Universität Bremen

Universität Bremen · Postfach 33 04 40 · 28334 Bremen

Prof. dr. Andrzej Tarlecki
President of the Academic Council for Mathematics and
Computer and Information Science
University of Warsaw

fachbereich 3
mathematik und informatik

**AG Datenbanken**

**Prof. Dr. Sebastian Maneth**

**Postanschrift:**
Universität Bremen
Fachbereich 3
Postfach 33 04 40
28334 Bremen

**Besucheradresse:**
Gebäude MZH
Raum 3350
Bibliothekstraße 1
28359 Bremen

(0421) 218 - 63600
maneth@uni-bremen.de
http://wwwdb.informatik.uni-bremen.de

Datum: 18.3.2022

**Review of the PhD thesis of Janusz Schmude**

Dear Madams and Sirs,

The following text contains my review of the PhD thesis of Janusz Schmude that was submitted in November 2021 at the University of Warsaw's Faculty of Mathematics, Informatics and Mechanics.

The thesis is concerned with proving that equivalence is decidable for certain classes of transducers. A transducer is typically a finite-state device that operates on finite inputs and outputs. The inputs and outputs are typically strings, trees, or graphs.

The question whether or not equivalence is decidable for a class of transducers (or algorithms) has been fascinating many researchers for many decades. We may ask, what is the most powerful formalism (say, class of transducers or class of algorithms) for which equivalence is decidable? It should be clear that being able to decide equivalence gives rise to manifold applications. To set the stage: even for the computationally rather weak formalism of context-free grammars, it is well-known that equivalence is undecidable. In contrast to that, for various types of transducers (which generate classes of output languages that are far beyond the class of context-free languages) it has been possible to prove decidability of equivalence.

One of the classic results in terms of transducers is the decidability of equivalence of (functional) top-down tree transducers which was proven in 1980 by Ézic. A related problem that had famously remained open since the 1980's is the equivalence problem for top-down tree to *string* transducers. Intuitively, there are vastly many alternative ways of producing the same output strings, so that no canonical normal form, or no generalization of the method of Ézic have ever been found. In 2018 this problem was finally shown to be decidable by Seidl, Maneth, and Kemper. Their idea was to simulate top-down tree transducers by so called "polynomial transducers" with output in the ring of integers. Using polynomial ideals and Hilbert's Basis Theorem they present two semi-algorithms, one searching for an invariant (the existence of which establishes equivalence) and one searching for a counter-example (i.e., an input which establishes in-equivalence). This approach was termed "the Hilbert Method" by Bojańczyk in 2019. Note that there are prior applications of this method, e.g., by Albert and Lawrence in 1985 (to give a proof of Ehrenfeucht's Conjecture), or in 2020 by Honkala (to give a (very) short proof of the HDT0L sequence equivalence

**Bankverbindungen:**
Bremer Landesbank
BLZ 290 500 00
Konto 1070 500 007
Swift-Code BRLADE22

Umsatzsteuernummer
71-545/22101

Umsatzsteuer-ID
DE811 245 070

problem).

The thesis applies the Hilbert Method to new transducer contexts and thereby establishes new results on the decidability of equivalence.

Chapter 1 introduces some background concerning commutative algebra and ideal membership. In Chapter 2 the particular transducer model is defined. A transducer is defined for a given (output) algebra (i.e, a universe together with a set of operations). The input of the transducer is determined by a ranked alphabet, i.e., the transducer takes ranked trees as input. The transducer is also equipped with a finite number of registers (which store intermediate results). A transition is of the form $(q_1, \ldots, q_k, \sigma, p, q)$ which means that if the machine has arrived (bottom-up) at an input node labeled $\sigma$ and in state $q_i$ at the $i$th child of the node, then the machine may change to state $q$ and apply the polynomial operation $p$ to the register entries of the children nodes (in order to obtain the new values of the registers at the current node). A final "output function" applies a polynomial operation to the registers at the input root node in order to obtain the final output (which may be a tuple of values). Since the polynomial operation $p$ determines the new values of all registers at once, the functionality problem of nondeterministic transducer and the the equivalence problem of functional transducers are interchangeable. Therefore, we will not mention anymore the functionality problem. The first basic result of Chapter 2 states that equivalence is decidable for functional transducers with output in a computable ring with no zero divisors. It is obtained by application of the Hilbert method.

