Report from Dagstuhl Seminar 23202

Regular Transformations

Rajeev Alur^{*1}, Mikołaj Bojańczyk^{*2}, Emmanuel Filiot^{*3}, Anca Muscholl^{*4}, and Sarah Winter^{$\dagger 5$}

- 1 University of Pennsylvania - Philadelphia, US. alur@cis.upenn.edu
- 2 University of Warsaw, PL. bojan@mimuw.edu.pl
- 3 UL - Brussels, BE. efiliot@gmail.com
- 4 University of Bordeaux, FR. anca@labri.fr
- 5 UL - Brussels, BE. Sarah.Winter@ulb.be

- Abstract -

This report documents the program and the outcomes of the Dagstuhl Seminar 23202 "Regular Transformations". The goal of this seminar was to advance on a list of topics about transducers that have gathered much interest recently, and to explore new connections between the theory of regular transformations and its applications in linguistics.

Seminar May 14-17, 2023 - https://www.dagstuhl.de/23202

2012 ACM Subject Classification Theory of computation \rightarrow Formal languages and automata theory

Keywords and phrases transducers; (poly-)regular functions; linguistic transformations Digital Object Identifier 10.4230/DagRep.13.5.96

1 **Executive Summary**

Rajeev Alur Mikołaj Bojańczyk Emmanuel Filiot Anca Muscholl

> License \bigcirc Creative Commons BY 4.0 International license Rajeev Alur, Mikołaj Bojańczyk, Emmanuel Filiot, and Anca Muscholl

Transducers, i.e. automata with outputs, are one of the oldest computational models in theoretical computer science. They even predate the usual boolean automata, going back to Church, Shannon, Moore, and Mealy. In spite of being considered too complex¹, transducers remained an active research topic ever since. Also connections to practical applications in efficient processing of streaming data have been established recently. The purpose of this Dagstuhl Seminar was to gather researchers to discuss recent developments on transducers and their applications.

The goal was twofold, to advance on a list of topics about transducers that have gathered much interest recently, and to explore new connections with researchers from linguistics. We enjoyed very interesting talks about:

- polyregular functions over trees and growth of regular functions
- data transducer synthesis and transducers over data words
- decomposition of finite-valued streaming string transducers

Rabin and Scott argued in their 1959 paper that they are better off by "doing away with a complicated output function and having our machines simply give 'yes' or 'no' answers.'



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Regular Transformations, Dagstuhl Reports, Vol. 13, Issue 5, pp. 96–113

Editor / Organizer

Editorial Assistant / Collector

Editors: Rajeev Alur, Mikołaj Bojańczyk, Emmanuel Filiot, Anca Muscholl, and Sarah Winter DAGSTUHL Dagstuhl Reports

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

- automata over series-parallel graphs and transducers for data management
- learning linguistic transformations and large language models as transducers
- tree transducers: class characterisations, macro tree transducers with semantic constraints
- transducers and complexity

The scientific programme was quite dense, given that we had only 3 days and almost all participants proposed to give a talk. Exchanges were very lively, and we are confident that new research directions and new collaborations emerged from this seminar.

Rajeev Alur, Mikołaj Bojańczyk, Emmanuel Filiot, Anca Muscholl

2 Table of Contents

Executive Summary Rajeev Alur, Mikolaj Bojańczyk, Emmanuel Filiot, and Anca Muscholl 9		
Overvi	iew of Talks	
Not Sha	ions of Symmetry in Transducers and Games ull Almagor	
Reg <i>Raj</i> e	ular languages of series-parallel graphs <i>eev Alur</i>	
Poly Mik	vregular functions on unordered trees of bounded height olaj Bojańczyk	
Fron Aut <i>Elis</i>	m Finite-Valued Nondeterministic Transducers to Deterministic Two-Tape omata abet Burjons	
A C Mic	Sircuit Complexity Approach to Transductions haël Cadilhac	
Det Oliv	erminization of transducers over real numbers	
$\operatorname{Rec} Rya$	urrent Neural Network LMs as weighted formal language recognizers <i>n Cotterell</i>	
From Luc	m Regular Transducer Expressions to Reversible Transducers <i>Dartois</i>	
A C Léo	Generic Solution to Register-bounded Synthesis for Systems over Data WordsExibard103	
$\begin{array}{c} \operatorname{Reg} \\ Em \end{array}$	ular functions of infinite words manuel Filiot	
Mac Pau	The Transducers with semantic constraints $l \ Gallot$	
Solv Mat	ving String Constraints with References using Streaming String Transducers thew Hague	
Reg	ular Transformations in Linguistics	
The Ism	Decomposition Theorem for streaming string transducers aël Jecker	
Tow Nat	vards regular functions with exponential growth <i>han Lhote</i>	
On Anc	finite-valuedness of streaming string transducers <i>a Muscholl</i>	
$\begin{array}{c} \text{High} \\ L\hat{e} \end{array}$	her-order phenomena in transducer theory Thành Dũng Nguyễn	
Lea: Jon	rning (sub)regular transformations Rawski	

