Algebraic Theory of Automata and Logic AUTOMATHA project, European Science Foundation

September 30 – October 1, 2006

Report

Z. Ésik, H. Straubing, P. Weil, Th. Wilke

1 Summary

The aim of the workshop was to provide a forum for researchers interested in the structure theory of automata and/or the application of the algebraic approach to logic to present their new results and to combine their efforts in the further development of the structure theory of finite automata, tree automata and related structures in connection with formal logic. Several schools and research groups were represented by leading experts including research groups from France, Germany, Portugal, Finland, Czech Republic, Poland, USA, and Hungary.

Altogether 20 talks were presented at the workshop (3 presentations of 40 minutes – J.-E. Pin, H. Straubing and W. Thomas – and 17 presentations of 25 minutes). In addition, there was an open problem session summarizing the open problems raised in the presentations.

2 Description of the Scientific Contents

Varieties of finite semigroups, monoids, and more recently C-varieties and varieties of finite automata have been widely used in the study of logical and combinatorial properties of words and the expressive power of logical systems on words. Using the Eilenberg correspondence, one can reduce the definability of a regular language to the decidability of the corresponding variety of finite semigroups or automata. The talks by Jean-Éric Pin, Laura Chaubard, Jorge Almeida, Michal Kunc and Manfred Kufleitner were all related to this topic. Pedro Silva showed how Turing machines can be described algebraically using block products and projective limits of bimachines. The aim of this research is to obtain an algebraic formulation of the P = NP problem and to develop algebraic tools for attacking this problem.

Algebraic machinery developed for dealing with properties of trees and the logical definability of tree languages and the algebraic characterization of the expressive power of logics on finite trees have been the subject matter of the talks by H. Straubing, M. Bojańczyk, S. Salehi, M. Steinby, Z. Ésik and Sz. Iván. H. Straubing and M. Bojańczyk talked about the emerging theory of "tree algebras" and some applications of this theory for obtaining decidable characterizations of certain tree language classes defined by logical and/or combinatorial properties. In contrast, the talks by S. Salehi, M. Steinby, Z. Ésik and Sz. Iván used the conventional theory of tree automata to provide general and specific results regarding tree language varieties.

Pebble automata and their connections to logic and XML formed the third major topic represented by the talks of T. Schwentick, B. ten Cate and Björklund. The three talks provided a very good overview of recent research in pebble automata related to logic, including several recent deep theorems.

Feferman-Vaught theorems provide a way of characterizing the first-order theory of a structure obtained by some algebraic operation from some component structures. The studied algebraic operations include, in addition to the classical sum and product constructions, the synchronized product of labeled transition systems. In the talk of Wolfgang Thomas, it was shown how the Feferman-Vaught theory can be used to derive decidability of the first-order theory of structures constructed using synchronized products, tree iterations and sums.

Olivier Carton compared four families of relations realized by multi-tape automata: recognizable, synchronous, deterministic, and rational relations.

The emerging field of weighted logics was represented by the talks of Manfred Droste and Ingmar Meinecke. Droste showed that weighted Muller automata, weighted Büchi automata and the sentences from a suitable restricted class monadic second-order weighted logic with discounting have the same expressive power. Meinecke gave a logical characterization of recognizable trace series.

The talk by Jelena Ignjatović developed a Myhill-Nerode theory for fuzzy languages and automata.

3 Assessment of the results and impact of the event on the future direction of the field

The workshop has very well served its purpose and aim by providing both an up to date account of a variety of recent work in algebra and the theory of automata in connection with formal logic, and a forum for promoting collaboration and the exchange of ideas among the participants.

Algebraic techniques based on the notions of syntactic semigroup and (pseudo)varieties of finite semigroups have been very successfully applied in the classical theory of automata and languages on finite and infinite words. The theory has recently been refined in order to be able to capture a wider class of combinatorial and algebraic properties of words and a wider class of logics on words (cf. C-varieties, varieties of automata, literal varieties, etc.). The workshop provided an excellent opportunity to look at these new developments in addition to the classical framework.

The development of algebraic tools for the study of combinatorial and logical properties of trees is current ongoing work by various researchers. The various approaches include semigroups, Wilke tree algebras, clones and preclones, algebraic theories, (pseudo)varieties of finite algebras and the recent "forest algebras" (also called tree algebra). Probably the most valuable contribution of the workshop was that it brought together the representatives of several schools working on the development of the above tools. The algebraic approach to trees was very nicely complemented by that recent work on pebble automata in relation to formal logic presented at the workshop. The workshop will certainly have a strong impact on the future research in this area.

Talks in the above areas were very nicely complemented by presentations giving an inside view of ongoing work in weighted logics and fuzzy automata, rational relations, and the Feferman-Vaught theory.

