DI Microconference AAMP VIII:

Words, Numeration and Automata

Vila Lanna, March 19–20, 2011

List of participants:

Petr Ambrož (Dept. of Math., FNSPE, CTU in Prague)
Ľubomíra Balková (Dept. of Math., FNSPE, CTU in Prague)
Karel Břinda (Dept. of Math., FNSPE, CTU in Prague)
Jozef Bobok (Faculty of Civil Engineering, Czech Technical University in Prague)
Daniel Dombek (Dept. of Math., FNSPE, CTU in Prague)
Tomáš Flouri (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Ondřej Guth (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Tomáš Hejda (Dept. of Math., FNSPE, CTU in Prague)
Jan Holub (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Štěpán Holub (Dept. of Algebra, Charles University in Prague)
Sirui Cheng (Dept. of Math., FNSPE, CTU in Prague)
Jan Janoušek (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Jakub Jaroš (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Karel Klouda (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Jakub Kolář (Dept. of Math., FNSPE, CTU in Prague)
Petr Kůrka (Center for theoretical study)
Zuzana Masáková (Dept. of Math., FNSPE, CTU in Prague)
Edita Pelantová (Dept. of Math., FNSPE, CTU in Prague)
Martin Plicka (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Martin Poliak (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Petr Procházka (Dept. of Theoretical Computer Science, FIT, CTU in Prague)
Štěpán Starosta (Dept. of Math., FNSPE, CTU in Prague)
Milena Svobodová (TIGR, Dept. of Math., FNSPE, CTU in Prague)
Jan Trávníček (Czech Technical University in Prague)
Tomáš Vávra (Dept. of Math., FNSPE, CTU in Prague)
Miloslav Znojil (Nuclear Physics Institute ASCR in Rez)

Program:

Saturday, March 19, 2011	
8:30–9:00	registration
9:00–9:05	opening
9:05–9:50	J. Bobok: On three results from discrete dynamical systems
9:50–10:35	Š. Holub: Variables, equations, and solutions in words
10:35–11:05	coffee break
11:05–11:50	P. Kůrka: Symbolic dynamics in Moebius number systems
11:50–12:15	T. Hejda: Moebius number systems with discrete groups
12:15–12:40	Š. Starosta: Rauzy triangle
12:40–14:00	lunch
14:00–14:25	D. Dombek: From Rényi to $(-\beta)$-integers
14:25–14:50	T. Vávra: Arithmetics in number systems with positive and negative non-integer base
14:50–15:15	J. Kolář: Generalized Continued Fractions
15:15–15:35	coffee break
15:35–16:00	S. Cheng: Beta-Power Series and Beta-Products
16:00–16:25	Ľ. Balková: Palindromes in infinite words
16:25–16:50	K. Břinda: Abelian complexity of infinite words
16:50–17:15	E. Pelantová: From quasicrystals to exchange of intervals

Sunday, March 20, 2011	
9:00–9:45	J. Holub: Current Trends in Stringology
9:45–10:10	P. Procházka: Blockwise natural language compression
10:10–10:55	J. Janoušek: Arbology
10:55–11:20	coffee break
11:20–11:45	T. Flouri: Tree Template Matching in Ranked, Ordered Trees
11:45–12:10	J. Trávníček: Nonlinear Tree Pattern Matching by Pushdown Automata
12:10–12:35	M. Plicka: Subtree Oracle Pushdown Automata for Trees in Prefix Notation
12:35–13:00	M. Svobodová: Parallel Addition in Non-standard Numeration Systems
13:00–14:20	closing and lunch

Abstracts:

Palindromes in infinite words

Ľubomíra Balková
Department of Mathematics, FNSPE CTU in Prague

In this talk, we will present palindromes – words that stay the same when read backwards – from several points of view. First, we will consider the properties based on palindromes characterizing Sturmian words, i.e., binary aperiodic infinite words with the lowest possible complexity. Possible generalizations of such characterizations to multiliteral alphabets will be discussed. Second, we will concentrate on palindromic richness of infinite words. In particular, we will present several equivalent descriptions of rich words, i.e., infinite words all of whose factors contain the maximum number of palindromes. Third, we will mention a recent conjecture concerning the so-called palindromic defect of infinite words and some other open problems related to palindromes.