	A transducer model for streaming enumeration problems Cristian Riveros
	Definability Results for Tree Transducers Sebastian Maneth
	When is a Bottom-up Deterministic Tree Transducer Top-down Deterministic?Helmut Seidl110
	An Algebraic Theory for Single-use Transducers Over Data Words Rafal Stefanski
	How to decide Functionality of Compositions of Top-Down Tree Transducers Martin Vu
	A Regular and Complete Notion of Delay for Streaming String Transducers Sarah Winter
O]	pen problems
	Unambiguous Single-use Register Automata Rafal Stefanski


```
3 Overview of Talks
```

3.1 Notions of Symmetry in Transducers and Games

Shaull Almagor (Technion – Haifa, IL)

License	© Creative Commons BY 4.0 International license
	© Shaull Almagor
Joint work of	Shaull Almagor, Antonio Abu-Nassar, Shai Guendelman
Main reference	Antonio Abu Nassar, Shaull Almagor: "Simulation by Rounds of Letter-To-Letter Transducers", in
	Proc. of the 30th EACSL Annual Conference on Computer Science Logic, CSL 2022, February 14-19,
	2022, Göttingen, Germany (Virtual Conference), LIPIcs, Vol. 216, pp. 3:1–3:17, Schloss Dagstuhl –
	Leibniz-Zentrum für Informatik, 2022.
URL	https://doi.org//10.4230/LIPIcs.CSL.2022.3

We consider various notions of "symmetry" for transducers and for games. The main focus is on process symmetry, whereby several processes interact, but may be unaware of the identification given to them by the system. They therefore must act interchangeably, in order to satisfy some specifications.

We examine definitions of process-symmetry, and the algorithms and decision problems pertaining to them. Specifically, we define symmetry in probabilistic transducers, symmetryby-rounds for deterministic transducers, and symmetry in games.

3.2 Regular languages of series-parallel graphs

Rajeev Alur (University of Pennsylvania – Philadelphia, US)

```
    License 
        Gerative Commons BY 4.0 International license
        Second Rajeev Alur

    Joint work of Rajeev Alur, Caleb Stanford, Christopher Watson
        Main reference
        Rajeev Alur, Caleb Stanford, Christopher Watson: "A Robust Theory of Series Parallel Graphs",
        Proc. ACM Program. Lang., Vol. 7(POPL), pp. 1058–1088, 2023.
        URL https://doi.org/10.1145/3571230
```

Motivated by distributed data processing applications, we introduce a class of labeled directed acyclic graphs constructed using sequential and parallel composition operations, and study automata and logics over them. We show that deterministic and non-deterministic acceptors over such graphs have the same expressive power, which can be equivalently characterized by Monadic Second-Order logic and the graded mu-calculus. We establish closure under composition operations and decision procedures for membership, emptiness, and inclusion. A key feature of our graphs, called "synchronized series-parallel graphs" (SSPG), is that parallel composition introduces a synchronization edge from the newly introduced source vertex to the sink. The transfer of information enabled by such edges is crucial to the determinization construction, which would not be possible for the traditional definition of series-parallel graphs.

SSPGs allow both ordered ranked parallelism and unordered unranked parallelism. The latter feature means that in the corresponding automata, the transition function needs to account for an arbitrary number of predecessors by counting each type of state only up to a specified constant, thus leading to a notion of "counting complexity" that is distinct from the classical notion of state complexity. The determinization construction translates a nondeterministic automaton with n states and k counting complexity to a deterministic automaton with 2^{n^2} states and kn counting complexity, and both these bounds are shown to be tight. Furthermore, for nondeterministic automata a bound of 2 on counting complexity suffices without loss of expressiveness.

3.3 Polyregular functions on unordered trees of bounded height

Mikołaj Bojańczyk (University of Warsaw, PL)

License
Creative Commons BY 4.0 International license
Mikołaj Bojańczyk
Joint work of Mikołaj Bojańczyk, Bartek Klin

We consider first-order interpretations that input and output trees of bounded height. The corresponding functions have polynomial output size, since a first-order interpretation can use a -tuple of input nodes to represent a single output node. We prove that the equivalence problem for such functions is decidable, i.e. given two such interpretations, one can decide whether for every input tree, the two output trees are isomorphic. This result is incomparable to the open problem about deciding equivalence for polyregular string-to-string functions. We also give a calculus, based on prime functions and combinators, which derives all first-order interpretations for unordered trees of bounded height. The calculus is based on a type system, where the type constructors are products, co-products and multisets. Thanks to the results about tree-to-tree interpretations, the equivalence problem is decidable for this calculus.As an application of our results, we show that the equivalence problem is decidable for first-order interpretations between classes of graphs that have bounded tree depth. In all cases studied in this paper, first-order logic and MSO have the same expressive power, and hence all results apply also to MSO interpretations.