There were a number of open problems raised during the talks which were collected by

the organizers. These open problems and the slides of the presentations are available from the workshop web site.

4 Final Scientific Programme

Saturday, September 30

8:55 Opening

Session 1

9:00 Jean-Éric Pin (Paris): The separation conjecture

9:40 Jorge Almeida (Porto): The profinite approach to decidability questions

Session 2

10:35 Michal Kunc (Brno): Equational description of pseudovarieties of homomorphisms

11:00 Laura Chaubard (Paris): C-varieties and first-order logic with modular predicates

11:25 Manfred Kufleitner (Stuttgart): Trace languages and the variety DA

11:50 Pedro Silva (Porto): Bimachines: an algebraic approach to Turing machine computation

Session 3

14:00 Howard Straubing (Boston): An algebraic approach to regular languages of unranked trees

14:40 Saeed Salehi (Turku): Characterizing Families of Tree Languages by Syntactic Monoids

15:05 Magnus Steinby (Turku): Tree Algebras and Varieties of Tree Languages

Session 4

16:00 Mikolaj Bojańczyk (Warsaw): Forest algebra

16:25 Zoltán Ésik (Szeged): Products of tree automata with an application to temporal logic

16:50 Szabolcs Iván (Szeged): Some varieties of finite tree automata related to restricted temporal logics

Sunday, October 1

Session 5

9:00 Wolfgang Thomas (Aachen): Product structures and logic: News on the Feferman-Vaught approach

9:40 Olivier Carton (Paris): Decidability of some classes of rational relations

Session 6

10:35 Thomas Schwentick (Marburg): Expressive power of pebble automata

11:00 Balder ten Cate (Amsterdam): Regular XPath: Algebra, Logic and Automata

11:25 Henrik Björklund (Dortmund): Towards Regular Data Languages

Session 7

14:00 Open problems

15:00 Break

Session 8

15:35 J. Ignjatović (Nis): Myhill-Nerode theory for fuzzy languages and automata

16:00 Manfred Droste (Leipzig): Weighted automata and weighted logics with discounting

16:25 Ingmar Meinecke (Leipzig): Weighted traces, their logics, and an extension to weighted MSCs

5 List of participants

1. Aehlig, Klaus; Swansea, United Kingdom
2. Almeida, Jorge; Porto, Portugal
3. Björklund, Henrik; Dortmund, Germany
4. Bojańczyk, Mikolaj; Warsaw, Poland
5. Carton, Olivier; Paris, France
6. Chaubard, Laura; Paris, France
7. Droste, Manfred; Leipzig, Germany
8. Ésik, Zoltán; Szeged, Hungary
9. Fülöp, Zoltán; Szeged, Hungary
10. Gécseg, Ferenc; Szeged, Hungary
11. Hella, Lauri; Tampere, Finland
12. Ignjatović, Jelena; Niš, Serbia
13. Iván, Szabolcs; Szeged, Hungary
14. Kaiser, Lukasz; Aachen, Germany
15. Kufleitner, Manfred; Stuttgart, Germany
16. Klíma, Ondřej; Brno, Czech Republic
17. Kuich, Werner; Wien, Austria
18. Kunc, Michal; Brno, Czech Republic
19. Maryns, Hendrik; Tübingen, Germany
20. Meinecke, Ingmar; Leipzig, Germany
21. Muzamel, Lóránd; Szeged, Hungary
22. Németh, Zoltán; Szeged, Hungary
23. Pin, Jean-Éric; Paris, France
24. Polák, Libor; Brno, Czech Republic
25. Ramanujam, Ramaswamy; Chennai, India
26. Salehi, Saeed; Turku, Finland
27. Schwentick, Thomas; Dortmund, Germany
28. Ségoufin, Luc; Paris, France
29. Silva, Pedro; Porto, Portugal
30. Steinby, Magnus; Turku, Finland
31. Straubing, Howard; Boston, USA
32. ten Cate, Balder; Amsterdam, Netherlands
33. Tesson, Pascal; Quebec, Canada
34. Thomas, Wolfgang; Aachen, Germany
35. Vágvölgyi, Sándor; Szeged, Hungary
36. Weil, Pascal; Bordeaux, France
37. Wilke, Thomas, Kiel, Germany

Except for one participant, all participants supported from the ESF fund came from countries contributing to the project.

6 Short abstracts

1. Jorge Almeida: The profinite approach to decidability questions.

Several questions about regular languages and their counterparts for finite semigroups admit a formulation in terms of relatively profinite semigroups. In this context, some combinatorial difficulties are in a sense untangled as more structure comes into play, and properties become more transparent. The talk presented a quick survey of how profinite methods have contributed to settle the decidability of various problems.

