



Editorial

StringMasters 2012 & 2013 special issue – volume 2



This is the second volume of the Special Issue on StringMasters 2012 & 2013, in which we publish selected papers on the topics of the StringMasters workshop series from the years 2012 and 2013, held in Rouen, London, Verona, and Prague. The topics of these open problems workshops are what is increasingly being called *stringology*: algorithms on strings, combinatorics on words, and their applications in any related research area, among these, bioinformatics, big data, compression, musicology, and internet issues, to name just a few. From these workshops, many fruitful collaborations have been born, the output of some of which is contained in this volume. We also accepted contributions whose authors have not yet participated in any of the workshops, but whose topics matched the areas of the StringMasters workshops. The fact that these workshops have had so much success, and that we received so many high quality submissions so as to merit two separate volumes, is evidence of the expansion and maturity of this area of research.

This volume contains eight regular papers and one brief reflective article, with which we start the volume. This first piece, “Perspectives” by W.F. (Bill) Smyth, one of the founding fathers of the StringMasters workshop series, analyses the history, development and outlook of the research field *stringology*, and argues that it has reached a high level of maturity and independent interest during the past few decades, since its early days.

We continue with two papers on reverse engineering of different data structures for strings: The first, by Alatabbi, Rahman, and Smyth, deals with indeterminate strings, a generalization of strings where the individual characters can be subsets of the alphabet. These have many applications, in particular in bioinformatics. This paper presents an algorithm for inferring an indeterminate string from its prefix graph, a recently introduced data structure for strings.

The second paper, by Starikovskaya and Vildhøj, studies the reverse engineering problem for suffix trees, one of the most celebrated data structures for strings: what properties does a tree have to exhibit in order to be the suffix tree of some string? The paper also gives an algorithm for special cases to construct such a string. This connects nicely to the first volume of this Special Issue, where a solution for a different variant of the same problem was given.

We continue with a substantial theoretical paper by Condon, Manuch and Thachuk, on the complexity of string partitioning, which establishes hardness results for a number of variants of a problem with important applications in synthetic biology.

Next follows a paper by Crochemore, Grossi, Kärkkäinen, and Landau, presenting a new in-place algorithm for computing the Burrows–Wheeler-Transform (BWT) of a string, which requires constant additional space and uses only combinatorial properties of the BWT. This can lead to efficiencies when a text occupies a large part of the main memory.

The fifth paper in this volume, by Gog, Navarro, and Petri, deals with another celebrated data structure on strings, compressed suffix arrays (CSA), and presents new techniques which allow reporting pattern occurrences in the same order as they occur in the text, as well as in particular text windows, in times which are much lower than the baseline techniques. The results are supported by extensive experimentation on different types of textual data.

We continue with compressed data structures: The paper by Gagie, Gawrychowski, and Puglisi presents a new efficient algorithm for an important pattern matching problem, that of finding all approximate occurrences of a pattern in an LZ77-compressed text.

Our penultimate article, by Navarro and Thankachan, introduces a new problem, bottom- k retrieval, with potential applications in many relevant data mining areas: retrieving from a document collection all those documents with the least number of occurrences of a pattern. The authors show that when the documents are compressed, the classical solutions for top- k retrieval cannot be easily extended, and give efficient solutions using the compressed suffix array.

We close this volume with a contribution by Yoshida and Kida, which gives a new encoding scheme for variable-to-fixed encoding using multiplexed parse trees, extending a recent encoding scheme. The authors provide experiments in which they show that the new scheme improves compressed pattern matching considerably on natural language texts.

As with Volume 1, all manuscripts underwent a rigorous review process, and we appreciate the generous spirit of the many anonymous reviewers who devoted much of their time and expertise to this Special Issue. We also warmly thank the staff at the Journal of Discrete Algorithms for their continued attention to high-quality scientific publication.

The StringMasters workshops continue to be held internationally, several times per year. We hope to expand the StringMasters community by attracting both frequent and especially new participants who are united in their interest in both fundamental and practical problems on strings. A new StringMasters venue, the Warsaw Center of Mathematics and Computer Science, University of Warsaw, hosts the first meeting in 2015. For information about past, present and future StringMasters meetings, see <http://www.stringmasters.org/>.

Special Issue Guest Editors

Maxime Crochemore^{a,b}

^a*King's College London, UK*

^b*Université Paris-Est, France*

Jacqueline W. Daykin

Royal Holloway & King's College University of London, UK

Zsuzsanna Lipták

University of Verona, Italy

E-mail address: zsuzsanna.liptak@univr.it

Available online 21 January 2015