On three results from discrete dynamical systems

Jozef Bobok
Department of Mathematics, FCE CTU in Prague

In the first part of this talk we discuss some properties of (universal) Banach spaces of real functions in the context of topological entropy. Among other things, we show that any subspace of $C([0,1])$ which is isometrically isomorphic to ℓ_1 contains a functions with infinite topological entropy. Also, for any $t \in [0,\infty]$, we construct a (one-dimensional) Banach space in which any nonzero function has topological entropy equal to t. In the second part we deal with polygonal billiards: Two polygons P, Q are code equivalent if there are billiard orbits u, v which hit the same sequence of sides and such that the projections of the orbits are dense in the boundaries $\partial P, \partial Q$. Our main results show when code equivalent polygons have the same angles, resp. are similar, resp. affinely similar. The third part is devoted to Parry's theorem claiming that any continuous piecewise monotone interval map f with the positive topological entropy $h(f)$ is semiconjugated to some piecewise affine map with the constant slope $e^{h(f)}$.

Abelian complexity of infinite words

Karel Břinda
Department of Mathematics, FNSPE CTU in Prague

In this presentation, we will introduce Abelian complexity of infinite words. Let $\mathbf{u}$ be an infinite word over an alphabet $\mathcal{A} = \{a_0, a_1, \ldots, a_{k-1}\}$. We define Abelian complexity as a mapping $\mathcal{AC} : \mathbb{N} \to \mathbb{N}$:

$$\mathcal{AC}(n) = \#\{\Psi(v) | v \in \mathcal{L}_n\},$$

where $\mathcal{L}_n$ is the set of all factors of $\mathbf{u}$ of length n and $\Psi(v) = (|v|_{a_0}, |v|_{a_1}, \ldots, |v|_{a_{k-1}})$ is the so-called Parikh vector of the finite word v. First, we will show the connection between Abelian complexity, Balance function

$$B(n) = \max_{a \in \mathcal{A}} \max_{v,w \in \mathcal{L}_n} (|v|_a - |w|_a),$$

and matrices of primitive substitutions. We will then present which classes of primitive substitutions are already well described concerning the Abelian complexity. Finally, we will mention some current open questions.

From Rényi to $(-\beta)$-integers

Daniel Dombek
Department of Mathematics, FNSPE CTU in Prague

The talk is devoted to positional numeration systems with both positive and negative bases. We start by recalling the definition and the most important properties of β-expansions, introduced by R'enyi in 1957. In particular, we mention the admissibility of digit strings and the definition of the set $\mathbb{Z}_\beta$ of β-integers. Then we present numeration system with $(-\beta) < -1$ as its base. We discuss the elementary features of such a system in a quite general setting and show the correspondence with $(-\beta)$-expansions introduced by Ito and Sadahiro in 2009. Although systems with positive and negative bases are in many aspects very similar, we show that certain properties are essentially different. We illustrate this phenomena on the example of the set of $(-\beta)$-integers, denoted by $\mathbb{Z}_{-\beta}$, whose description is much more complicated than that of $\mathbb{Z}_\beta$.

Tree Template Matching in Ranked, Ordered Trees

Tomáš Flouri
Department of Theoretical Computer Science, FIT CTU in Prague

We consider the problem of tree template matching and propose a solution based on the bottom-up technique. Specifically, we solve the tree pattern matching problem as a string matching problem, by transforming the tree template and the subject tree to strings representing their postfix notation, and using pushdown automata as the computational model. The method is analogous to the construction of string pattern matchers: For a given tree template, a nondeterministic pushdown automaton is created, and it is then transformed to an equivalent deterministic pushdown automaton. Although the upper theoretical bound of space complexity is exponential to the size of the tree template, the lower bound is linear to the size of the tree template, in case the tree template does not consist of pairwise independent subtrees. The matching time (for both cases) is linear to the size of the subject tree.