2. Henrik Björklund and Thomas Schwentick: Towards Regular Data Languages.

The authors compared three different classes of automata: register automata, data automata, and deterministic automata derived from an alternative description of data automata.

They showed that in terms of expressivity, 1-way nondeterministic register automata are, somewhat surprisingly, strictly weaker than data automata, rather than incomparable. For data automata, they showed that a couple of simplifications and additions can be made without changing the expressive power, and that the membership problem is NP-complete. The deterministic automaton model they introduced was shown to have some nice closure properties, and they investigated it's expressive power.

Finally, they discussed some possible extensions of the automata models, and investigated decidability of the emptiness problem for these extensions.

3. Mikołaj Bojańczyk: Forest algebra.

Forest algebra is an algebraic formalism for describing languages of finite unranked trees. The hope is that forest algebra will be a convenient tool to solve the decision problem: "can a given regular tree language be defined in first-order logic?".

The author presented some evidence justifying this hope, including simplified characterizations of the temporal logics EF and EF + EX.

4. Olivier Carton: Decidability of some classes of rational relations.

This talk focused on rational relations which are the relations realized by multi-tape automata. The four families of recognizable, synchronous, deterministic, and rational relations were considered. They form a strict hierarchy. The problem of whether a given relation of a family belongs to a smaller family was studied. Known results were recalled and some new result were also given.

5. Laura Chaubard: C-varieties and first-order logic with modular predicates.

The theory of C-varieties extends the classical setting of Eilenberg's varieties. This generalization now finds new fields of applications, including logic, where it helps deciding some levels of the Σ_n hierarchy that counts the number of quantifiers alternations. The first part of the talk consisted of a general introduction of C-varieties and their motivations. Then, the author turned to first-order logic with modular predicates and stated two new results regarding the decidability for the first levels of the Σ_n hierarchy.

6. Manfred Droste: Weighted automata and weighted logics with discounting.

The author presented two weighted automata models (Büchi and Muller) with discounting. The weights are non-negative real numbers. The infinitary behavior of such automata was described by assigning to each infinite word a weight, the max-

imal discounted weight of the paths realizing the given word. The author showed that weighted Muller automata, weighted Büchi automata and the sentences from a suitable weighted logic with discounting describe exactly the same class of functions.

7. Zoltán Ésik and Szabolcs Iván: Products of tree automata with an application to temporal logic.

In an earlier paper by the first author, an (extended) temporal logic FTL($\mathcal{L}$) was associated with each class $\mathcal{L}$ of regular tree languages. It was shown under a mild condition that for logics in which the next modalities are expressible there is a bijective correspondence between the language classes definable in the logics FTL($\mathcal{L}$) and those (pseudo)-varieties of finite tree automata containing the definite tree automata which are closed under the cascade product. In this talk authors removed from this result the assumption that the next modalities are expressible. This was achieved by replacing the cascade product with the Moore-product introduced in the talk.

8. Zoltán Ésik and Szabolcs Iván: Some varieties of finite tree automata related to restricted temporal logics.

The authors provided simple polynomial time decidable characterizations of several varieties of finite tree automata related to certain fragments of CTL. Some of the results provided an improvement on earlier results by Bojanczyk and Walukiewicz.

9. Jelena Ignjatović and Miroslav Ćirić: Myhill-Nerode Theory for Fuzzy Languages and Automata.

The authors developed a Myhill-Nerode theory for fuzzy languages and automata studying recognizability of fuzzy languages through fuzzy right congruences and fuzzy congruences on the free monoid X^*.

10. Manfred Kufleitner: Trace languages and the variety DA.

Traces were introduced by Mazurkiewicz in 1977 as a generalization of words to describe the behavior of concurrent processes. A lot of characterizations for the class of recognizable trace languages and for the the class of star-free trace languages are known. A result from the early 90s is that a trace-language is star-free if and only if its syntactic monoid is aperiodic. An important subclass of aperiodic monoids is the variety DA. Over words, many characterizations have been found for languages whose syntactic monoid is in DA. The authors showed which of them also hold for traces and which of them do not.

11. Michal Kunc: Equational description of pseudovarieties of homomorphisms.

The notion of pseudovarieties of homomorphisms onto finite monoids was introduced by Howard Straubing as an algebraic characterization for certain classes of regular languages more general than varieties. The author described an appropriate generalization of implicit operations which can be used to define these pseudovarieties by means of equations.

12. Ingmar Meinecke: Weighted Traces, Their Logics, and an Extension to Weighted MSCs.

The author gave a logical characterization of recognizable trace series. By introducing a weighted monadic second order logic, the author showed that any recognizable trace series can be defined by some restricted weighted MSO-formula and, vice versa, every definable trace series is recognizable.

The author was able to transfer results for formal power series over words to trace series using a translation technique between formulas for words and those for traces. Moreover, the author obtained a general unambiguity result for formulas over rela-

tional structures.