Section 2.3 focuses on *unordered* output trees. The idea is to simulate transducers with unordered output trees by transducers with output in the ring $\mathbb{Z}[x]$ (i.e., the update operations are polynomials with coefficients in $\mathbb{Z}[x]$). Roughly speaking, in the simulation each node on a path of the output tree raises the power of $x$ in such a way, that via Eisenstein's Criterion the proposed "simulation function" can be shown to be injective. As a consequence, equivalence is decidable for transducers with output in an unordered forest algebra. From my point of view this is a novel and nice result. — It could be nice to explicitly state whether this result has consequences for conventional types of tree transducers such as deterministic top-down or bottom-up tree transducers (either for ranked, or for unranked unordered output trees).

In Chapter 3 it is proven that equivalence is decidable for (walk-functional) MSO (graph) transductions for input graphs of bounded treewidth and output graphs that are equivalent up to "walk-equivalence". Note that from previous results it had been known that if the output graphs are restricted to string or trees, then the corresponding equivalence problems are decidable. Two graphs are walk-equivalent, if for every $n$, they contain the same number of length $n$ walks. As before, it is shown that the walk-functionality problem is interchangeable with the equivalence problem of walk-functional MSO transducers (on bounded treewidth input graphs). One essential idea of the proof (of the main result) is to use rational power series (i.e., certain restricted infinite polynomials) to represent the numbers of walks of a graph; in such a power series, the coefficient of $x^n$ is the number of length-$n$ walks of the graph. The proof proceeds by presenting a mapping from the $k$-sourced graph algebra to such power series that is "injective up to walk-equivalence". In a nutshell, it is then shown that the algebra of power series can be polynomially simulated by the enrichment of the field $\mathbb{Z}(x)$ by the division operator, and finally, that zeroness of register transducers with the output in this enrichment can be reduced to zeroness of register transducers with output in the ring $\mathbb{Z}[x]$. I was quite intrigued by the ideas of this proof. It could be nice to investigate whether it is decidable if the range of an MSO transducer does not contain two non-isomorphic graphs that are walk-equivalent, so that in this case equivalence up to isomorphism could be obtained.

Chapter 4 studies the effect of adding substitution to the output ring of polynomials of a transducer. In terms of polynomials $p, q$, substitution is the same as the composition $p \circ q$. The first main result states that for a transducer and two of its registers $R$ and $S$ it is undecidable to determine if $R \circ S = 0$ for every accepting run. The proof is done via a reduction to the reachability problem of reset vector addition systems with states. Then two restrictions on the use of substitution are identified under which zeroness is decidable.

Chapter 5 deals with finitely presented monoids. In particular, monoids $\mathbb{M}$ are considered whose relations permute the letters of a word. The chapter establishes a necessary criterion under which $\mathbb{M}$-equivalence of transducers with output in the monoid $\Sigma^*$ is decidable. The feasibility of the method is demonstrated on an example. The same strategy is then applied to the “monoid of queue actions” (which is well known from the literature). It is shown that unfortunately the criterion does not hold and hence the method is not applicable. The chapter closes with ideas on automatizing this strategy.

I deem this thesis fully sufficient to grant a PhD. The author has demonstrated the ability to carry out in-depth research. The particular topic is mathematically demanding. The author has not only been able to master this topic, but also to represent the outcomes in a genuinely uniform representation. The representation is dense and compact, yet, it contains enough information to conveniently follow the arguments. The notation is chosen very carefully not to overwhelm but to allow to concentrate on the main aspects of the constructions and proofs. This is carried out in a very consistent and pleasing manner throughout the entire thesis.

It was a pleasure for me to read this thesis and I congratulate the author for his achievement.

With best regards,

Sebastian Maneth