3.4 From Finite-Valued Nondeterministic Transducers to Deterministic Two-Tape Automata

Elisabet Burjons (York University – Toronto, CA)

License
 Gereative Commons BY 4.0 International license
 Selisabet Burjons

 Joint work of Elisabet Burjons, Fabian Frei, Martin Raszyk

 Main reference Elisabet Burjons, Fabian Frei, Martin Raszyk: "From Finite-Valued Nondeterministic Transducers to
 Deterministic Two-Tape Automata", in Proc. of the 36th Annual ACM/IEEE Symposium on Logic
 in Computer Science, LICS 2021, Rome, Italy, June 29 – July 2, 2021, pp. 1–13, IEEE, 2021.

 URL https://doi.org/10.1109/LICS52264.2021.9470688

Every nondeterministic finite transducer defines a binary relation associating each input word with all output words that the transducer can successfully produce on the given input. Finite-valued transducers are those for which there is a finite upper bound on the number of output words that the relation associates with every input word. We characterize finite-valued, functional, and unambiguous nondeterministic transducers whose relations can be verified by a deterministic two-tape automaton and show how to construct such an automaton if one exists. The necessary and sufficient property for a finite-valued transduction to be verifiable by a deterministic two-tape transducer is called bounded trailing, which is a more refined condition than bounded variation. We prove the undecidability of the criterion, i.e., it is undecidable to know whether a transduction has bounded trailing or not.

3.5 A Circuit Complexity Approach to Transductions

Michaël Cadilhac (DePaul University – Chicago, US)

License	© Creative Commons BY 4.0 International license
	© Michaël Cadilhac
Joint work of	Michaël Cadilhac, Andreas Krebs, Michael Ludwig, Charles Paperman
Main reference	Michaël Cadilhac, Andreas Krebs, Michael Ludwig, Charles Paperman: "A Circuit Complexity
	Approach to Transductions", in Proc. of the Mathematical Foundations of Computer Science 2015 -
	40th International Symposium, MFCS 2015, Milan, Italy, August 24-28, 2015, Proceedings, Part I,
	Lecture Notes in Computer Science, Vol. 9234, pp. 141–153, Springer, 2015.
URL	https://doi.org//10.1007/978-3-662-48057-1_11

Low circuit complexity classes and regular languages exhibit very tight interactions that shade light on their respective expressiveness. We propose to study these interactions at a functional level, by investigating the deterministic rational transductions computable by constant-depth, polysize circuits. To this end, a circuit framework of independent interest that allows variable output length is introduced. Relying on it, there is a general characterization of the set of transductions realizable by circuits. It is then decidable whether a transduction is definable in AC^0 and, assuming a well-established conjecture, the same for ACC^0 .

3.6 Determinization of transducers over real numbers

Olivier Carton (Université Paris Cité, FR)

```
License 
Creative Commons BY 4.0 International license
Olivier Carton
Joint work of Olivier Carton, Alexis Bès, Christian Choffrut
```

We consider one-way transducers that realize real number functions. We show that if the real function is continuous, the transducer can be made input-deterministic up to a change of the digit set used in the output. This is very similar to Avižienis numeration system used to perform sequences of additions.

3.7 Recurrent Neural Network LMs as weighted formal language recognizers

Ryan Cotterell (ETH Zürich, CH)

License ☺ Creative Commons BY 4.0 International license © Ryan Cotterell Joint work of Ryan Cotterell, Anej Svete

Studying language models (LMs) in terms of well-understood formalisms allows us to precisely characterize their abilities and limitations. Previous work has investigated the expressive power of recurrent neural network (RNN) LMs in terms of their capacity to recognize unweighted formal languages. However, LMs do not describe unweighted formal languages rather, they define *probability distributions* over strings. In this work, we study what classes of such probability distributions RNN LMs can represent, which allows us to make more direct statements about their capabilities. We show that simple RNNs are equivalent to a subclass of probabilistic finite-state automata, and can thus model a strict subset of probability distributions expressible by finite-state models. Furthermore, we study the space complexity of representing finite-state LMs with RNNs. We show that, to represent an arbitrary deterministic finite-state LM with N states over an alphabet Σ , an RNN requires $\Omega(N|\Sigma|)$ neurons. These results present a first step towards characterizing the classes of distributions RNN LMs can represent and thus help us understand their capabilities and limitations.

3.8 From Regular Transducer Expressions to Reversible Transducers

Luc Dartois (University Paris-Est – Créteil, FR)

License
 © Creative Commons BY 4.0 International license
 © Luc Dartois

 Joint work of Luc Dartois, Paul Gastin, R. Govind, Shankaranarayanan Krishna
 Main reference Luc Dartois, Paul Gastin, R. Govind, Shankara Narayanan Krishna: "Efficient Construction of Reversible Transducers from Regular Transducer Expressions", CoRR, Vol. abs/2202.04340, 2022.
 URL https://arxiv.org/abs/2202.04340

The class of regular transformation enjoy several equivalent characterizations (MSO Transductions, Deterministic two-way transducers, Streaming String Transducers). Regular Transducer Expressions (RTE) can be seen as a lift of Rational Expressions to string functions. They offer a combinator system to describe regular transformations. In this talk, we present a way to construct, given an RTE, an equivalent operational machine, in our case a reversible, i.e. deterministic and co-deterministic, two-way transducer. The procedure is doubly exponential for the class of regular functions, and simply exponential for the rational functions.

