UNIVERSITÉ LIBRE DE BRUXELLES

ULB

Dr Emmanuel Filiot
Computer Science Department, NO Building, 8th floor
Campus Plaine
Bd du Triomphe
1050 Brussels

Report on the Dissertation

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

by Rafał Stefański

The thesis of Rafał Stefański deals with automata models processing finite words over infinite alphabets, whose elements are called atoms and can be compared for equality. Over finite alphabets, the class of regular languages is one of the most robust class of languages, as it admits many equivalent definitions of various nature: just to cite a few, computational definitions such as finite automata (deterministic, non-deterministic, one-way, two-way ...), logical definitions based on monadic-second order logic with one successor, or algebraic definitions such as recognizability by finite monoids. Most equivalences are however lost when trying to extend naturally those notions to infinite alphabets. One such and well-known extension is register automata: they are automata equipped with a finite set of registers in which incoming atoms can be stored, and later on compared for equality with other registers or new incoming atoms. Later on, a theory of orbit-finite automata was developed by M. Bojańczyk, which provides a natural way of extending several automata models from finite to infinite alphabets. For example, deterministic finite-automata (DFA) are extended to deterministic orbit-finite automata, defined as DFA with a possibly infinite set of states, which is however orbit-finite (intuitively, such a set is finite up to atom renaming). Accordingly, the transition function is also required to be finite up to (this is called equivariance). Similarly, other classical notions over finite alphabets can be lifted to infinite alphabets by requiring orbit-finiteness, such as non-deterministic orbit-finite automata, two-way orbit-finite automata, recognizability by orbit-finite monoids, etc. However, those notions turn out to be non-equivalent.

The goal of the present work is to define a restriction for which most of the aforementioned notions coincide with orbit-finite monoids, and to study its impact on computational models for functions of finite words (transducers). This is achieved by introducing the so called *single-use restriction*. Intuitively, this restriction for a register automaton requires that whenever a register is tested, its content is erased (it can be tested a *single* time). More generally, the thesis introduces the notion of *single-use functions,* which are functions between polynomial orbit-finite sets with atoms (such sets are isomorphic to disjoint unions of sets of atom tuples), and which are syntactically obtained by combining a set of basic functions (projection, co-projection, symmetry, identity etc.) with three operations: composition, sum and product. As a matter of fact (see Section 2.2.3), extending single-use functions with the basic function copy which duplicates its input, one obtains a class of functions into which the transition functions of (multiple-use) register automata fall. I now describe the contributions of the thesis.

Chapter 1 provides a smooth introduction to the main notions that are used in the thesis: sets with atoms, automata models over such sets, orbit-finite monoids, ... In particular, it contains a proof (Theorem 1) that deterministic register automata corresponds to deterministic orbit-finite automata. Many examples are provided to help the reader, as well as useful intuitions. This is clearly a section I would recommend to read to anyone starting on the topic of automata models over infinite alphabets.

Chapter 2 introduces the single-use restriction and in particular the class of single-use functions. Once again, this is done in a very pedagogical way, with many examples (see for instance Section 2.2.2). The main result of this chapter is Theorem 6, which proves the correspondence between recognizability by orbit-finite monoids (restricted to polynomial orbit-finite alphabets), one-way deterministic single-use automata and two-way deterministic single-use automata (such automata have in particular a single-use state-to-state transition function). Dropping the polynomial orbit-finite restriction of the alphabet is left as future work. One-way deterministic single-use automata slightly generalize deterministic single-use *register* automata (the models are equivalent when the alphabet is exactly the set of all atoms). The main ingredients of the proof are Theorem 5 of the same chapter, which proves that set of all single-use functions is orbit-finite, and a decomposition result for single-use Mealy machines, which is the main purpose of Chapter 3. The latter result is perhaps the most difficult and technical of the thesis.

