UNIVERSITÄT
LEIPZIG

Fakultät für
Mathematik und Informatik
Institut für Informatik
Dr. habil. Karin Quaas

Universität Leipzig, Fakultät für Mathematik und Informatik, 04081 Leipzig

Prof. Andrzej Tarlecki
Chairman of the Academic Council
of Mathematics and Computer
and Information Sciences

Telefon +49 341 97-32172
Telefax +49 341 97-32241
quaas@informatik.uni-leipzig.de

Review of the Doctoral Thesis by Mr Mohnish Pattathurajan

2. Mai 2023

Dear Prof. Andrzej Tarlecki, dear members of the Academic Council,

Mr Mohnish Pattathurajan has submitted a doctoral thesis with the title "Commutative Images of Languages Over Infinite Alphabets" to the Academic Council of Mathematics and Computer and Information Sciences of the University of Warsaw. The thesis comprises around 100 pages and 9 chapters, and it is written in English.

The thesis studies the model of *register automata*, that is, finite automata operating on finite words over some infinite alphabet (like the natural numbers or the rational numbers). In order to accept interesting languages over such infinite alphabets, the automaton is equipped with a finite set of registers that can store current input letters (also called *data*) for later comparison. As such, register automata accept languages called *data languages*, and the class of such recognizable data languages has interesting properties, which may deviate from the class of *regular* languages (that is languages recognizable by finite automata defined over some finite alphabet). For instance, it is well known that the class of recognizable data languages is not equal to the class of data languages recognizable by *deterministic* register automata.

The main research question of the thesis is to study whether the famous Parikh theorem for contextfree languages also holds for data languages accepted by register automata. Parikh's theorem states that the Parikh image of every language (that is the set of Parikh vectors of all words in that language) that can be defined by a contextfree grammar is equal to the Parikh image of a regular language. Moreover such a set of vectors is semilinear, that is, definable by a finite union of linear sets.

The question whether this can be generalized to the class of data languages (that are recognizable by register automata) is a natural and challenging research question.

The thesis starts with a chapter giving an informal introduction to the main concepts and an overview over the main results of the thesis. The next two chapters formally define the concepts of *sets of atoms*, and *register automata* and *register grammars* (which are defined over sets of atoms). The formal definitions are accompanied by lots of illustrating examples.

In chapter 4, the authors proposes a generalization of the well known concept of *rational expressions* for describing recognizable languages (like, e.g., $(a + bb)^*$) to orbit-finite (but infinite) alphabets. The definition is very natural; however, the data languages that can be defined by such rational expressions cannot capture the set of recognizable data languages (a counterexample is presented in Prop 4.1.3). He also proposes a notion of *rational sets of data vectors* that can be used to characterize the Parikh image of data languages. This is useful, because already in the next chapter, the notion of semilinear sets is generalized to orbit-finite alphabets, and it is proved that - in opposition to regular/contextfree languages in the classical setting - the Parikh images of data languages accepted by register automata are not semilinear, and this holds even for deterministic one-register automata. Indeed, the author proves formally in chapter 6 (certainly the main technical chapter) that Parikh images of data languages accepted by one-register automata are rational. This result is then generalized to languages defined by one-register context-free grammars. Last but not least it is proved that for every one-register context-free language there exists a register automaton with the same Parikh image.

As a conclusion, this thesis makes some first steps into partially proving that Parikh's theorem holds for data languages accepted by register automata. It is conjectured that one can generalize to the case with more than one register. As a first step into this direction, the results from the previous chapters are generalized to so-called hierarchical register automata (chapter 7). I do not know yet this particular class of register automata, but it looks natural to me. Also, in chapter 8, more generalisations with respective to context-free grammars are presented.

The thesis ends with some conclusion and remarks on possible future research directions.

The thesis makes a very good impression on me. It contains everything a good Phd thesis in the field of theoretical computer science needs: a good introduction containing motivation and background of the work, as well as related literature; interesting results with solid technical proofs, and some outlook on future research.

The citation is correct. (I especially like the bibliographical footnotes, which are actually unusual in our field of research, but they make the thesis easier to read!)

The topic of the thesis is up-to-date and certainly of interest in the research community. This is also proved by the acceptance of one of the contributions to LICS 2021 (one of the best conferences in our field), and another part is accepted at FSTTCS 2021. Register automata are surely in the focus of active research in the fields of automata theory and verification; hence the thesis is a contribution to an active resarch field.

In my point of view, the thesis is remarkably well written. The text is very readable – it was actually a pleasure to read the thesis –, and the layout is supporting the reader very much. Already Chapter 1 guides the reader in a very convenient way to the topic of the thesis. The author gives an easy-to-read introduction to the automata models and concepts studied in the thesis, as well as mentioning and explaining the main results of the thesis along with some technical difficulties and challenges that arose during the research work.

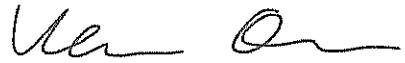
The proofs are very well written as well, and accompanied with examples. The two chapters that form the technical core of the results (chapters 6, 7 and 8) contain very clear concepts (e.g. altering loops, anti-paths, anti-cycles in chapter 6) that are extremely helpful in understanding the technically involved proofs. I did not find any mistakes in the reasoning.

I spotted only few typos.

Altogether I do not only deem the thesis as sufficient to grant a PhD, but I find the thesis exceptionally good. I would like to congratulate the author to this very

good thesis.

With best regards



Karin Quaas