論文の内容の要旨

論文題目　Complexity and Expressiveness of Models of XML Translations

(XML 変換モデルの計算複雑性と表現力)

氏名　稲　葉　　一　浩

XML has become widely used in computer industry, and the importance of static analysis and verification of applications manipulating XML documents is increasing. For analyzing or proving any properties on XML manipulating programs, it is essential to have some model with theoretically well-defined semantics. After a long history of researches on models of tree-to-tree translations, a recent trend is to regard macro tree transducers (mtts) as a standard model for XML manipulation. Mtts are known to cover tree translations expressible by other models such as attributed tree transducers, MSO-definable tree translations, or pebble tree transducers, and the high expressiveness allows representing a vast range of practical XML translations. Yet, they ensure various good properties such as exact typechecking, streaming, decidable emptiness, and so on. Nevertheless, mtts still lack some properties desired for modeling XML translations. In particular, (1) mtts have poor closure properties on composition and (2) computational complexities for many problems on mtts are still unknown. A consequence of the first point is that, for example, an mtt composed with even a very simple pre- or post- tree translation cannot be represented by a single mtt. The lack of the composability implies difficulty of modular modeling; even if we could construct a model for each smaller subpart of a program separately, it is in general impossible to compose them up to obtain the model for the whole program. For the second point, the decidability is proved for many problems on mtts or their compositions, while the complexities and concrete algorithms solving the problems have been left open. This makes it difficult to estimate the computational hardness of each verification problem. One example of such a case is the membership problem of output languages, i.e., the problem determining whether a tree is a valid output with respect to the translation and a given input regular type. Although the problem is essential for, e.g., verifying that a program never generates `wrong' outputs, its complexity had not been analyzed.

The goal of the thesis is to improve these shortcomings of mtts. First, to address the composability issue, a new model of tree-to-tree translation called multi-return macro tree transducer (mr-mtt) is introduced. As its name shows, an mr-mtt is an mtt extended with the capability of returning multiple tree fragments simultaneously, in contrast to an mtt that can return only one. We show that mr-mtts are closed under pre- and post- compositions with arbitrary deterministic total top-down tree transducers, and therefore enable modular modeling of tree-translations. We also show that mr-mtts are strictly more expressive than mtts, which at the same time formally proves as a corollary the folklore conjecture that mtts are not closed under post-compositions with top-down tree transducers.

Second, the thesis investigates complexity on compositions of mtts. We show that the data-complexity of the membership problem of output languages is in the class DSPACE(n) and is NP-complete. The crucial lemma of the proof is that any composition of finite number of mtts can be transformed into so-called garbage-free forms, meaning that any subtrees of intermediate results are actually used for generating the final output. Besides the complexity of output languages, we give another application of the garbage-free form: the complexity is shown for the 'translation membership' problem which determines whether a given pair of trees is an input-output pair of the translation.