

Formal Methods of Transformations

Edited by

Emmanuel Filiot¹, Sebastian Maneth², and Helmut Seidl³

¹ Free University of Brussels, BE, efiliot@ulb.ac.be

² Universität Bremen, DE, maneth@uni-bremen.de

³ TU München, DE, seidl@in.tum.de

Abstract

The goal of this Dagstuhl seminar was to gather researchers working on the theory and practice of transformations (also known as transductions) of word and tree structures, which are realised by transducers (automata with outputs). This seminar was motivated by recent advances and breakthrough results, both in the settings of words and trees.

Seminar April 2–5, 2017 – <http://www.dagstuhl.de/17142>

1998 ACM Subject Classification F.1.1 Models of Computation, F.4.3 Formal Languages

Keywords and phrases string transducers, tree transducers, expressiveness, complexity

Digital Object Identifier 10.4230/DagRep.7.4.23


Edited in cooperation with Ismaël Jecker

1 Executive Summary

Emmanuel Filiot

Sebastian Maneth

Helmut Seidl

License  Creative Commons BY 3.0 Unported license
© Emmanuel Filiot, Sebastian Maneth, and Helmut Seidl

The Dagstuhl seminar 17142 “Formal Methods of Transformations” was a short two and half day seminar that took place from April 3rd to 5th, 2017. The aim of this seminar was to bring together researchers working on theory and applications of formal models of transformations (also known as transductions) of strings and trees. A model of transformation which has been central in this seminar is that of a transducer, i.e., an automaton extended with output. Transducers were introduced in the 1960s as formal models for linguistics and syntax-directed translation in compilers. Today, research on string and tree transducers is an active field with various new applications in databases, document processing, natural language processing, software engineering, and verification. To make the seminar more focused, we had identified six research directions as key topics for the seminar:

- Canonical Normal Forms: it is well-known that regular languages admit a unique minimal (and canonical) deterministic finite automaton. Similarly, is it possible to characterize classes of transformations by canonical normal form for transducers?
- Transducer-Logic Relationships: there are well-known automata-logic correspondences in the theory of languages. What is known for transformations and can we obtain similar connections?



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Formal Methods of Transformations, *Dagstuhl Reports*, Vol. 7, Issue 04, pp. 23–37

Editors: Emmanuel Filiot, Sebastian Maneth, and Helmut Seidl



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- Subclass Definability: a fundamental question, which requires a deep understanding of the manipulated objects, is that of subclass definability: given (an effective description of) an object in some class C , does this object belong to a given subclass C' of C ? What are the recent results and open problems with respect to subclass definability for transformations?
- New Decidability Results: what are the recent decidability breakthroughs in the theory of transformations and what are the important open problems?
- New Transducer Models: is there a need for new transducer models tailored to new applications?
- Model-Checking Data-Centric Systems: what are the potential applications of transducer theory to the verification of systems that transform data?

The seminar gave a large overview of recent results and open problems with respect to these research directions. It was a follow-up of Dagstuhl seminar 13192 “Tree Transducers and Formal Methods”, which now included researchers on string transducers. String transducers have indeed received a lot of attention in the recent years and new important results have been obtained. The aim of this seminar was also to gather researchers from the string and tree transducer communities. There were 31 participants from 10 countries (Sweden, France, Poland, Germany, US, Belgium, UK, Japan, Portugal, and India). These participants were invited by the organizers Emmanuel Filiot, Sebastian Maneth, and Helmut Seidl to give survey talks and shorter talks on their current research.

The seminar started with a survey talk by Emmanuel Filiot on recent results on string transformations, intended to motivate the need to address the theory of string transformations in this seminar, and to introduce some of its main recent breakthroughs. Mikolaj Bojanczyk then gave a survey talk on the notion of *origins* in transformations. Origins are inherent to all known transducer models and making them explicit in the semantics of transducers gave rise to new decidability and definability results. Mikolaj’s talk was followed by a session on string transducers. Luc Dartois presented a new and expressive logic to define string transformations. The logic offers good decidability properties with respect to satisfiability and equivalence (under the origin semantics). String transformations, under the origin semantics, can be seen as sets of graphs, called origin graphs. Bruno Guillon presented a characterization of the class of origin graphs generated by known string transducer models. Finally, Jacques Sakarovitch closed this session by presenting a contribution, and an open problem, on rational base numeration systems and their analysis by means of automata and transducers.

