

Challenges to automated reasoning in decidable fragments of first-order logic: trees, orders and numerals

First-order logic is a formal language widely used to describe structured ensembles of objects and data. Its combination of mathematical precision and intuitive readability have made it a standard medium of representation in a range of academic disciplines. In Computer Science, the use of formal languages, including first-order logic, to describe, query or manipulate structured data is firmly embedded in day-to-day practice.

The idea of a mechanical procedure for reasoning with formal languages arguably dates back to Classical times; however, the question of the existence of an algorithm for performing inference in first-order logic—known as the *Entscheidungsproblem*—was first formulated at the start of the twentieth century. It was soon shown that no such algorithm exists: first-order logic is undecidable. There are two responses to this situation, both of which have achieved notable practical success in recent years. The first is to develop algorithms to reason with arbitrary collections of first-order formulas, accepting that, however well they generally work in practice, there will always be instances that defeat them. The second is to restrict attention to a subset—or, as we say, *fragment*—of first-order logic, for which the problem of determining logical validity is algorithmically decidable, exploiting the fact that, in many real-life situations, the formulas we encounter will belong to such decidable fragments.

Where the problem of reasoning in a logical fragment is decidable, it makes sense to investigate its computational complexity; and the computational complexity of reasoning in fragments of first-order logic has become an important area of research in theoretical Computer Science. The broad pattern that emerges is one of a trade-off between expressiveness and computational manageability: as the expressive power of a logical fragment increases, so does its computational complexity. More colloquially: the more you can say in a logic, the harder it is to reason with. Research into decidable fragments of first-order logic attempts to establish the exact terms of this trade.

The practical importance of these questions resides in their application to information technology. Many IT systems require access to databases that is mediated by general knowledge concerning the domain in question. Querying databases here is not simply of matching a query to data stored in tables, but rather, inferring the answers to a query from the data together with a background theory, expressed in some logic. To make reasoning computable, the fragment of logic in question must be decidable. We mention three particularly salient such fragments here: the *two-variable fragment*, the *guarded two-variable fragment* (the basis of most *description logics*), and the *fluted fragment* (a currently relatively neglected logic, originally identified in the 1960s). The trade-off alluded to above then becomes one between answering queries efficiently and being able to express the relevant general knowledge in the background theory. Two commonly occurring expressive requirements in particular are known to pose problems in respect of computability and efficiency: that of dealing with *ordered* data (i.e. where one data-value is, in some respect, *greater than* another), and that of accessing data organized into tree-like structures (as, for example in XML documents).

The proposed workplan seeks to resolve various open reasoning problems concerning the fragments mentioned above and centred around the concepts of ordering and trees, including the ability to express numerical restrictions. These questions will be tackled by application of a range of mathematical techniques that have been used, in various combinations, to other problems of decidability and complexity of first-order fragments. Answering them will provide designers of knowledge-mediated information-retrieval systems with additional possibilities in regard to the logics that may be considered.