The single-use restriction for register automata and transducers over infinite alphabets

Thesis summary

Author: Rafał Stefański

PhD Supervisor: prof. dr hab. Mikołaj Bojańczyk

June 20223

The thesis studies the effects of restricting register automata and transducers with the *single-use restriction*. It also structures and clarifies the results published in [5].

Register automata have been defined and studied in [7], which introduces them as a model that is powerful enough to handle infinite alphabets but is still similar enough to regular languages, so that it can be reasoned about using the techniques from automata theory (e.g. their emptiness problem is decidable). The same paper [7] notices that the expressive power of register automata is not as stable as the one of register automata – deterministic one-way, deterministic two-way and nondeterministic one-way register automata all recognize different classes of languages (in case of finite automata, all of those variants recognize the class of regular languages). Further studies (such as [11]) have confirmed this lack of robustness – virtually all variants of register automata are pairwise nonequivalent, whereas their finite-state counterparts are all equivalent to regular languages.

This lack of robustness does not change the fact that the register automata (and their variants) are important both theoretically and practically. Because of that they have been thoroughly studied. One approach, started in [2] and in [9], is to study register automata using *sets with atoms* (also called *nominal sets*). The paper [2] defines the class of languages recognized by *orbit-finite monoids* (weaker than register automata) and studies its connections to the $\mathrm{MSO}^{\sim}$ logic and to the $\mathrm{FO}^{\sim}$ logic. The class of languages was further studied in [6], which defines *rigidly-guarded* $\mathrm{MSO}^{\sim}$ and proves that it is equivalent to orbit-finite monoids. Papers [13] (my master's thesis) and [5] continue to study the class of languages recognized by orbit-finite monoids. They introduce the *single-use restriction* for register automata (stating that every read access to a register should have the side result of erasing that register's contents), and show that *single-use register automata* are equivalent to orbit-finite monoids. This means that there is a class of languages over infinite alphabets, which has three substantially different definitions:

- the algebraic definition – through orbit-finite monoids;
- the logical definition – through rigidly guarded $\mathrm{MSO}^{\sim}$;
- the automaton definition – through single-use register automata.

(There is also an ongoing work developing the topological definition [14].) This strongly suggests that this class is theoretically important and that it is worth studying.

The structure of the thesis

The thesis consists of five chapters. The first two of them discuss automata, the next two discuss transducers and the last one presents possible directions for further studies.

1. Infinite alphabets

The first chapter presents the fragments of the existing theory of register automata that are relevant to the thesis. In particular, it defines register automata (and some of their variants) and introduces the theory of sets with atoms. It is mainly based on [3] and [12].

2. Single-use restriction

The second chapter defines the single-use restriction for register automata. Following that, it introduces the concept of *single-use functions* (as a generalization of the single-use condition) and studies their properties. In the end it proves that orbit-finite monoids are at least as expressive as single-use register automata (the other inclusion is proven in Chapter 3). This chapter is based mainly on [5] and on [13], but it also presents some new results.

3. Single-use Mealy machines and their Krohn-Rhodes decomposition

The third chapter starts with a definition of *single-use Mealy machines* (an infinite-alphabet extension of the classical Mealy machines defined in [10]). After that it presents the formulation of the original Krohn-Rhodes theorem, it generalizes it to infinite alphabets, and proves the generalized version. In the process, it defines the model of *a local monoidal transduction* and proves the outstanding inclusion from the previous chapter. It is mainly based on [5], but it also presents some new results.

4. Two-way transductions with atoms

The fourth chapter defines the following models of transducers:

- two-way single-use register automata;
- single-use and copyless SST's (a model from [1] additionally restricted by the single-use restriction);
- regular list functions over infinite alphabets (an extension of a model from [4]);
- two-way Krohn-Rhodes decompositions for infinite alphabets (an extension of a model from [8]).

Following that, it proves that all of those models are pairwise equivalent. This chapter is based on [5].

References

[1] Rajeev Alur and Pavol Černý. "Streaming transducers for algorithmic verification of single-pass list-processing programs". In: *Proceedings of the 38th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. 2011, pp. 599–610.

[2] Mikołaj Bojańczyk. "Nominal monoids". In: *Theory of Computing Systems* 53.2 (2013), pp. 194–222.

[3] Mikołaj Bojańczyk. *Slightly infinite sets*. 2019. URL: `https://www.mimuw.edu.pl/~bojan/paper/atom-book`.

[4] Mikołaj Bojańczyk, Laure Daviaud, and Shankara Narayanan Krishna. "Regular and first-order list functions". In: *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*. 2018, pp. 125–134.

[5] Mikołaj Bojańczyk and Rafał Stefański. *Single use register automata for data words*. 2020. arXiv: `1907.10504 [cs.FL]`.

[6] Thomas Colcombet, Clemens Ley, and Gabriele Puppis. "Logics with rigidly guarded data tests". In: *Log. Methods Comput. Sci.* 11.3 (2015). DOI: `10.2168/LMCS-11(3:10)2015`. URL: `https://doi.org/10.2168/LMCS-11(3:10)2015`.

[7] Michael Kaminski and Nissim Francez. "Finite-memory automata". In: *Theoretical Computer Science* 134.2 (1994), pp. 329–363. ISSN: 0304-3975. DOI: `https://doi.org/10.1016/0304-3975(94)90242-9`. URL: `https://www.sciencedirect.com/science/article/pii/0304397594902429`.

[8] Kenneth Krohn and John Rhodes. "Algebraic theory of machines. I. Prime decomposition theorem for finite semigroups and machines". In: *Transactions of the American Mathematical Society* 116 (1965), pp. 450–464.

[9] Sławomir Lasota, Bartek Klin, and Mikołaj Bojańczyk. "Automata theory in nominal sets". In: *Logical Methods in Computer Science* 10 (2014).

[10] George H Mealy. "A method for synthesizing sequential circuits". In: *The Bell System Technical Journal* 34.5 (1955), pp. 1045–1079.

[11] Frank Neven, Thomas Schwentick, and Victor Vianu. "Finite state machines for strings over infinite alphabets". In: *ACM Transactions on Computational Logic (TOCL)* 5.3 (2004), pp. 403–435.

[12] Andrew M. Pitts. *Nominal Sets: Names and Symmetry in Computer Science*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139084673.

[13] Rafał Stefański. "An automaton model for orbit-finite monoids". MA thesis. University of Warsaw - Faculty of Mathematics, Informatics and Mechanics, 2018, p. 29.

[14] Henning Urbat. *Nominal Topology for Data Languages*. A Short Contribution at 16th IFIP WG 1.3 International Workshop on Coalgebraic Methods in Computer Science. Apr. 2022. URL: https://www.coalg.org/cmcs22/programme/.