The afternoon session started with a survey talk by Christof Löding on the automatic synthesis of deterministic transducers of strings and trees, from specifications given by non-deterministic transducers. This problem, called the uniformization problem, can be seen as a variant of the classical Church synthesis problem. Two short talks given by Nathan Lhote and Michaël Cadilhac presented their recent contributions on definability problems for string transducers. Nathan Lhote presented decidability results for checking whether a given rational function is first-order definable, and, more generally, whether it is definable by some string transducer whose (input) transition monoid (which disregards the outputs) belongs to some given class, thus lifting to transformations well-known results from the theory of regular languages. Michaël Cadilhac then introduced the notion of C -continuity for string transformations and showed effectiveness of this notion for particular classes C . This notion gives an alternative and machine-independent way, which also takes outputs into account, of defining meaningful subclasses of transformations.

Day 2 (morning) was a session devoted to applications. It started with a survey talk on symbolic transducers, by Margus Veanes. Symbolic transducers have been introduced to

address practical issues when dealing with very large alphabets (even infinite). Transition labels are replaced by predicates defining sets of allowed labels. Fundamental decidability results and practical applications were presented. Based on symbolic string transducers, Loris d’Antoni then presented a fully automatic method to invert a practical class of functional (and injective) transformations, implemented in the tool GENIC. Adrien Boiret then addressed the problem of checking equivalence for symbolic top-down tree transducers. Keisuke Nakano presented some fundamental results on B-terms, which model function composition. Finally, the morning session ended with a talk by Olivier Carton about the notion of infinite word compression by transducers and its relation to normality.

Andreas Maletti started the afternoon session with a survey talk on tree transducers and their applications in linguistics. He presented transducer models well-suited to some linguistic applications such as automatic natural language translation, and showed a comparison with deep learning approaches. Frank Drewes then discussed the notion of graph transformations, and especially of DAGs, and proposed in his talk a definition of DAG transducers. Motivated by natural-language interface applications, Johanna Björklund presented a study of the expressivity of checking stack transducers, which extend with outputs the well-known model of checking stack automata. Day 2 ended with a talk by Helmut Seidl on a result obtained together with Sebastian Maneth and Gregor Kemper on the decidability of equivalence for deterministic top-down tree-to-string transducers, which had been a long standing open problem.

The last day was dedicated to results on the class of string transformations defined by two-way transducers, mostly definability problems. There were two talks, one by Olivier Gauwin and one by Jean-Marc Talbot, presenting two different techniques, for deciding whether a transformation in this class can be defined by a one-way transducers, i.e., is rational. It was open whether this problem was solvable in elementary time, and the two presented approaches answer this question positively. Another definability problem about the minimization of the number of registers in subclasses of SSTs (a model introduced by Alur and Cerny, with the same expressivity as two-way transducers), was discussed by Pierre-Alain Reynier. Ismaël Jecker presented a class of transducers with strong constraints on their structures, but still as expressive as two-way transducers, and at the same time enjoying many good algorithmic properties, in particular improving by one exponential a famous result by Hopcroft and Ullman. Finally, Krishna S. concluded the seminar by sketching an alternative proof of a result by Alur et. al. on regular expressions for string transformations.

We had long after-lunch breaks (till 4pm) to give the opportunity to the participants to discuss. It was greatly appreciated by the participants, and some of them initiated new collaborations. The discussions inspired new ideas and we hope that joint papers will be published by the participants.

We warmly thank Schloss Dagstuhl for making this fruitful event possible, and for their help in the organization. It is highly appreciated as organizers, and allowed us to focus only on the scientific aspects of the seminar.