Chapter 3 is the longest chapter of the thesis, it starts on page 90 and ends on page 173, and contains many deep and interesting results. It deals with automata models (transducers) to compute functions of words over infinite alphabets. After a brief overview of existing transducer models over finite alphabets, it introduces single-use Mealy machines, which extend single-use automata with outputs. Such machine compute length-preserving functions of words over a (polynomial orbit-finite) set with atoms. The main result is Theorem 8, which proves a decomposition theorem in the spirit of Krohn-Rhodes decomposition: functions computed by single-use Mealy machines can be obtained as compositions of prime functions which include, for instance, functions computed by Mealy machines over finite alphabets (which are themselves known to be decomposable into other prime functions over finite alphabets), or iterations of equivariant letter-to-letter functions. This is a strong and difficult result, published by the author and M. Bojańczyk, but the thesis provides an alternative proof. First, single-use Mealy machines are translated into a more algebraic model based on orbit-finite semigroups, called local semigroup transductions. This model is interesting in itself and proved to be equivalent to single-use Mealy machines. To show that any local semigroup transduction is equivalent to some single-use Mealy machine, it is first decomposed into prime functions. Compositions of prime functions are then easily shown to fall into the class of functions computed by single-use Mealy machines.
The decomposition result of local semigroup transductions is the most technical, and is based on a powerful tool, factorization forests. Factorization forests are well-known for finite semigroups: roughly, they decompose any finite sequence of semigroup elements as a tree which explicits some structure between idempotent infixes of the sequence. Idempotent infixes play a crucial role in automata theory as they can be iterated. The celebrated Simon's factorization theorem states that any sequence admits a decomposition tree whose height is bounded by a constant which only depends on the semigroup, and not on the sequence. Simon's forest factorization theorem does not generalize to orbit-finite semigroups (see Example 35), but a factorization theorem is obtained by the author (Theorem 11) by relaxing the idempotency requirement in the definition of factorization forests. The author even proves a stronger result: (bounded) factorization trees can be constructed by composing prime functions. This general result is of independent interest, and its proof is quite involved. Then, it is shown how to exploit bounded factorization trees to decompose local semigroup transductions as primes (Section 3.5).
The chapter then conducts an independent study of local semigroup transductions. Its main contribution is a characterization (Lemma 69) of local semigroup transductions within the class of equivariant functions, based on a notion of syntactic semigroup transduction, which echoes the notion of syntactic monoid in language theory. This characterization allows to decide whether a given (not necessarily local) orbit-finite semigroup transduction is equivalent to a local orbit-finite semigroup transduction (Theorem 12).
The chapter concludes with a study of local rational semigroup transductions, which strictly extend local semigroup transductions. Their finite alphabet counterpart are (length-preserving) rational

functions, a class of functions central in transducer theory. The main result is a far non-trivial Krohn-Rhodes decomposition result for this class, useful for the last chapter of the thesis, but also of independent interest.

The last chapter, Chapter 4, pushes the computational power of the single-use transducer models of the previous chapters. It introduces single-use *two-way* transducers, and other models which are proved to be equivalent. In particular, its main result, Theorem 15, states the equivalence between single-use two-way transducers, single-use streaming transducers, regular list transductions with atoms, and compositions of single-use two-way primes. Similar equivalences are now well-known for finite alphabets, but most of them cannot be extended to infinite alphabets without the single-use restriction, thus showing once again how robust is the single-use restriction.

The thesis of Rafał Stefański is an outstanding body of work which contains many deep, interesting, new and important results and lay the foundations for a theory of regular languages and functions of words over infinite alphabets. It is particularly interesting and to my opinion quite surprising that a simple restriction such as the single-use restriction, which is syntactic by nature, turns out to be a very robust notion, supported by the many model equivalences proved in the thesis, even with more algebraic models (single-use Mealy machines and local semigroup transductions for instance). The thesis contains original results or proofs which were not published before by the author, and requires to master involved notions from mathematics and automata theory. The writing quality is excellent: it is written with a lot of rigour with respect to technical details, as well as with high-level intuitions and pedagogical introductions to the mathematical concepts that are used or defined, showing the expertise and maturity of the author.

For all these reasons, I enthusiastically recommend to accept the work of Rafał Stefański as a doctoral dissertation, and propose to award it an "honorary distinction".

Brussels, 30/08/2023

Emmanuel Filiot