Final report on the project Extensions of the Theory of Automata and Languages supported by the NKFI grant nr. 108448

Formal power series. The papers [1, 2, 3] were submitted for publication during the time frame between the submission of the project proposal and the start date of the project. However, the revision of the papers took place after the project started. Support from the project has been acknowledged.
In [4] it is observed that the axioms of iteration semirings capture many important identities of regular languages and rational power series. The authors gave a description of the free iteration semirings using rational power series.
In [5] power series over a graded monoid M of finite type are considered. It is shown first that, under certain conditions, the equivalence problem of power series over M with coefficients in the semiring $\mathbb{N}$ of nonnegative integers can be reduced to the equivalence problem of power series over $\{x\}^*$ with coefficients in $\mathbb{N}$. This result is then applied to rational and recognizable power series over M with coefficients in $\mathbb{N}$, and to rational power series over Σ^* with coefficients in the semiring $\mathbb{Q}_+$ of nonnegative rational numbers, where Σ is an alphabet.

Kleene algebras for energy problems. In [6] the authors define and study basic properties of $*$-continuous Kleene ω-algebras that involve a $*$-continuous Kleene algebra with a $*$-continuous action on a semimodule and an infinite product operation that is also $*$-continuous. They show that $*$-continuous Kleene ω-algebras give rise to iteration semiring-semimodule pairs, and they show how their work can be applied to solve certain energy problems for hybrid systems. These results are used in [7] that energy problems concerning both the finite and infinite behaviour of energy automata can be solved by algebraic manipulations on the transition matrices.

Semigropus associated to finite state automata. In [9] a new proof is given for the known result of Brzozowski et al that a reduced n-state automaton M recognizes a maximally atomic language if and only if its transition semigroup $T(M)$ contains a transformation of rank $n-1$, moreover, the permutation subgroup $P(M)$ of $T(M)$ is set-transitive. The approach of [9] allows for a significantly more compact presentation than the previous attempts.
In [10] an upper bound of $n((n-1)! - (n-3)!)$ is given for the possible largest size of a subsemigroup of the full transformational semigroup over n elements consisting only of nonpermutational transformations. As an application the same upper bound is gained for the syntactic complexity of definite languages as well. Moreover, the same upper bound is proved for the syntactic complexity of generalized definite languages.

Generalizations of context-free languages. In [11] Muller context free languages are studied, which are a generalization of context-free languages to infinite (countable) words where we allow infinite derivation trees with a Muller-type acceptance condition. Generalizing our former results, we showed that it is decidable for an MSO-definable property and a Muller context-free language (given by a grammar) whether every word in the language satisfies the given property. This implies, for example, that it is decidable whether every word in the language is scattered or dense,

etc.
In [16] the authors gave a survey on Muller context-free languages and shown an elementary and original characterization of MCFL in terms of regular expression-like operations.
In [23] we showed that the following five classes of weighted languages are the same: (i) the class of proper weighted context-free languages, (ii) the class of weighted languages recognized by proper weighted tree automata, (iii) the class of weighted languages recognized by deterministic and proper top-down weighted tree automata, (iv) the class of weighted languages recognized by deterministic and proper bottom-up weighted tree automata, and (v) the class of weighted languages obtained by taking the yield of proper local weighted tree languages.
In [26] we combine three extensions of context-free grammars: (a) associating its nonterminals with storage configurations, (b) equipping its rules with weights, and (c) controlling its derivations. For a commutative semiring K, we introduce the class of weighted languages generated by K-weighted linear context-free grammars with storage S and with derivations controlled by (S, K)-recognizable weighted languages. The control on the derivations can be iterated in a natural way. We characterize the n-th iteration of the control in terms of the n-th iteration of the one-turn pushdown operator on the storage S of the control weighted language. Moreover, for each proper semiring we prove that iterating the control yields an infinite, strict hierarchy of classes of weighted languages.