2 Table of Contents

Executive Summary

<i>Emmanuel Filiot, Sebastian Maneth, and Helmut Seidl</i>	23
--	----

Overview of Talks

Deterministic Stack Transducers – On the Regularity of Stack Languages <i>Johanna Björklund</i>	28
Transducers with origin information <i>Mikolaj Bojanczyk</i>	28
Continuity and Rational Functions <i>Michaël Cadilhac</i>	29
Compression with transducers <i>Olivier Carton</i>	29
Automatic Program Inversion using Symbolic Transducers <i>Loris d’Antoni</i>	29
A decidable logic for string transductions <i>Luc Dartois</i>	30
Recent Results on Word Transductions <i>Emmanuel Filiot</i>	30
Untwisting two-way transducers in elementary time <i>Olivier Gauwin and Anca Muscholl</i>	31
Which classes of origin graphs are generated by transducers? <i>Bruno Guillon</i>	31
On Reversible Transducers <i>Ismaël Jecker</i>	32
Towards algebraic and logical characterizations of rational functions <i>Nathan Lhote, Emmanuel Filiot, and Olivier Gauwin</i>	32
Tree Automata in Parsing and Machine Translation <i>Andreas Maletti</i>	33
On repetitive right application of B-terms <i>Keisuke Nakano</i>	33
Two twinning properties with applications to register minimisation in streaming string transducers <i>Pierre-Alain Reynier</i>	33
Transducers as odometers in rational base numeration systems <i>Jacques Sakarovitch</i>	34
Equivalence of Deterministic Tree-to-String Transducers is Decidable <i>Helmut Seidl, Gregor Kemper, and Sebastian Maneth</i>	35
Combinator Expressions for Regular Transformations <i>Krishna Shankaranarayanan</i>	35
From Two-way to One-way Finite State Transducers: a Shepherdson’s Approach <i>Jean-Marc Talbot</i>	35

Participants 37

3 Overview of Talks

3.1 Deterministic Stack Transducers – On the Regularity of Stack Languages

Johanna Björklund (University of Umeå, SE)

License © Creative Commons BY 3.0 Unported license
© Johanna Björklund

Joint work of Suna Bensch, Johanna Björklund, Martin Kutrib

Main reference S. Bensch, J. Björklund, M. Kutrib, “Deterministic Stack Transducers”, in Proc. of the Int’l Conf. on Implementation and Application of Automata (CIAA 2016), pp. 27–38, Springer, 2016.

We introduce and investigate stack transducers, which are one-way stack automata with an output tape. A one-way stack automaton is a classical pushdown automaton with the additional ability to move the stack head inside the stack without altering the contents. For stack transducers, we distinguish between a digging and a non-digging mode. In digging mode, the stack transducer can write on the output tape when its stack head is inside the stack, whereas in non-digging mode, the stack transducer is only allowed to emit symbols when its stack head is at the top of the stack. These stack transducers have a motivation from natural-language interface applications, as they capture long-distance dependencies in syntactic, semantic, and discourse structures. We study the computational capacity for deterministic digging and non-digging stack transducers, as well as for their non-erasing and checking versions. We finally show that even for the strongest variant of stack transducers the stack languages are regular.

3.2 Transducers with origin information

Mikolaj Bojanczyk (University of Warsaw, PL)

License © Creative Commons BY 3.0 Unported license
© Mikolaj Bojanczyk

Joint work of Laure Daviaud, Bruno Guillon, Vincent Penelle

Main reference M. Bojanczyk, “Transducers with Origin Information”, in Proc. of the 41st Int’l Colloquium on Automata, Languages, and Programming (ICALP 2014), LNCS, Vol. 8573, pp. 26–37, Springer, 2014.

URL http://dx.doi.org/10.1007/978-3-662-43951-7_3

This talk is about origin semantics for transducers. The idea is to change the semantics of a transducer, by adding information about which positions in the output originate from which positions in the input. Under such semantics, called origin semantics, fewer transducers are equivalent to each other, e.g. the reverse and identity transformations over a one letter alphabet are not equivalent under origin semantics, but they are equivalent under standard semantics. I will discuss how, with origin semantics, many technical and combinatorial problems go away. There is also an implicit and debatable philosophical point, which is that origin semantics better reflects the “true essence” of a transducer, because when thinking of a transducer one also typically has in mind some origin information.