3.9 A Generic Solution to Register-bounded Synthesis for Systems over Data Words

Léo Exibard (Gustave Eiffel University – Marne-la-Vallée, FR)

License
 Creative Commons BY 4.0 International license
 © Léo Exibard

 Joint work of Léo Exibard, Emmanuel Filiot, Ayrat Khalimov
 Main reference Léo Exibard, Emmanuel Filiot, Ayrat Khalimov: "A Generic Solution to Register-bounded Synthesis with an Application to Discrete Orders", CoRR, Vol. abs/2205.01952, 2022.

URL https://doi.org//10.48550/arXiv.2205.01952

In this talk, we consider synthesis of reactive systems interacting with environments using an infinite data domain. A popular formalism for specifying and modelling those systems is register automata and transducers. They extend finite-state automata by adding registers to store data values and to compare the incoming data values against stored ones. Synthesis from nondeterministic or universal register automata is undecidable in general. However, its register-bounded variant, where additionally a bound on the number of registers in a sought transducer is given, is known to be decidable for universal register automata which can compare data for equality, i.e., for data domain $(\mathbb{N}, =)$.

After briefly reviewing this result, we extend it to the domain $(\mathbb{N}, <)$ of natural numbers with linear order. Our solution is generic: we define a sufficient condition on data domains (regular approximability) for decidability of register-bounded synthesis. It allows one to use simple language-theoretic arguments and avoid technical game-theoretic reasoning. Further, by defining a generic notion of reducibility between data domains, we show the decidability of synthesis in the domain $(\mathbb{N}^d, <^d)$ of tuples of numbers equipped with the component-wise partial order and in the domain (Σ^*, \prec) of finite strings with the prefix relation.

3.10 Regular functions of infinite words

Emmanuel Filiot (UL – Brussels, BE)

- License O Creative Commons BY 4.0 International license
 - © Emmanuel Filiot
- Joint work of Rajeev Alur, Olivier Carton, Gaëtan Douéneau-Tabot, Emmanuel Filiot, Sarah Winter
- Main reference Olivier Carton, Gaëtan Douéneau-Tabot, Emmanuel Filiot, Sarah Winter: "Deterministic Regular Functions of Infinite Words", in Proc. of the 50th International Colloquium on Automata, Languages, and Programming, ICALP 2023, July 10-14, 2023, Paderborn, Germany, LIPIcs, Vol. 261, pp. 121:1–121:18, Schloss Dagstuhl Leibniz-Zentrum für Informatik, 2023.
 INPL https://doi.org/10.4230/UPLes.ICALP.2023.121
 - URL https://doi.org//10.4230/LIPIcs.ICALP.2023.121

Regular functions of infinite words are (partial) functions realized by deterministic twoway transducers with infinite look-ahead. Equivalently, Alur et. al. have shown that they correspond to functions realized by deterministic Muller streaming string transducers, and to functions defined by MSO-transductions [1]. Regular functions are however not computable in general (for a classical extension of Turing computability to infinite inputs), and we consider in this talk the class of deterministic regular functions of infinite words, realized by deterministic two-way transducers without look-ahead. We prove in [2] that it is a well-behaved class of functions: they are computable, closed under composition, characterized by the guarded fragment of MSO-transductions, by deterministic Büchi streaming string transducers, by deterministic two-way transducers with finite look-ahead, and by finite compositions of sequential functions and one fixed basic function called map-copy-reverse.

References

- Rajeev Alur, Emmanuel Filiot, Ashutosh Trivedi. Regular transformations of infinite strings. LICS 2012: 65-74.
- 2 Olivier Carton, Gaëtan Douéneau-Tabot, Emmanuel Filiot, Sarah Winter. *Deterministic regular functions of infinite words*. ICALP 2023: 121:1-121:18.

3.11 Macro Tree Transducers with semantic constraints

Paul Gallot (Universität Bremen, DE)

License
Creative Commons BY 4.0 International license
Paul Gallot
Joint work of Charles Peyrat, Paul Gallot, Sebastian Maneth

Macro Tree Transducers (MTT) are a very expressive model of tree transducers, but its restriction to tree-to-tree functions of Linear Size Increase (LSI) has been shown to be as expressive as MSO tree transductions. In this talk we study variations of this LSI semantic contraint, namely Linear Height Increase (LHI) and Linear input Size to output Height Increase (LSHI), and provide a new normal form for MTTs called Depth-normal form. This new normal form allows an elegant characterization of the LHI and LSHI constraints, which in turn gives an algorithm for deciding if a MTT respects these constraints.

3.12 Solving String Constraints with References using Streaming String Transducers

Matthew Hague (Royal Holloway, University of London, GB)

License

 Creative Commons BY 4.0 International license
 Matthew Hague

 Main reference Taolue Chen, Alejandro Flores-Lamas, Matthew Hague, Zhilei Han, Denghang Hu, Shuanglong Kan, Anthony W. Lin, Philipp Rümmer, Zhilin Wu: "Solving string constraints with Regex-dependent functions through transducers with priorities and variables", Proc. ACM Program. Lang., Vol. 6(POPL), pp. 1–31, 2022.