Moebius number systems with discrete groups

Tomáš Hejda
Department of Mathematics, FNSPE CTU in Prague

We study number systems generated by Moebius transformations (MT) of the hyperbolic plane $\mathbb{U} = \{z \in \mathbb{C} | \Im z \geq 0\}$. We are concerned about finitely generated groups of MTs that are discrete in the group of all MTs. Any MT is a map $z \to \frac{az+b}{cz+b}$ with parameters $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$. We will try to answer the question of existence of a discrete group of MTs such that all its elements have rational parameters and corresponding number system is covergent.

Current Trends in Stringology

Jan Holub
Department of Theoretical Computer Science, FIT CTU in Prague

The stringology started in seventies with simple pattern matching tasks. The problems being solved in the stringology got more and more complicated. Many different approaches were used. Nowadays the stringology is closely connected to application fields like bioinformatics, data compression, and musicology. We show the overview of approaches in stringology and current trends in stringology.

Variables, equations, and solutions in words

Štěpán Holub
Department of Algebra, Charles University in Prague

I give a survey of results and open problems regarding equations in words and related topics like test sets and equality sets. Included will be questions of deciding whether an equation is solvable, finding its solution or determining the structure of the set of all solutions. Focus will be on the state of the art in simplest cases, namely involving one, two and three variables.

Beta-Power Series and Beta-Products

Sirui Cheng
Department of Mathematics, FNSPE CTU in Prague

First, we take elementary functions which can be written as power series or infinite product and define beta-power series and beta-product by replacing the rational integers with beta-integers. Later, we examine the newly defined power series and products, i.e. convergence radius, error. Finally, we will show the graphs of our beta-power series and beta-products in comparison to the rational ones.

Generalized Continued Fractions

Jakub Kolář
Department of Mathematics, FNSPE CTU in Prague

We study a generalization of the classical continued fraction expansion, with coefficients in a more general discrete set M. We present the work of J. Bernat, in which he used τ-integers for the set M, where τ is the golden ratio. We study the convergence and other properties of such generalized continued fractions. Finally we expose a number of open problems.

Symbolic dynamics in Moebius number systems

Petr Kůrka
Center for Theoretical Study, Prague

We characterize interval Moebius number systems with sofic subshifts and show that they can be obtained as factors of interval Moebius number systems with subshifts of finite types. The endpoints of cylinders of such systems can be computed by a fast algorithm which generalizes the computation of the Farey fractions in the Stern-Brocot tree. We treat in detail the bimodular number system which is the simplest system which complies to our theory and exhibits many symmetries.

Arbology

Jan Janoušek
Department of Theoretical Computer Science, FIT CTU in Prague

Many various algorithms were developed for operations on trees used in many applications. Stringology represents a systematic approch to the construction of algorithms for processing strings. The computation model of stringology algorithms can be a finite automaton. Arbology, a new algorithmic discipline, aims to be an analogous systematic approach to the construction of algorithms for processing trees. Instead of finite automaton, arbology uses a standard pushdown automaton which reads a linear notation of the tree. Basic arbology algorithms for matching and indexing trees and finding repeats in trees are presented.

From quasicrystals to interval exchange

Edita Pelantová
Department of Mathematics, FNSPE CTU in Prague

Cut-and-project sets are commonly accepted as a model for physical materials called quasicrystals. Cut-and-project sets arise by projection of chosen lattice points from higher dimensional space to a suitable oriented physical subspace. The symbolic dynamical system coding exchange of two or three intervals can be interpreted as cut-and-project sets. This point of view brings a powerful tool to the study of combinatorial properties of words coding 3-interval exchange (3-iet words). Among many, we mention substitution invariance, palindromic and factor complexity and enumeration of factors occurring in 3-iet words.