3.3 Continuity and Rational Functions

Michaël Cadilhac (Universität Tübingen, DE)

License © Creative Commons BY 3.0 Unported license
© Michaël Cadilhac

Joint work of Michaël Cadilhac, Olivier Carton, Charles Paperman

Main reference M. Cadilhac, O. Carton, C. Paperman, “Continuity and Rational Functions”, in Proc. of the 44th Int’l Colloquium on Automata, Languages, and Programming (ICALP 2017), LIPIcs, Vol. 80, pp. 115:1–115:14, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2017.

A word-to-word function is continuous for a class of languages V if its inverse maps V -languages to V . This notion provides a basis for an algebraic study of transducers, and was integral to the characterization of the sequential transducers computable in some circuit complexity classes.

Here, we report on the decidability of continuity for functional transducers and some standard classes of regular languages. Previous algebraic studies of transducers have focused on the structure of the underlying input automaton, disregarding the output. We propose a comparison of the two algebraic approaches through two questions: When are the automaton structure and the continuity properties related, and when does continuity propagate to superclasses?

3.4 Compression with transducers

Olivier Carton (University Paris-Diderot, FR)

License © Creative Commons BY 3.0 Unported license
© Olivier Carton

Joint work of Verónica Becher, Olivier Carton, Pablo Ariel Heiber

We strengthen the theorem that establishes that deterministic finite transducers can not compress normal infinite words. We prove that, indeed, non-deterministic finite transducers, even augmented with a fixed number of counters, can not compress normal infinite words. However, there are push-down non-deterministic transducers that can compress normal infinite words. We also obtain new results on the preservation of normality with automata selectors. Complementing Agafonov’s theorem for prefix selectors, we show that suffix selectors also preserve normality. However, there are simple two-sided selectors that do not preserve normality.

3.5 Automatic Program Inversion using Symbolic Transducers

Loris d’Antoni (University of Wisconsin – Madison, US)

License © Creative Commons BY 3.0 Unported license
© Loris d’Antoni

URL <http://pages.cs.wisc.edu/~loris/papers/pldi17inversion.pdf>

We propose a fully-automated technique for inverting functional programs that operate over lists such as string encoders and decoders. We consider programs that can be modeled using symbolic extended finite transducers (s-EFTs), an expressive model that can describe complex list-manipulating programs while retaining

several decidable properties. Concretely, given a program P expressed as an s-EFT, we propose techniques for: 1) checking whether P is injective and, if that is the case, 2) building an s-EFT P^{-1} describing its inverse. We first show that it is undecidable to check whether an s-EFT is injective and propose an algorithm for checking injectivity for a restricted, but a practical class of s-EFTs. We then propose an algorithm for inverting s-EFTs based on the following idea: if an s-EFT is injective, inverting it amounts to inverting all its individual transitions. We leverage recent advances in program synthesis and show that the transition inversion problem can be expressed as an instance of the syntax-guided synthesis framework. Finally, we implement the proposed techniques in a tool called GENIC and show that GENIC can invert 13 out of 14 real complex string encoders and decoders, producing inverse programs that are almost identical to manually written ones.

3.6 A decidable logic for string transductions

Luc Dartois (Free University of Brussels, BE)

License © Creative Commons BY 3.0 Unported license
© Luc Dartois

Joint work of Luc Dartois, Emmanuel Filiot, Nathan Lhote

Main reference L. Dartois, E. Filiot, N. Lhote, “Decidable Logics for Transductions and Data Words”, arXiv:1701.03670 [cs.FL], 2017.