URL https://doi.org//10.1145/3498707

String constraints arise naturally when analysing string-manipulating programs. For example, during symbolic execution, determining whether an identified program path is feasible means checking if a set of constraints (from program branches, assignments, and loop conditions) has some satisfying assignment. That, there is some input that will lead to the given path being executed. This then requires constraint satisfaction tools to be able to handle complex string operations, as well as basics such as concatenation and containment in a regular language.

One such operation is the "replaceall" operation, where a matched pattern is replaced in one string to obtain another. The replacement may itself use "references" to the matched text. For example, an author list in the format "first-name last-name" may be converted to "last-name first-name".

Streaming String transducers provide an elegant model for such string operations. Moreover, they have an important property that the pre-image of a regular language under a replaceall remains regular. Since we additionally need to ensure that matches are "left-most longest", we introduce "prioritised streaming string transducers" along the lines of Berglund and van der Merwe. Using these techniques, we extend the string constraint solver OSTRICH to support this replaceall operation.

In this talk, we discuss the application of streaming string transducers in constraint satsifaction, and briefly summarise some theoretical and experimental results.

3.13 Regular Transformations in Linguistics

Jeffrey Heinz (Stony Brook University, US)

License
 Creative Commons BY 4.0 International license
 I Jeffrey Heinz

 Joint work of Jeffrey Heinz, Dakotah Lambert
 Main reference Dakotah Lambert: "Unifying Classification Schemes for Languages and Processes with Attention to Locality and Relativizations Thereof". Ph.D. thesis, Stony Brook University. 2022.

URL https://vvulpes0.github.io/PDF/dissertation.pdf

Recent work in theoretical computational linguistics highlights the importance of regular transformations over strings and trees in characterizing linguistic generalizations. This talk reviews some of the main results of this research program. Regular transformations provide an upper bound on morpho-phonological processes in natural languages. Particular subclasses appear to be sufficient, and these subclasses are generally (much) less than FO(<). Additionally, these subclasses are learnable with relatively low time and data complexity in ways the full class of regular transformations is not.

3.14 The Decomposition Theorem for streaming string transducers

Ismaël Jecker (University of Warsaw, PL)

License
Creative Commons BY 4.0 International license
Ismaël Jecker
Joint work of Ismaël Jecker, Emmanuel Filiot, Christof Löding, Anca Muscholl, Gabriele Puppis, Sarah Winter

The recently introduced notion of delay for streaming string transducers (SST) has various applications, one of which is demonstrated through the DecompositionTheorem: This theorem states that every k-valued SST (where each input is mapped to at most k outputs) can be decomposed into a finite union of k single-valued SST. While similar results have been established in other settings such as finite state transducers, tree transducers and weighted automata, the specific case of SST remained open up to this day. We present a solution that relies on the notion of delay to extract, one after the other, k single-valued SST from a given k-valued SST, thus creating the desired decomposition. The Decomposition Theorem holds significant interest due to its implications: it reveals that k-valued SST recognise a robust class of functions, for instance equivalent to those recognised by finite-valued transducers. Moreover, it shows that this class enjoys good algorithmic properties. For instance, equivalence between finite-valued SST can be decided efficiently.

3.15 Towards regular functions with exponential growth

Nathan Lhote (Aix-Marseille University, FR)

License
Creative Commons BY 4.0 International license
Nathan Lhote
Joint work of Nathan Lhote, Emmanuel Filiot, P.-A. Reynier

We present a class of transductions with exponential growth based on logical interpretations, but with monadic second-order parameters instead of first-order ones. We study some of the closure properties of this class as well as potential alternative characterizations.

3.16 On finite-valuedness of streaming string transducers

Anca Muscholl (University of Bordeaux, FR)

© Anca Muscholl

License \bigcirc Creative Commons BY 4.0 International license

Joint work of Anca Muscholl, Emmanuel Filiot, Ismaël Jecker, Christof Löding, Gabriele Puppis, Sarah Winter

While one-valued streaming string transducers define a very robust class of transductions, finite-valued ones still enjoy some good properties. Recent results (see talk by Ismaël Jecker) show that they are equivalent to finite-valued two-way transducers. Moreover, jointly with G. Puppis, we have shown their equivalence problem to be decidable (ICALP 2019).

In this talk I presented a conjecture towards deciding whether a streaming string transducer is finitely valued. The characterisation conjectured reminds of the one for one-way transducers obtained by A. Weber, and exploits ideas about state invariants involved in the equivalence decision procedure for finitely-valued streaming string transducers.

3.17 Higher-order phenomena in transducer theory

Lê Thành Dũng Nguyễn (ENS – Lyon, FR)

License ☺ Creative Commons BY 4.0 International license © Lê Thành Dũng Nguyễn Joint work of Lê Thành Dũng Nguyễn, Cécilia Pradic

This talk was about the connections between transducer theory and functional programming. An important role is played by type systems inspired from linear logic – linearity in programming language theory provides counterparts to various "copyless" or "single-use" restrictions on automata.