Tree transducers and weighted tree transducers. Multi bottom-up tree transducers (MBOT) are a formal model of syntax directed translation based on tree automata. They are used in machine translation of natural languages. In [12], we proved linking theorems for MBOT. The linking theorems can be used to establish that certain translations can not be performed by MBOT.
Some of these results are used in [15], where we consider the closure under composition of linear extended top-down tree transducers (or synchronous tree-substitution grammars). We prove that the composition hierarchy of linear and ε-free (all rules consume input) extended top-down tree transducers with regular look-ahead collapses at power 3. The composition hierarchy of the same kind of tree transducers without regular look-ahead collapses at power 4. The composition hierarchy of linear, simple (all rules produce output), and epsilon-free extended top-down tree transducers with and without regular look-ahead collapses at power 2. The numbers 3, 4, and 2 are proved to be the least such numbers. We prove that the composition hierarchy of linear, nondeleting, and ε-free extended top-down tree transducer is infinite. The same holds for linear extended top-down tree transducers and for linear extended top-down tree transducers with regular look-ahead.
In [27] we generalize the results of [15] to the weighted case. We present linear weighted extended top-down tree transducers in the framework of synchronous grammars. We prove that the composition hierarchy of linear and epsilon-free weighted extended top-down tree transducers with regular look-ahead collapses at power 3. The composition hierarchy of the same kind of tree transducers without regular look-ahead collapses at power 4. The composition hierarchy of linear, simple, and ε-free weighted extended top-down tree transducers with and without regular look-ahead collapses at power 2.

Tree automata and weighted tree automata. In [8], generalizing the notion of local tree languages, we introduced local weighted tree languages over a semiring. We

proved that a weighted tree language over an arbitrary semiring is recognizable by a finite weighted tree automaton if and only if it can be obtained as the image of a local weighted tree language under a deterministic relabeling.
In [13] we proved a Kleene theorem for weighted tree automata over tree valuation monoids (tv-wta). In order to be able to define the semantics of a tv-wta inductively we enriched the tree valuation monoid by a family of decomposition operations. This enriched structure is called a Cauchy tree valuation monoid. By these decomposition operations we defined the rational operations concatenation and Kleene-star. We used variables as additional labels for leaves of trees and assumed that the tree valuation monoid has a unit element. We represented rational trees series by rational expressions. We showed that weighted rational expressions and wta over Cauchy unital tree valuation monoids are expressively equivalent. The journal version of [13] contains the full proofs of the theorems and more natural examples. A further novelty is that we gave a new, simpler definition of a unital tree valuation monoid (the weight structure). With this our model became more general and, among others, it generalizes wta over semirings. Hence our main result provides an alternative Kleene-type result for wta over semirings.
In [25] we give a general definition of weighted tree automata and define three instances which differ in the underlying weight algebras: semirings, multi-operator monoids, and tree-valuation monoids. We also define a general concept of weighted expressions based on monadic second-order logic. In the same way as for weighted tree automata, we define three instances corresponding to the above mentioned weight algebras. We prove that weighted tree automata over semirings are equivalent to weighted expressions over semirings, and establish the equivalence between weighted tree automata and weighted expressions over tree-valuation monoids. For weighted tree automata over semirings and over tree-valuation monoids we prove characterizations in terms of bimorphisms.
In [24] we introduce weighted regular tree grammars with storage. We prove that, for multioperator monoids canonically associated to particular strong bimonoids, the support of the generated weighted tree languages can be generated by (unweighted) regular tree grammars with storage. We characterize the class of all generated weighted tree languages by the composition of three basic concepts. Moreover, we prove results on the elimination of chain rules and of finite storage types, and we characterize weighted regular tree grammars with storage by a new weighted MSO-logic.
In [18] we consider offline sensing top-down tree automata in which the state transition is computed by bimachines. We give a polynomial time algorithm for minimizing such tree automata when they are state-separated.

Theory of classical formal languages. Permitting semi-conditional grammars (pSCGs) are extensions of context-free (cf) grammars. Since 1984 it was an open question whether they can generate every context-sensitive language. In [19] we showed that these grammars without erasing rules are strictly weaker than context-sensitive grammars. To show this we used standard pumping arguments based on the properties of the m-embedding relation (which is a generalization of the scattered subword relation) defined in [19]. In [20] we considered a generalization of pSCGs (called gpSCGs to be short) where each rule r is associated with a set of words P, and r is applicable only if every word in P occurs in the current sentential form. Generalizing the ideas in [19] we showed that gpSCGs are still strictly weaker than context-sensitive grammars. Forbidding random context grammars (fRCGs) are extensions of cf grammars in which each rule r is associated with a set of nonterminals P, and

r is applicable only if no nonterminal in P occurs in the current sentential form. In [20] we could compare the generative capacity of gpSCGs and that of fRCGs as follows. We showed that for every positive integer m, there is a language L_m such that L_m can be generated by an fRCG but cannot by any gpSCG having permitting words with length at most m.