URL <https://arxiv.org/abs/1701.03670>

We introduce a logic, called LT, to express properties of transductions, i.e. binary relations from input to output (finite) words. In LT, the input/output dependencies are modeled via an origin function which associates with any position of the output word, the input position from which it originates. The logic LT can express all MSO-definable functions, and is incomparable with MSO-transducers for relations. Despite its high expressive power, we show, among other interesting properties, that LT has decidable satisfiability and equivalence problems. The transduction logic LT is shown to be expressively equivalent to a logic for data words, LD, up to some bijection from transductions with origin to data words (the origin of an output position becomes the data of that position). The logic LD, which is interesting in itself and extends in expressive power known logics for data words, is shown to have decidable satisfiability.

3.7 Recent Results on Word Transductions

Emmanuel Filiot (Free University of Brussels, BE)

License © Creative Commons BY 3.0 Unported license
© Emmanuel Filiot

Main reference E. Filiot, P.-A. Reynier, “Transducers, logic and algebra for functions of finite words”, SIGLOG News, Vol. 3(3), pp. 4–19, ACM, 2016.

URL <https://doi.org/10.1145/2984450.2984453>

This talk surveys some important recent results from the theory of word transductions. It motivates the need to have a Dagstuhl seminar on transductions that includes word transductions, and gives a very brief overview of the talks of this seminar related to word transductions. It is organised around three main pillars for word transductions: computational models (transducers), logical models, and algebraic characterizations. Both the cases of functional and relational transductions are presented.


Some parts of the talks are based on the following surveys:

References

- 1 Emmanuel Filiot and Pierre-Alain Reynier. *Transducers, logic and algebra for functions of finite words*, SIGLOG News, Vol. 3(3), pages 4-19, 2016.
- 2 Emmanuel Filiot. *Logic-Automata Connections for Transformations*. Proceedings of the Indian Conference of Logic and Applications (ICLA). Pages 30-57. 2015.

3.8 Untwisting two-way transducers in elementary time

Olivier Gauwin (University of Bordeaux, FR) and Anca Muscholl (University of Bordeaux, FR)

License  Creative Commons BY 3.0 Unported license
© Olivier Gauwin and Anca Muscholl

Functional transductions realized by two-way transducers (equivalently, by streaming transducers and by MSO transductions) are the natural and standard notion of “regular” mappings from words to words. It was shown recently (LICS’13) that it is decidable if such a transduction can be implemented by some one-way transducer, but the given algorithm has non-elementary complexity. We provide an algorithm of different flavor solving the above question, that has double exponential space complexity. We further apply our technique to decide whether the transduction realized by a two-way transducer can be implemented by a sweeping transducer, with either known or unknown number of passes.

3.9 Which classes of origin graphs are generated by transducers?

Bruno Guillon (University of Warsaw, PL)

License  Creative Commons BY 3.0 Unported license
© Bruno Guillon

Joint work of Mikołaj Bojańczyk, Laure Daviaud, Bruno Guillon, Vincent Penelle

Main reference M. Bojańczyk, L. Daviaud, B. Guillon, V. Penelle, “Which classes of origin graphs are generated by transducers?”, in Proc. of the 44th Int’l Colloquium on Automata, Languages, and Programming (ICALP 2017), LIPIcs, Vol. 80, pp. 114:1–114:13, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2017.

URL <http://dx.doi.org/10.4230/LIPIcs.ICALP.2017.114>

We consider the origin semantics of transducers (2-way automata with outputs, streaming string transducers, string-to-string MSO-transduction) which extends relations by providing a mapping from positions of the output into positions of the input, saying how the output originates from the input. In the other hand, every relation on words can be extended by such a mapping, leading to the notion of origin transduction, viewed as set of particular graph, namely the origin graphs. We exhibit structural properties of origin transductions which are necessary and sufficient for characterising the origin semantics of streaming string transducers.

3.10 On Reversible Transducers

Ismaël Jecker (Free University of Brussels, BE)

License © Creative Commons BY 3.0 Unported license
© Ismaël Jecker

Joint work of Dartois, Luc; Fournier, Paulin; Lhote, Nathan

Main reference L. Dartois, P. Fournier, I. Jecker, N. Lhote, “On Reversible Transducers”, in Proc. of the 44th Int’l Colloquium on Automata, Languages, and Programming (ICALP 2017), LIPIcs, Vol. 80, pp. 113:1–113:12, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2017.