- 1. In our "implicit automata in typed λ -calculi" research programme [1, 2], Cécilia Pradic and I show that natural questions concerning the expressive power of theoretical functional programming languages turn out to have automata-theoretic answers. In particular, the study of a linear λ -calculus led us to discover a robust subclass of polyregular functions [3].
- 2. Meanwhile, machine models manipulating higher-order data (functions as first-class values, that may themselves take functions as arguments), taking inspiration from the higher-order grammars of the 1970s, arise from the study of composition hierarchies of transducers [4]. In this context, transducers whose memory stores linear λ-terms have been used to characterize MSO transductions of trees [5] which is very close to one of the results of [2].

An example of phenomenon (2) is the composition of streaming string transducers (SSTs, the topic of many other talks): their registers store concatenable strings, which should be seen as order-1 functions on appendable strings, so composing n SSTs results in a machine using order-n registers.

With this point of view, the natural counterpart of SSTs on trees is a bottom-up transducer whose registers store tree *contexts*, as noted in [6, 7]. But in fact, this model is none other than the much older *macro tree transducer* (MTT) [8]! A recursive equation such as $q_1\langle a(t, u)\rangle = b(q_2\langle u\rangle, q_3\langle t\rangle)$, read as a top-down propagation of states in the old-style presentation of MTTs (= becomes \rightarrow), can also be understood as a bottom-up computation of registers (= becomes :=). This remark was the subject of its own lightning talk.

References

- 1 Lê Thành Dũng Nguyễn and Cécilia Pradic. Implicit automata in typed λ-calculi I: aperiodicity in a non-commutative logic. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), volume 168 of LIPIcs, pages 135:1–135:20. Schloss Dagstuhl Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ICALP.2020.135.
- 2 Lê Thành Dũng Nguyên. Implicit automata in linear logic and categorical transducer theory. PhD thesis, Université Paris XIII (Sorbonne Paris Nord), December 2021. URL: https://nguyentito.eu/thesis.pdf.
- 3 Lê Thành Dũng Nguyễn, Camille Noûs, and Cécilia Pradic. Comparison-Free Polyregular Functions. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, 48th International Colloquium on Automata, Languages, and Programming (ICALP 2021), volume 198 of Leibniz International Proceedings in Informatics (LIPIcs), pages 139:1–139:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.139.
- 4 Joost Engelfriet and Heiko Vogler. High level tree transducers and iterated pushdown tree transducers. Acta Informatica, 26(1/2):131–192, 1988. doi:10.1007/BF02915449.

108 23202 – Regular Transformations

- 5 Paul Gallot, Aurélien Lemay, and Sylvain Salvati. Linear high-order deterministic tree transducers with regular look-ahead. In Javier Esparza and Daniel Král', editors, 45th International Symposium on Mathematical Foundations of Computer Science, MFCS 2020, August 24-28, 2020, Prague, Czech Republic, volume 170 of LIPIcs, pages 38:1–38:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.MFCS.2020.38.
- 6 Rajeev Alur and Loris D'Antoni. Streaming Tree Transducers. *Journal of the ACM*, 64(5):1–55, August 2017. doi:10.1145/3092842.
- 7 Mikołaj Bojańczyk and Amina Doumane. First-order tree-to-tree functions. In Holger Hermanns, Lijun Zhang, Naoki Kobayashi, and Dale Miller, editors, LICS '20: 35th Annual ACM/IEEE Symposium on Logic in Computer Science, Saarbrücken, Germany (online conference), July 8-11, 2020, pages 252–265. ACM, 2020. doi:10.1145/3373718.3394785.
- 8 Joost Engelfriet and Heiko Vogler. Macro tree transducers. Journal of Computer and System Sciences, 31(1):71–146, 1985. doi:10.1016/0022-0000(85)90066-2.

3.18 Learning (sub)regular transformations

Jon Rawski (San José State University, US)

License
 $\textcircled{\mbox{\scriptsize C}}$ Creative Commons BY 4.0 International license
 $\textcircled{\mbox{\scriptsize C}}$ Jon Rawski

This talk highlights foundational and recent results on learning regular transformations. Restrictions on the class of grammars, combined with various ingredients in a learning paradigm, show positive results. The talk overviews positive learning results on subsequential string functions, as well as the subclass of input and output strictly local string functions. On the empirical side, using classes of regular string transformations allows us to study the inference capabilities of black box modes such as neural networks. The talk showcases several negative results in this area.

3.19 A transducer model for streaming enumeration problems

Cristian Riveros (PUC – Santiago de Chile, CL)

License
 Creative Commons BY 4.0 International license
 Cristian Riveros

 Joint work of Cristian Riveros, Antoine Amarilli, Marco Bucchi, Alejandro Grez, Louis Jachiet, Martin Muñoz, Stijn Vansummeren

In this talk, we present Annotated automata (AnnA), a non-deterministic transducer model that outputs a subsequence of increasing positions annotated from a finite alphabet. We show that this model is suitable for encoding computational models for complex event recognition (CER) and information extraction. We show that AnnA allows for streaming enumeration algorithms, which implies the evaluation problems in the previously mentioned settings. Further, one can extend the annotation approach to visibly pushdown automata, context-free grammars, or compressed words, with similar efficient algorithmic results. Finally, we present some implementations of AnnA and the algorithms in the context of CER and information extraction.

