a defined semantic, the paradoxical consequences of uncareful formulations can be detected and/or removed automatically.

Since the introduction of Description Logic, the big collections of concepts, definitions, statements about the various areas of knowledge has been constructed. They are called ontologies. Some of the ontologies deal with health and medical services. They contain anatomical knowledge about human body, definitions of illnesses and medical procedures. They are used in facilitating the administrative tasks, lowering the burden of documentation for doctors in hospitals, providing them with useful reference, answering queries in a smart way, all of it using the inferences performed in the underlying logic.

The project "Automation of the Unification Problem in Description Logics" proposes a research in this direction: to facilitate the management of medical ontologies like SNOMED CT (361 907 concepts), GALEN (23 141 concepts), FMA (104 721 concepts). We want to design algorithms that would allow detection and removal of redundancies from such knowledge bases. Redundancies, i.e. concepts that denote the same thing, but differ between themselves by using different primitive notions, due to different approaches of ontology designers, may cause errors or at least unnecessarily slow down inferences made over a given ontology. Unification algorithm should provide an excellent tool to decide if some concepts are redundant. Given two complex concepts, the algorithm attempts to construct a set of definitions for the undefined names that appear in these concepts in such a way, that the two become identical or equivalent w.r.t. to these new definitions. The set of definitions computed by the algorithm may or may not make sense for an expert in the area. It should be assessed by such an expert and then the redundancy may be repaired by removing one of the culprit concepts and extending the ontology with the set of new definitions.

Such algorithms has already been developed by other researchers as well as by the PI of the project, but they can work on only a small fragment of Description Logics, which is not expressive enough to deal with a full ontology. We will attack the unification problem for the extended fragment of Description Logic, which contains enough constructors, to express the knowledge represented in these big medical ontologies. In order to do this, we will use the intuitions from our previous work on unification on smaller fragments of Description Logic. Such research is being conducted in research centers all over the world, we hope that the results obtained during our work on the project will provide new tools for solving theoretical problems as well as helping the practical management of the medical knowledge bases.