URL <http://dx.doi.org/10.4230/LIPIcs.ICALP.2017.113>

Deterministic two-way transducers define the robust class of regular functions which, among other good properties, is closed under composition. However, the best known algorithms for composing two-way transducers cause a double exponential blow-up in the size of the inputs. We expose a class of transducers for which the composition has polynomial complexity. It is the class of reversible transducers, for which the computation steps can be reversed deterministically. While in the one-way setting this class is not very expressive, we prove that any two-way transducer can be made reversible through a single exponential blow-up. As a consequence, we obtain that the composition of two-way transducers can be done with a single exponential blow-up in the number of states.

3.11 Towards algebraic and logical characterizations of rational functions

Nathan Lhote (University of Bordeaux, FR), Emmanuel Filiot (Free University of Brussels, BE), and Olivier Gauwin (University of Bordeaux, FR)

License © Creative Commons BY 3.0 Unported license
© Nathan Lhote, Emmanuel Filiot, and Olivier Gauwin

Regular languages are characterized by several different formalisms including monadic second-order logic (MSO) and also congruences of finite index. Many links have been established between fragments of MSO and varieties of congruences (sets of congruences with good closure properties). These characterization often yield an effective way to decide if a language can be expressed in a given logic, the most prominent example being the Schutzenberger/McNaughton-Papert theorem which characterizes the first-order languages as the languages recognized by aperiodic congruences. We present how to lift some of these results to word-to-word transductions, with some assumptions on the logical fragment amounting to having access to the linear order over words.

3.12 Tree Automata in Parsing and Machine Translation

Andreas Maletti (*Universität Leipzig, DE*)

License © Creative Commons BY 3.0 Unported license
© Andreas Maletti

Joint work of Fabienne Braune, Joost Engelfriet, Zoltan Fülöp, Daniel Quernheim, Nina Seemann

Main reference J. Engelfriet, Z. Fülöp, A. Maletti, “Composition Closure of Linear Extended Top-down Tree Transducers”, *Theory Comput. Syst.* 60(2):129–171, 2017.

URL <https://doi.org/10.1007/s00224-015-9660-2>

We will discuss how tree automata were rediscovered in the area of statistical parsing of natural language sentences and demonstrate that some techniques developed in that area might also be beneficial in automata theory. On the example of syntax-based machine translation, we will demonstrate the other direction showing how automata theory can provide solutions to problems in natural language processing. With the identification of the exact expressive power in terms of standard models and known closure properties for them, we developed a syntax-based translation system that can translate not only a single parse, but rather the full parse forest delivered by the parser.

3.13 On repetitive right application of B-terms

Keisuke Nakano (*The University of Electro-Communications – Tokyo, JP*)

License © Creative Commons BY 3.0 Unported license
© Keisuke Nakano

Joint work of Mirai Ikebuchi, Keisuke Nakano

Main reference M. Ikebuchi, K. Nakano, “On repetitive right application of B-terms”, arXiv:1703.10938 [cs.LO], 2017.

URL <http://arxiv.org/abs/1703.10938>

B-terms are built from the B combinator alone defined by $B f g x = f (g x)$, which is well-known as a function composition operator. This talk introduces an interesting property of B-terms, that is, whether repetitive right applications of a B-term circulates or not. Conditions for B-terms to and not to have the property are discussed. Specifically, we show that some of infinitely many B-terms which do not have the property can be given as a regular tree language.

3.14 Two twinning properties with applications to register minimisation in streaming string transducers

Pierre-Alain Reynier (*Aix-Marseille University, FR*)

License © Creative Commons BY 3.0 Unported license
© Pierre-Alain Reynier

Joint work of Laure Daviaud, Ismaël Jecker, Jean-Marc Talbot and Didier Villevalois

Main reference L. Daviaud, P.-A. Reynier, J.-M. Talbot, “A Generalised Twinning Property for Minimisation of Cost Register Automata”, in *Proc. of the 31st Annual ACM/IEEE Symp. on Logic in Computer Science (LICS 2016)*, pp. 857–866, ACM, 2016.