Term rewriting systems. In [31] it is shown that for every recognizable tree language L, finite set Z of regular path languages, and Z-prefix constrained linear monadic term rewriting system S, the position cutting descendant of L for S is recognizable as well. For every recognizable tree language L and finite set Z of regular path languages, we construct finitely many Z-prefix constrained linear monadic term rewriting systems generating the position cutting descendants of L for all Z-prefix constrained linear monadic term rewriting systems.
A binary relation on a set of terms is called a rewrite relation if it is closed both under context application (the“replacemen” or “monotonicity” property) and under substitutions (the “fully invariant property”). The intersection of the reflexive transitive closures of two rewrite relations induced by term rewriting system is also a rewrite relation. In [32] it is shown to be undecidable whether the intersection of the reflexive transitive closures of two rewrite relations induced by term rewriting systems is equal to the reflexive transitive closure of a rewrite relation induced by a term rewriting system.

Reversible regular languages. In [17] we characterize the class of reversible regular languages which have only finitely many minimal reversible automata by means of a forbidden pattern for the minimal automaton of the langauge in question. Based on these results, we showed in [30] that it is NL-complete to decide whether an automaton recognizes a reversible language having only finitely many reduced reversible automata. Also, we give an explicit construction for a family of reversible regular languages having only finitely many reduced reversible automata such that the largest such reduced reversible automaton is exponentially larger (more precisely, the blow-up rate is $\Omega(1.61^n)$) than the minimal automaton of the language in question. Both results hold already for the case of binary alphabets.

Miscellaneous related works. A rectifier network is a Boolean network computing a mapping of binary words of length n to binary words of length m for some integers n and m using only disjunction gates. The smallest number of edges of a rectifier network computing a function (given by a Boolean matrix) is a classic complexity measure of matrices. In [14] the authors obtained new (both lower and upper) bounds of these measures for two concrete cases using linear programming methods. First, they improved the upper bound for the depth-2 complexity of the disjointness matrix, contributing towards the solution of the open problem 7.3 of the recent book of Jukna and Sergeev. Second, they confirmed a conjecture of Find et al. from 2013 up to a logarithmic factor, on the lower bound for the complexity of tensor products.

In the works [21, 22, 29] we showed that the recognition problem of Union-Find trees is NP-complete also in the case when one uses either the so-called union-by-rank or the union-by-size building strategy. The former has two subcases: (a) the rank info is also provided, and (b) the input is only the untagged structure of the tree. The three papers cover all the two problems being NP-complete.
In [28] we investigate dinamyc graph algorithms and prove the following two main results. First, depth-first search has no sparse strong certificates, thus its dynamic variant cannot be sped up by the sparsification technique. Second, we gave an algo-

rithm that maintains a maximal matching of a dynamic graph in $O(\log n)$ amortized update time, using a lookahead of $O(m)$.

Szeged, 18. 01. 2019.

Zoltán Fülöp

References

[1] M. Droste, Z. Ésik, and W. Kuich, Conway and iteration hemirings Part 1, Int. J. Algebra and Computation, 24(4) pp. 461-482, 2014.

[2] M. Droste, Z. Ésik, and W. Kuich, Conway and iteration hemirings Part 2, Int. J. Algebra and Computation, 24(4) pp. 483-513, 2014.

[3] Z. Ésik, S. Iván, Operational characterization of scattered MCFLs, Int. J. Foundations of Computer Science 25(8) pp. 1001-1016, 2014.

[4] Z. Ésik, T. Hajgató, On the structure of free iteration semirings, J. Automata, Languages and Combinatorics, Vol. 19 pp. 57-66, 2014.

[5] Z. Ésik, W. Kuich, On Power Series over a Graded Monoid, in: Computing with New Resources, LNCS, Vol. 8808, pp. 49-55, Springer, 2014. 49–55.

[6] Z. Ésik, U. Fahrenberg, and A. Legay, $*$-Continuous Kleene ω-Algebras, in: Proc. of DLT 2015, LNCS, Vol. 9168, pp. 240-250, Springer, 2015.

[7] Z. Ésik, U. Fahrenberg, and A. Legay, $*$-Continuous Kleene ω-Algebras for Energy problems, in: Proc. of Fixed Points in Computer Science, Berlin, Electronic Proceedings in Theoretical Computer Science, Vol. 191, pp. 48-59, 2015.

[8] Z. Fülöp, Local Weighted Tree Languages, Acta Cybernetica, 22 pp. 393-402, 2015.

[9] S. Iván, Complexity of atoms, combinatorially, Information Processing Letters, 116(5) pp. 356-360, 2016.