13. Jean-Éric Pin: The separation conjecture.

A language is Σ_n-definable [Π_n-definable] if it can be defined by a Σ_n-sentence [a Π_n-sentence]. A language is Δ_n-definable if it is both Σ_n-definable and Π_n-definable. A language K *separates* two disjoint languages L_1 and L_2 if either $L_1 \subseteq K \subseteq A^* \setminus L_2$ or $L_2 \subseteq K \subseteq A^* \setminus L_1$.

The separation conjecture states that two disjoint Σ_n-definable languages can be separated by a Δ_n-definable language. The author presented an algebraic extension of the separation conjecture and discussed some partial results.

14. John Rhodes and Pedro V. Silva: Bimachines: an Algebraic Approach to Turing Machine Computation.

Turing machines constitute the standard model of computation, upon which is built computational complexity theory. By studying the concept of bimachine, a structure combining actions of semigroups on two automata with a single output function, the authors developed a new algebraic model for Turing machine computation that will hopefully open new perspectives into the study of computational complexity.

15. Saeed Salehi: Characterizing Families of Tree Languages by Syntactic Monoids.

When trees are defined as terms over a ranked alphabet, syntactic algebras become the most natural structures for characterizing families of tree languages. However, several other syntactic structures have been introduced by some authors, and in most cases an a la Eilenberg variety theorem has been provided for them. The author sketched a variety theorem for syntactic semigroups/monoids. The author also showed that definite tree languages are not definable by syntactic monoids (neither by semigroups), which refutes a result of M. Nivat & A. Podelski for the first time after its publication in 1989.

16. Thomas Schwentick: Expressive power of pebble automata.

Two variants of pebble tree-walking automata on trees were considered that had been introduced in the literature. It was shown that for each number of pebbles, the two models have the same expressive power both in the deterministic case and in the nondeterministic case. Furthermore, nondeterministic (resp. deterministic) tree-walking automata with $n+1$ pebbles can recognize more languages than those with n pebbles. Moreover, there is a regular tree language that is not recognized by any tree-walking automaton with pebbles. As a consequence, $FO + posTC$ is strictly included in MSO over trees.

(Joint work with Mikołaj Bojańczyk, Mathias Samuelides, and Luc Ségoufin.)

17. Magnus Steinby: Tree Algebras and Varieties of Tree Languages.

The author considered several aspects of the tree algebra formalism of Thomas Wilke.

Wilke's axioms for tree algebras were turned into a convergent term rewriting system for which we can describe the normal form representations of trees and contexts by Γ-terms. Hence, a simple computational method for deciding the equivalence of two Γ-terms was obtained.

Varieties of binary tree languages (VBTLs) and varieties of finite tree algebras (VFTAs) were defined as expected, but the natural maps between them via syntactic tree algebras do not yield the anticipated bijection. However, by replacing VFTAs with varieties of finite *reduced tree algebras*, the author obtained a Variety Theorem for VBTLs.

(Joint work with S. Salehi.)

18. Howard Straubing: An algebraic approach to regular languages of unranked trees.

The author provided an outline of an algebraic approach to deal with unranked trees based on the notion of "tree prealgebra" or "tree algebra" consisting of a vertical semigroup V and a horizontal semigroup H which is equipped with an action of V on H. A tree algebra also has actions of H on V. Tree (pre)algebras were first defined by M. Bojanczyk and I. Walukiewicz in an unpublished paper in order to provide algebraic characterizations of the expressive power of logics on unranked trees.

(Joint work with M. Bojanczyk and I. Walukiewicz.)

19. Balder ten Cate: Regular XPath: Algebra, Logic and Automata.

The author discussed the XML path language *Regular XPath* and its connections to transitive closure logics and tree walking automata. Regular XPath is a language for describing binary relations in XML documents. It was shown that positive Regular XPath has the same expressive power (in terms of definable binary relations) as tree walking automata with return pebbles. This result was compared to the automata theoretic characterization of positive Regular XPath$^{\approx}$ by Goris and Marx, and to the characterization of $FO + posTC^1_{np}$ by Engelfriet and Hoogeboom.

20. Wofgang Thomas: Product structures and logic: News on the Feferman-Vaught approach

The talk provided an introduction to the Feferman-Vaught theory and discussed some of its applications in Computer Science.

Feferman-Vaught theorems provide a way of characterizing the first-order theory of a structure obtained by some algebraic operation from some component structures. In computer science, the algebraic operations include, in addition to the classical sum and product constructions, the synchronized product of labeled transition systems. It was shown how the Feferman-Vaught theory can be used to derive decidability of the first-order theory of structures constructed using the operations of (synchronized) products, tree iterations and sums.