URL <http://dx.doi.org/10.1145/2933575.2934549>

In this talk, I present recent results on transducers of finite words. In this setting, input-deterministic machines are less expressive than non-deterministic ones. Non-deterministic transducers admitting an equivalent deterministic one are characterized by the twinning property introduced by Choffrut in 1977.

I will present two modifications of this property allowing to solve two different minimization problems :

- first, a simple restriction of the recently introduced model of streaming string transducers, called appending SST, is expressively equivalent to non-deterministic functional transducers. The twinning property of order k , introduced in [1], precisely characterizes the transducers that can be realized by an appending SST with k registers.
- second, though not realizable by an input-deterministic transducer, some functions, called multi-sequential, may be realized as finite unions of such machines. I present the branching twinning property of order k , introduced in [2], allowing to characterize functions that can be realized as a union of k input-deterministic transducers.

References

- 1 Laure Daviaud, Pierre-Alain Reynier, Jean-Marc Talbot. A Generalised Twinning Property for Minimisation of Cost Register Automata. LICS 2016: 857-866
- 2 Laure Daviaud, Ismaël Jecker, Pierre-Alain Reynier, Didier Villevalois: Degree of Sequentiality of Weighted Automata. FoSSaCS 2017: 215-230

3.15 Transducers as odometers in rational base numeration systems

Jacques Sakarovitch (*Telecom ParisTech, FR*)

License  Creative Commons BY 3.0 Unported license

© Jacques Sakarovitch

Joint work of Shigeki Akiyama, Victor Marsault, Jacques Sakarovitch

The definition of numeration systems with rational base, in a joint work with S. Akiyama and Ch. Frougny, has allowed to make some progress in a number theoretic problem, by means of automata theory and combinatorics of words. At the same time, it raised the problem of understanding the structure of the sets of the representations of the integers in these systems from the point of view of formal language theory.

At first sight, these sets look rather chaotic and do not fit well in the classical Chomsky hierarchy of languages. They all enjoy a property that makes them defeat, so to speak, any kind of iteration lemma. On the other hand, these sets also exhibit remarkable regularity properties.

During the recent years, these regularities have been studied in a series of joint papers with my student V. Marsault. In particular, we have shown that periodic signatures are characteristic of the representation languages in rational base numeration systems. For given rational base p/q , the representation language naturally define an infinite tree $T_{p/q}$ the nodes of which are the integers. Every integer n is then associated with the infinite minimal branch whose label is an infinite word w_n which is characteristic of n . The next remarkable property is that the function that maps w_n onto w_{n+1} is realized by a letter-to-letter transducer whose underlying graph is precisely $T_{p/q}$.

References

- 1 S. Akiyama, Ch. Frougny, and J. Sakarovitch. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math.*, 168:53–91, 2008.
- 2 S. Akiyama, V. Marsault, and J. Sakarovitch. Auto-similarity in rational base number systems. *Proc. of WORDS 2013*, LNCS 8079:34–45, 2013.
- 3 V. Marsault and J. Sakarovitch. Trees and languages with periodic signature *Indagationes Mathematicae* 28:221–246, 2017.

3.16 Equivalence of Deterministic Tree-to-String Transducers is Decidable

Helmut Seidl (TU München, DE), Gregor Kemper, and Sebastian Maneth (University of Edinburgh, GB)

License © Creative Commons BY 3.0 Unported license
 © Helmut Seidl, Gregor Kemper, and Sebastian Maneth
Main reference H. Seidl, S. Maneth, G. Kemper, “Equivalence of Deterministic Top-down Tree-to-String Transducers is Decidable”, in Proc. of the 56th Ann. Symp. on Foundations of Computer Science (FOCS 2015), pp. 943–962, IEEE, 2015.
URL <https://doi.org/10.1109/FOCS.2015.62>

We show that equivalence of deterministic top-down tree-to-string transducers is decidable, thus solving a long standing open problem in formal language theory. For our main result, we prove that equivalence can be certified by means of inductive invariants using polynomial ideals. This allows us to construct two semi-algorithms, one searching for a proof of equivalence, one for a witness of non-equivalence. Furthermore, we extend our result to deterministic top-down tree-to-string transducers which produce output not in a free monoid but in a free group.

