

Report on CPM 2003
14th Annual Symposium on Combinatorial Pattern Matching
Morelia, Michoacán, Mexico, June 25–27, 2003

Gabriel Valiente, Technical University of Catalonia, Barcelona, Spain

The 14th Annual Symposium on Combinatorial Pattern Matching, CPM 2003, was held on June 25–27, 2003, at the Universidad Michoacana in Morelia, Michoacán, Mexico. Over the past decade, this symposium has gathered leading experts in the field of searching and matching all kind of patterns, ranging from strings and trees to regular expressions, point sets, arrays, and graphs, providing an excellent arena for discussing their combinatorial properties as well as their applications in areas such as data mining, computational biology, computer vision, image processing, and multimedia systems. Previous meetings were held in Paris (1990), London (1991), Tucson (1992), Padova (1993), Asilomar (1994), Helsinki (1995), Laguna Beach (1996), Aarhus (1997), Piscataway (1998), Warwick (1999), Montréal (2000), Jerusalem (2001), and Fukuoka (2002).

The conference was organized by Edgar Chávez of the Universidad Michoacana in Morelia, Michoacán, Mexico and was supported by the Universidad Michoacana and the Consejo Nacional de Ciencia y Tecnología (CONACyT). There were over 50 participants in total, including 16 from good old Europe, 16 from America and 10 from Middle and Far East.

The conference started on Wednesday, June 25 with an invited lecture by Vladimir Levenshtein of the Keldysh Institute for Applied Mathematics, Russia, about *Efficient Reconstruction of Sequences*. Levenshtein is well known among the CPM community for having presented in *Binary Codes capable of Correcting Deletions, Insertions, and Reversals* (Cybernetics and Control Theory, 10(8):707–710, 1966) the earliest use of a sequence distance function which is appropriate in the presence of deletion and insertion errors.

This was followed by a session on *Genome Rearrangements*, chaired by Dan Gusfield, in which Elad Verbin talked about *Efficient Data Structures and a New Randomized Approach for Sorting Signed Permutations by Reversals* (joint work with Haim Kaplan), Maria Emilia M. T. Walter talked about *Working on the Problem of Sorting by Transpositions on Genome Rearrangements* (joint work with Luiz Reginaldo A. F. Curado and Adilton G. Oliveira), and Tzvika Hartman talked about *A Simpler 1.5-Approximation Algorithm for Sorting by Transpositions*.

After a coffee break, the second session of the morning was on *Alignments*, chaired by Vladimir Levenshtein, in which Gad M. Landau talked about *Sparse LCS Common Substring Alignment* (joint work with Baruch Schieber and Michal Ziv-Ukelson), Bin Ma talked about *Alignment between Two Multiple Alignments* (joint work with Zhuozhi Wang and Kaizhong Zhang), and Mohamed Ibrahim Abouelhoda talked about *Multiple Genome Alignment: Chaining Algorithms Revisited* (joint work with Enno Ohlebusch).

The generous lunch break, long enough to accommodate a Spanish “siesta,” was followed by a *Miscellaneous* session, chaired by Gad M. Landau, in which François Nicolas talked about *Complexities of the Centre and Median String Problems* (joint work with Eric Rivals), and a session on *Compression*, also chaired by Gad M. Landau, in which Marinella Sciortino talked about *Optimal Partitions of Strings: A New Class of Burrows-Wheeler Compression Algorithms* (joint work with Raffaele Giancarlo), and Hiroshi Sakamoto talked about *A Fully Linear-Time Approximation Algorithm for Grammar-Based Compression*.

After a coffee break, the second session of the afternoon was on *Trees*, chaired again by Gad M. Landau, in which H el ene Touzet talked about *Analysis of Tree Edit Distance Algorithms* (joint work with Serge Dulucq), and Gabriel Valiente talked about *Constrained Tree Inclusion*.

Thursday, June 26 started straight with a first session on *Computational Biology*, chaired by Kaizhong Zhang, in which Bin Ma talked about *An Effective Algorithm for the Peptide De Novo Sequencing from MS/MS Spectrum* (joint work with Kaizhong Zhang and Chengzhi Liang), Behshad Behzadi talked about *An Improved Algorithm for Generalized Comparison of Minisatellites* (joint work with Jean-Marc Steyaert), Giulio Pavesi talked about *Pattern Discovery in RNA Secondary Structure Using Affix Trees* (joint work with Giancarlo Mauri), and Olivier Elemento talked about *An Exact and Polynomial Distance-Based Algorithm to Reconstruct Single Copy Tandem Duplication Trees* (joint work with Olivier Gascuel).