References

- Martin Muñoz, Cristian Riveros. Streaming Enumeration on Nested Documents. ICDT 2022.
- 2 Antoine Amarilli, Louis Jachiet, Martin Muñoz, Cristian Riveros. Efficient Enumeration for Annotated Grammars. PODS 2022.
- 3 Martin Muñoz, Cristian Riveros. Constant-Delay Enumeration for SLP-Compressed Documents. ICDT 2023.
- 4 Marco Bucchi, Alejandro Grez, Andrés Quintana, Cristian Riveros. *CORE: a COmplex event Recognition Engine*. VLDB 2022.

3.20 Definability Results for Tree Transducers

Sebastian Maneth

License
Creative Commons BY 4.0 International license
Sebastian Maneth
Joint work of Sebastian Maneth, Helmut Seidl, Martin Vu

We give a summary of three definability results that were established recently. First, we discuss how to decide whether a given bottom-up tree translation can be realised by a top-down tree transducer. This is somewhat similar to deciding whether a left-to-right finite state string transducer can be realised by a right-to-left one (simply by checking the "twinning" property), but turns out to be far more complex in our tree case. This result was presented with Helmut Seidl at ICALP'2020. In fact, we cannot yet prove this result in its entirety, but we can decide whether for a given "uniform-copying" top-down tree transducer with look-ahead an equivalent transducer exists that has no look-ahead (but is uniform copying). It is well-known that every bottom-up tree transducer can be realised by a top-down tree transducer with look-ahead. Next, we show that for a given top-down tree transducer with look-ahead it is decidable whether or not an equivalent linear top-down transducer (without look-ahead) exists. This result will appear in IJFCS with Helmut Seidl and Martin Vu. Lastly, we present preliminary results concerning the conjecture (of Joost Engelfriet), that a macro tree translation can be realised by an attributed transducer (with look-around) if and only if the translation has linear increase of the number of distinct output subtrees with respect to the input tree size.

References

- Sebastian Maneth, Helmut Seidl, Martin Vu. Definability Results for Top-Down Tree Transducers. Int. J. Found. Comput. Sci. 34(2&3): 253-287 (2023).
- 2 Sebastian Maneth, Helmut Seidl. When Is a Bottom-Up Deterministic Tree Translation Top-Down Deterministic?. ICALP 2020: 134:1-134:18.

3.21 When is a Bottom-up Deterministic Tree Transducer Top-down Deterministic?

Helmut Seidl (TU München – Garching, DE)

License

 © Creative Commons BY 4.0 International license
 © Helmut Seidl

 Joint work of Sebastian Maneth, Helmut Seidl: "When Is a Bottom-Up Deterministic Tree Translation Top-Down Deterministic?", in Proc. of the 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), LIPIcs, Vol. 168, pp. 134:1–134:18, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.
 URL https://doi.org//10.4230/LIPIcs.ICALP.2020.134

We consider two natural subclasses of deterministic top-down tree-to-tree transducers, namely, linear and uniform-copying transducers. For both classes we show that it is decidable whether the translation of a transducer with look-ahead can be realized by a transducer from the same class without look-ahead.

The transducers constructed in this way, may still make use of inspection, i.e., have an additional top-down deterministic tree automaton restricting the domain. We provide a second procedure which decides whether inspection can be removed and if so, constructs an equivalent transducer without inspection.

The construction relies on a precise abstract interpretation of inspection requirements and a dedicated earliest-normal form for linear as well as uniform-copying transducers which can be constructed in polynomial time. As a consequence, equivalence of these transducers can be decided in polynomial time.

Applying these results to deterministic bottom-up transducers, we obtain that it is decidable whether or not their translations can be realized by deterministic uniform-copying top-down transducers without look-ahead (but with inspection) - or without both look-ahead and inspection.

3.22 An Algebraic Theory for Single-use Transducers Over Data Words

Rafal Stefanski (University College London, GB)

 License

 Creative Commons BY 4.0 International license
 Rafal Stefanski

 Joint work of Mikołaj Bojańczyk, Rafal Stefanski
 Main reference Mikołaj Bojańczyk, Rafal Stefanski: "Single-Use Automata and Transducers for Infinite Alphabets" in Proc. of the 47th International Colloquium on Automata, Languages, and Programming, ICALP
 Content of the 47th International Colloquium on Automata, Languages, and Programming, ICALP

2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), LIPIcs, Vol. 168,

pp. 113:1–113:14, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.

 $\textbf{URL}\ https://doi.org//10.4230/LIPIcs.ICALP.2020.113$

In a recent paper (ICALP, 2020) Bojańczyk and I have shown that single-use register automata are equivalent to orbit-finite monoids. In this talk I am going to extend this result to transducers. For this purpose, I introduce a new algebraic model called local semigroups transductions, and I show that it is equivalent to single-use register Mealy machines.