3.17 Combinator Expressions for Regular Transformations

Krishna Shankaranarayanan (Indian Institute of Technology – Mumbai, IN)

License © Creative Commons BY 3.0 Unported license
 © Krishna Shankaranarayanan
Joint work of Vrunda Dave, Paul Gustin, Krishna Shankaranarayanan

MSO transductions, functional transductions realized by two-way transducers, as well as streaming string transducers all capture regular transformations of languages. In this work, we show that every regular function captured by a two-way transducer can be constructed from constant functions using the combinators of choice, split-sum and iterated sum, that are analogs of unambiguous union, concatenation and Kleene-star respectively. This is achieved by unambiguously factorizing the input language into forests of bounded height.

Our proof extends to finite as well as infinite words.

3.18 From Two-way to One-way Finite State Transducers: a Shepherdson’s Approach

Jean-Marc Talbot (Aix-Marseille University, FR)

License © Creative Commons BY 3.0 Unported license
 © Jean-Marc Talbot
Joint work of Benjamin Monmege, Pierre-Alain Reynier, Jean-Marc Talbot

We study functions over finite words defined by one-way and two-way finite state transducers (1FST and 2FST, respectively). 2FST are strictly more powerful than 1FST and a natural question is then to decide, given a 2FST describing a function, whether there exists an equivalent 1FST. Filiot, Gauwin, Reynier and Servais showed in 2013 that this problem is decidable, using a procedure based on the construction of Rabin and Scott for transforming two-way into one-way automata, yielding a non-elementary complexity. In this article, we

prove that the problem is in 2-EXPSPACE. Our proof technique relies on a new representation of the transducer based on the crossing sequences introduced in Shepherdson's construction (an alternative to the one of Rabin and Scott), and on a normalisation of this transducer. In case of one-way definability, we produce an equivalent 1FST (deterministic with look-ahead) of size at most triply-exponential in the size of the original 2FST.

Participants

- Johanna Björklund
University of Umeå, SE
- Adrien Boiret
INRIA Lille, FR
- Mikolaj Bojanczyk
University of Warsaw, PL
- Michaël Cadilhac
Universität Tübingen, DE
- Olivier Carton
University Paris-Diderot, FR
- Loris d'Antoni
University of Wisconsin –
Madison, US
- Luc Dartois
Free University of Brussels, BE
- Frank Drewes
University of Umeå, SE
- Emmanuel Filiot
Free University of Brussels, BE
- Olivier Gauwin
University of Bordeaux, FR
- Bruno Guillon
University of Warsaw, PL
- Florent Jacquemard
IRCAM & INRIA – Paris, FR
- Ismaël Jecker
Free University of Brussels, BE
- Aurélien Lemay
INRIA Lille, FR
- Nathan Lhote
University of Bordeaux, FR
- Christof Löding
RWTH Aachen, DE
- Andreas Maletti
Universität Leipzig, DE
- Sebastian Maneth
University of Edinburgh, GB
- Anca Muscholl
University of Bordeaux, FR
- Keisuke Nakano
The University of
Electro-Communications –
Tokyo, JP
- Joachim Niehren
INRIA Lille, FR
- Chih-Hao Luke Ong
University of Oxford, GB
- Raphaela Palenta
TU München, DE
- Rogério Reis
University of Porto, PT
- Pierre-Alain Reynier
Aix-Marseille University, FR
- Jacques Sakarovitch
Telecom ParisTech, FR
- Helmut Seidl
TU München, DE
- Frédéric Servais
Free University of Brussels, BE
- Krishna Shankaranarayanan
Indian Institute of Technology –
Mumbai, IN
- Jean-Marc Talbot
Aix-Marseille University, FR
- Margus Veanes
Microsoft Corporation –
Redmond, US