Subtree Oracle Pushdown Automata for Trees in Prefix Notation

Martin Plicka
Department of Theoretical Computer Science, FIT CTU in Prague

Oracle modification of subtree pushdown automata for trees in prefix notation and prefix bar notation is presented. Subtree pushdown automata represent a complete index of a tree for subtrees. Subtree pushdown automata accept all subtrees which match the tree. Oracle pushdown automata, as inspired from factor oracle automata from stringology, have number of states equal to n+1 where n is length of input string representing the tree in prefix or prefix bar notation, respectively. This makes space complexity minimal. The presented pushdown automata are input-driven and therefore they can be determinised. Generally, oracle versions of these automata, analogously to factor oracle automata in stringology, also accepts trees which are not present in given subject tree. In comparison to factor oracle, the set of false positive matches tends to be smaller because not all substrings of a tree in prefix notation represent tree.

Block-wise natural language compression

Petr Procházka
Department of Theoretical Computer Science, FIT CTU in Prague

We address the problem of blockwise natural language compression. The blockwise compression is semi-adaptive in terms of one block but it is adaptive in terms of whole input. Our blockwise compression method is based on Dense Code idea. It achieves very good compression ratio around 32% on natural language text and at the same time it has some interesting properties

which could be applied in digital libraries. The compression method allows direct searching on compressed text. Moreover the vocabulary can be used as a block index which makes some kinds of searching very fast. Another property is that the compressor can send single blocks with corresponding vocabulary which is considerate to limited bandwidth. In addition the compressed file can be continuously extended without need of previous decompression.

Parallel Addition in Non-standard Numeration Systems

Milena Svobodová
Department of Mathematics, FNSPE CTU in Prague

We consider numeration systems where digits are integers and the base is an algebraic number β such that $|\beta| > 1$ and the modulus of all the conjugates of β differs from 1. For this broad class of bases β, we can find an alphabet of signed-digits on which addition is realizable by a parallel algorithm in constant time. An important property of this algorithm is that it is a p-local function. We also discuss the question of cardinality of the used alphabet, and we are able to modify our algorithm in order to work in a smaller alphabet. For the particular case of the base β being the Golden Mean, we further refine the construction to obtain a parallel algorithm (again a p-local function) on the alphabet $-1, 0, 1$, which is the minimum alphabet allowing parallel addition in this base.

Rauzy triangle

Štěpán Starosta
Department of Mathematics, FNSPE CTU in Prague

P. Arnoux and G. Rauzy described in 1991 a symbolic dynamical system where trajectories are coded by a special class of ternary infinite words nowadays called Arnoux-Rauzy or strict episturmian words. We will show a dual system and expose its relation with frequencies of letters in ternary Arnoux-Rauzy words. Using the dual system, we will show that the set of all frequencies of letters of a ternary Arnoux-Rauzy word is of Lebesgue measure 0, as already conjectured in 1991.

Nonlinear Tree Pattern Matching by Pushdown Automata

Jan Trávníček
Department of Theoretical Computer Science, FIT CTU in Prague

A new approach for nonlinear tree pattern matching is presented. The approach uses nonlinear tree pattern pushdown automaton, which represents a full index of tree for nonlinear tree patterns. This automaton can be determinised and the time of finding all locations of a tree pattern is linear with size of the tree pattern. Several modifications to this automaton are presented for more complex use.

Arithmetics in number systems with positive and negative non-integer base

Tomáš Vávra
Department of Mathematics, FNSPE CTU in Prague

We consider positional number system with base $\beta > 1$ defined by Rényi and its analogue with negative base recently introduced by Ito and Sadahiro. We give an overview of known arithmetical properties of such systems. In particular we describe the set $\mathrm{Per}(\beta)$ and $\mathrm{Per}(-\beta)$ of numbers with periodic expansions; we study the algebraic structure of the set $\mathrm{Fin}(\beta)$ and $\mathrm{Fin}(-\beta)$ of numbers with finite expansions. We also question the complexity of arithmetical operations.