After a coffee break, the second session of the morning was a first session on *Pattern Matching*, chaired by Gabriel Valiente, in which Gonzalo Navarro talked about *Average-Optimal Multiple Approximate String Matching* (joint work with Kimmo Fredriksson), Leena Salmela talked about *Tuning String Matching for Huge Pattern Sets* (joint work with Jari Kyt ojoki and Jorma Tarhio), Veli M akinen talked about *On Minimizing Pattern Splitting in Multi-track String Matching* (joint work with Kjell Lemstr om), and Amihoud Amir talked about *Two-Dimensional Pattern Matching with Rotations* (joint work with Ayelet Butman, Maxime Crochemore, Gad M. Landau, and Malka Schaps).

Another generous lunch break was followed by a second session on *Pattern Matching*, chaired by Gonzalo Navarro, in which Johann Pelfr ene talked about *Extracting Approximate Patterns* (joint work with Sa id Abdedda im and Jo el Alexandre). Neither Nadia Nedjah nor Luiza de Macedo Mourelle were present to talk about *More Efficient Left-to-Right Pattern Matching in Non-Sequential Equational Programs*.

This was immediately followed by a second session on *Computational Biology*, chaired by Amihoud Amir, in which Dan Gusfield talked about *Haplotype Inference by Pure Parsimony*, Brona Brejov a talked about *Optimal Spaced Seeds for Hidden Markov Models, with Application to Homologous Coding Regions* (joint work with Daniel G. Brown and Tom as Vinar), and Stefano Lonardi talked about *Efficient Selection of Unique and Popular Oligos for Large EST Databases* (joint work with Jie Zheng, Timothy J. Close, and Tao Jiang).

This second day finished with a Conference Dinner which gathered all partic-

ipants at a restaurant in the city of Morelia. The brave ones dared to try typical Mexican food, which was delicious, and an excellent selection of fine Spanish wines accompanied the peer review of scientific stories until late in the evening.

Friday, June 27 started with an invited lecture by Ian Munro of the University of Waterloo, Canada, about *Succint Representations of Combinatorial Objects*.

This was followed by a session on *Suffix Trees and Arrays*, chaired by Ian Munro, in which Dong Kyue Kim talked about *Linear-Time Construction of Suffix Arrays* (joint work with Jeong Seop Sim, Heejin Park, and Kunsoo Park), Pang Ko talked about *Space Efficient Linear Time Construction of Suffix Arrays* (joint work with Srinivas Aluru), Juha Kärkkäinen talked about *Fast Lightweight Suffix Array Construction and Checking* (joint work with Stefan Burkhardt) and last, but not least, Raphaël Clifford talked about *Distributed and Paged Suffix Trees for Large Genetic Databases* (joint work with Marek J. Sergot).

A short lunch break was the necessary prelude to a half-day excursion into the Pátzcuaro region, visiting the town of Tzintzuntzan, located on the shores of the Pátzcuaro lake and home of both a Franciscan convent and some of the oldest olive trees in the world, and the city of Pátzcuaro itself, a main pre-Hispanic religious center. Most participants managed to spend a few more days in Mexico, visiting the National Museum of Anthropology in the city of Mexico itself and even the Maya pyramids of Teotihuacan.

The proceedings of CPM 2003, edited by Ricardo Baeza-Yates, Edgar Chávez, and Maxime Crochemore, have been published by Springer-Verlag as volume 2676 of Lecture Notes in Computer Science and can be found in the CPM archive at <http://www.cs.ucr.edu/~stelo/cpm/>, together with the PDF or PPT files of all contributed talks. A special issue of the Journal of Discrete Algorithms, dedicated to combinatorial pattern matching and its applications and edited by Ricardo Baeza-Yates and Maxime Crochemore, is scheduled to appear in 2004.

The 15th Annual Symposium on Combinatorial Pattern Matching, CPM 2004, will be held in Istanbul, Turkey, June 25–27, 2004, organized by Cenk Sahinalp of Case Western Reserve University. For details, see the CPM 2004 web page at <http://genomics.cwru.edu/cpm/>.