3.23 How to decide Functionality of Compositions of Top-Down Tree Transducers

Martin Vu (Universität Bremen, DE)

 License

 Creative Commons BY 4.0 International license
 Martin Vu

 Joint work of Sebastian Maneth, Helmut Seidl, Martin Vu
 Main reference Sebastian Maneth, Helmut Seidl, Martin Vu: "How to Decide Functionality of Compositions of Top-Down Tree Transducers", in Proc. of the Algebraic Informatics – 9th International Conference, CAI 2022, Virtual Event, October 27-29, 2022, Proceedings, Lecture Notes in Computer Science, Vol. 13706, pp. 175–191, Springer, 2022.
 URL https://doi.org//10.1007/978-3-031-19685-0_13

Given a nondeterministic tree transducer, it is a natural question to ask whether or not it is functional. i.e., whether its induced tree transformation is a (partial) function? ésik has shown that this problem is decidable even in the presence of look-ahead. We extend this result by showing that even for compositions of tree transducers the question of functionality is decidable. We prove this result by reducing the problem to the functionality of a single top-down tree transducer with look-ahead.

3.24 A Regular and Complete Notion of Delay for Streaming String Transducers

Sarah Winter (UL – Brussels, BE)

License

 © Creative Commons BY 4.0 International license
 © Sarah Winter

 Joint work of Emmanuel Filiot, Ismaël Jecker, Christof Löding, Sarah Winter
 Main reference Emmanuel Filiot, Ismaël Jecker, Christof Löding, Sarah Winter: "A Regular and Complete Notion of Delay for Streaming Transducers", in Proc. of the 40th International Symposium on Theoretical Aspects of Computer Science, STACS 2023, March 7-9, 2023, Hamburg, Germany, LIPIcs, Vol. 254, pp. 32:1–32:16, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.

 URL https://doi.org//10.4230/LIPIcs.STACS.2023.32

The notion of delay between finite transducers is a core element of numerous fundamental results of transducer theory. In this talk we present a new notion of delay tailored to measure the similarity between streaming string transducers (SSTs). We show that our notion is regular: we design a finite automaton that can check whether the delay between any two SSTs executions is smaller than some given bound. Moreover, we show that our notion has good completeness properties: we prove that two SSTs are equivalent if and only if they are equivalent up to some (computable) bounded delay. Together with the regularity of our delay notion, it provides an alternative proof that SST equivalence is decidable. Finally, the definition of our delay notion is machine-independent, as it only depends on the origin semantics of SSTs. As a corollary, the completeness result also holds for equivalent machine models such as deterministic two-way transducers, or MSO transducers.



4.1 Unambiguous Single-use Register Automata

Rafal Stefanski (University College London, GB)

 License

 Creative Commons BY 4.0 International license
 Rafal Stefanski

 Joint work of Mikołaj Bojańczyk, Rafal Stefanski
 Main reference Mikołaj Bojańczyk, Rafal Stefanski: "Single-Use Automata and Transducers for Infinite Alphabets", in Proc. of the 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), LIPIcs, Vol. 168, pp. 113:1–113:14, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. URL https://doi.org/10.4230/LIPIcs.ICALP.2020.113

A recent result by Bojańczyk and me (ICALP 2020) shows that deterministic one-way singleuse register automata (over data words) are equivalent to deterministic two-way single-use register automata. On the other hand, it is known that nondeterministic one-way register automata are stronger than their deterministic version. In this open question, we ask if unambiguously nondeterministic one-way single-use register automata are equivalent to the deterministic one-way single-use register automata are equivalent to the deterministic one-way single-use register automata.



Participants

 Shaull Almagor
 Technion – Haifa, IL
 Rajeev Alur
 University of Pennsylvania – Philadelphia, US
 Mikołaj Bojańczyk

University of Warsaw, PL

Elisabet Burjons York University – Toronto, CA

Michaël Cadilhac DePaul University – Chicago, US

Olivier Carton
 Université Paris Cité, FR

Ryan Cotterell ETH Zürich, CH

Luc Dartois
 University Paris-Est –
 Créteil, FR

■ Gaëtan Douéneau-Tabot Université Paris Cité, FR

Léo Exibard Gustave Eiffel University -Marne-la-Vallée, FR Emmanuel Filiot $\rm UL$ – Brussels, BE Paul Gallot Universität Bremen, DE Matthew Hague Royal Holloway, University of London, GB Jeffrey Heinz Stony Brook University, US Ismaël Jecker University of Warsaw, PL Sandra Kiefer University of Oxford, GB Nathan Lhote Aix-Marseille University, FR Sebastian Maneth Universität Bremen, DE

Anca Muscholl University of Bordeaux, FR Le Thanh Düng Nguyen ENS – Lyon, FR Cécilia Pradic Swansea University, GB Gabriele Puppis University of Udine, IT Jon Rawski San José State University, US Cristian Riveros PUC – Santiago de Chile, CL Helmut Seidl TU München – Garching, DE Rafal Stefanski University College London, GB Martin Vu Universität Bremen, DE Sarah Winter $\rm UL$ – Brussels, BE