[10] S. Iván, J. Nagy-György, On nonpermutational transformation semigropus with an application to syntactic complexity, Acta Cybernetica, 22 pp. 687-701, 2016.

[11] Z. Ésik, S. Iván, MSO-definable Properties of Muller Context-Free Languages are Decidable, in Proc. of DCFS 2016, LNCS, Vol. 9777, pp. 87-97, Springer, 2016.

[12] Z. Fülöp, and A. Maletti, Linking theorems for tree transducers. Journal of Computer and System Sciences 82 pp. 1201-1222, 2016.

[13] Z. Fülöp, M. Droste, and D. Götze, A Kleene Theorem for Weighted Tree Automata over Tree Valuation Monoids. In: Proc. of LATA 2016, Lecture Notes in Computer Science, Vol. 9618, pp. 452-463, Springer, 2016. Journal version will appear in Information and Computation.

[14] D. Chistikov, S. Iván, A. Lubiw, and J. Shallit: Fractional Coverings, Greedy Coverings, and Rectifier Networks. STACS 2017: 23:1-23:14. http://drops.dagstuhl.de/opus/volltexte/2017/7010/pdf/LIPIcs-STACS-2017-23.pdf

[15] J. Engelfriet, Z. Fülöp, and A. Maletti, Composition Closure of Linear Extended Top-down Tree Transducers, Theory of Computing Systems, 60 pp. 129-171, 2017.

[16] K. Gelle, S. Iván, Regular Expressions for Muller Context-Free Languages. Acta Cybernetica, 23(1) pp. 349-369, 2017.

[17] K. Gelle, S. Iván, Reversible languages having finitely many reduced automata, in Proc. of AFL 2017, Electronic Proceedings in Theoretical Computer Science, Vol. 257, pp. 114-127, 2017.

[18] Z. Fülöp, S. Vágvölgyi, Minimization of Deterministic Top-down Tree Automata, Acta Cybernetica, 23 pp. 379-401, 2017.

[19] Z. Gazdag, K. Tichler, On the power of permitting semiconditional grammars, in Proc. of DLT 2017, Lecture Notes in Computer Science, Vol. 10396, pp. 173-184, Springer, 2017.

[20] Z. Gazdag, K. Tichler, and E. Csuhaj-Varjú, A Pumping Lemma for Permitting Semi-conditional Languages. To appear in International Journal of Foundations of Computer Science.

[21] K. Gelle, S. Iván, Recognizing Union-Find Trees Built Up Using Union-by-rank Strategy is NP-Complete. in Proc. of DCFS 2017, LNCS, Vol. 10316, pp. 152-163, Springer, 2017.

[22] K. Gelle, S. Iván, Recognizing Union-Find Trees is NP-Complete. Information Processing Letters, 131 pp. 7-17, 2018.

[23] Z. Fülöp, Z. Gazdag, Weighted Languages Recognizable by Weighted Tree Automata, Acta Cybernetica, 23(3) pp 867-886, 2018.

[24] Z. Fülöp, L. Herrmann, H. Vogler, Weighted Regular Tree Grammars with Storage, Discrete Mathematics & Theoretical Computer Science 21(1) pp. 1-44, 2018.

[25] Z. Fülöp, H. Vogler, Characterizations of Recognizable Weighted Tree Languages by Logic and Bimorphisms, Soft Computing, 22 pp. 1035-1046, 2018.

[26] Z. Fülöp, H. Vogler, Weighted Iterated Linear Control, Acta Informatica, to appear.

[27] Z. Fülöp, and A. Maletti, Composition Closure of Linear Weighted Extended Top-down Tree Transducers, submitted.

[28] K. Gelle, S. Iván, DFS is Unsparsable and Lookahead Can Help in Maximal Matching. Acta Cybernetica 23(3), p. 887-902, 2018.

[29] K. Gelle, S. Iván, Recognizing Union-Find Trees is NP-complete, even without Rank Info. To appear in International Journal on Foundations of Computer Science.

[30] K. Gelle, S. Iván, More on reversible languages having finitely many reduced reversible automata. To appear in International Journal on Foundations of Computer Science.

[31] S. Vágvölgyi, Descendants of a recognizable tree language for prefix constrained linear monadic term rewriting with position cutting strategy, Theoretical Computer Science, 732 pp. 60-72, 2018.

[32] S. Vágvölgyi, Intersection of the reflexive transitive closures of two rewrite relations induced by term rewriting systems, Information Processing Letters, 134 pp. 47-51, 2018.