



## Editorial

Combinatorial pattern matching addresses issues of searching and matching strings and more complicated patterns such as trees, regular expressions graphs, point sets and arrays. The goal is to derive non-trivial combinatorial properties of such structures and to exploit these properties in order to achieve superior performance for the corresponding computational problems.

Over recent years a steady flow of high-quality research on this subject has changed a sparse set of isolated results into a fully-fledged area of algorithmics. This area is continuing to grow even further due to the increasing demand for speed and efficiency that comes from important and rapidly expanding applications such as the World Wide Web, computational biology and multimedia systems, involving requirements for information retrieval, data compression and pattern recognition. Starting in 1990, the CPM series of annual conferences has provided an international forum for research in combinatorial pattern matching and related applications.

The general organization and orientation of CPM conferences is coordinated by a steering committee composed of Alberto Apostolico, Maxime Crochemore, Zvi Galil and Udi Manber.

After the first meeting in Paris, a selection of papers appeared as a special issue of *Theoretical Computer Science* in Volume 92. Extended abstracts of the third to fourteenth symposia, in Tucson (1992), Padua (1993), Asilomar (1994), Helsinki (1995), Laguna Beach (1996), Aarhus (1997), Piscataway (1998), Warwick (1999), Montreal (2000), Jerusalem (2001), Fukuoka (2002) and Morelia (2003) appeared as volumes 644, 684, 807, 937, 1075, 1264, 1448, 1645, 1848, 2089, 2373 and 2676, respectively, of the Springer Lecture Notes in Computer Science.

After the eleventh symposium, held June 21–23, 2000 at the Université de Montréal, we undertook to select and publish full versions of papers exemplifying the diverse ongoing progress in the field. We solicited the participation of authors of 13 of the 25 contributed papers (already chosen by the program committee from among 44 extended abstracts submitted in response to the original call for papers) and three invited talks. Nine of the manuscripts were accepted after the refereeing process and appear in this special issue of JDA.

1. Ian Witten, *Adaptive text mining: inferring structure from sequences* gives a presentation of how text compression techniques can be successfully used for text mining, a rather novel direction in data mining. A number of examples are presented.
2. R. Backofen, *A polynomial time upper bound for number of contacts in the HP-model on the face-centered-cubic lattice (FCC)*, provides an extensive treatment of this topic, central to protein folding studies, including algorithms for the computation of the bound.

3. D. Liben-Nowell and J. Kleinberg, *Structural properties and tractability results for linear synteny*, show that many useful properties of syntenic distance carry over to linear syntenic distance. Moreover, the first polynomial time algorithm exactly solving linear synteny for a non-trivial class of instances is provided.
4. D. Bryant, *A lower bound for the breakpoint phylogeny problem*, derives lower bounds for two breakpoint problems. Those are of great significance for the assessment of performance of practical breakpoint phylogeny algorithms. Methods for the computation of the bounds are part of the GoTree software package.
5. T. Jiang, Guo-Hui Lin, Bin Ma and K. Zhang, *The longest common subsequence problem for arc-annotated sequences*, give algorithmic and complexity results for this novel problem formulation, of great significance for the representation of structural information of RNA and protein sequences.
6. D. Fernández-Baca, T. Seppäläinen and G. Slutzki, *Parametric multiple sequence alignment and phylogeny construction*, give bounds on the size of the parameter-space decomposition induced by multiple sequence alignment problems where phylogenetic information may be given or inferred.
7. L. Gaşieniec, J. Jansson and A. Lingas, *Approximation algorithms for Hamming clustering problems*, study Hamming versions of two classical clustering problems, show that some versions of those problems can be solved in polynomial time, while others cannot be approximated within any factor smaller than two, unless  $P = NP$ .
8. S.-R. Kim and K. Park, *A dynamic edit distance table*, consider the incremental/decremental version of the classic edit distance problem and provides a very simple and intuitive algorithm for it.
9. S. Skiena and P. Sumazin, *Shift error detection in standardized exams*, provide the first extensive study of how to correct errors in multiple-choice exams. They propose algorithms that accurately detect and correct those errors and describe an extensive experimental study on undergraduate Stony Brook exam sets and over 100,000 Scholastic Aptitude Tests.

The 11th Symposium on Combinatorial Pattern Matching was hosted and supported by the Centre de recherches mathématiques of the Université de Montréal, in the context of a thematic year on Mathematical Methods in Biology and Medicine (2000–2001). It was attended by 73 researchers and students, from 15 countries. Nadia El-Mabrouk and Louis Pelletier collaborated on the local arrangements.

Raffaele Giancarlo would like to acknowledge kind support to his research, including editing this special issue, from the Italian MIUR Projects PRIN “Bioinformatica e Ricerca Genomica” and FIRB “Bioinformatica per la Genomica e la Proteomica”. Part of this work was done while visiting Institut Gaspard Monge, Université Marne La Vallée, France, under the auspices of CNRS.

Raffaele Giancarlo

*Dipartimento di Matematica, University of Palermo, Italy*

*E-mail address: raffaele@math.unipa.it (R. Giancarlo)*

David Sankoff

*Department of Mathematics and Statistics, University of Ottawa, Canada*

*E-mail address: sankoff@uottawa.ca (D. Sankoff)*