Preface

This volume contains the papers presented at the Fifth International Conference on Similarity Search and Applications (SISAP 2012), which took place on August 9–10, 2012 at the Fields Institute for Research in Mathematical Sciences, Toronto, Ontario, Canada.

SISAP is a conference devoted to similarity searching, with emphasis on metric space searching. It aims to fill in the gap left by the various scientific venues devoted to similarity searching in spaces with coordinates, by providing a common forum for theoreticians and practitioners around the problem of similarity searching in general spaces (metric and non-metric) or using distance-based (as opposed to coordinate-based) techniques in general. Four types of contributions are welcome: (1) fundamental techniques to handle general similarity search problems, (2) applied techniques to solve particular similarity search problems of wide interest, (3) new similarity search problems, where their features and challenges are studied, and (4) actual systems for similarity search, in the form of demos. SISAP is not only seen as a forum for exchanging new indexing techniques and real-world applications, but also common testbeds and benchmarks, and source code. Authors are expected to use the testbeds and code from the SISAP web site (`www.sisap.org`) for comparing new applications, databases, indexes and algorithms.

This year we received 19 full-paper and two demo submissions, from Argentina, Chile, Czech Republic, France, Japan, Mexico, Norway, Russia, Spain, Switzerland, United Kingdom, and United States. Each submission was assigned, in double-blind mode, to three PC members, who reviewed them themselves and/or supervised subreviews. Submissions received 2 to 5 reviews (3.14 on average). Then the PC chairs and involved members discussed the articles where no obvious agreement had been reached. The final decisions of acceptance or rejection were made by the PC chairs. Finally, 14 full papers and the two demos were selected to be presented in the Conference and to appear in the Proceedings.

From the full papers accepted, 9 refer to techniques to handle general similarity search problems, improving upon the state of the art on topics like parallelism, dynamism, secondary memory, approximation techniques, optimized construction, combinations of data structures, and novel scenarios, like streams of related searches and inferring factual space properties from the data. Further, two accepted papers refer to applied techniques, to similarity searching in string dictionaries and in images. The other three papers study the properties of specific spaces like sequences under time-warping distance and factorized tensors, and propose and study new distances for vector spaces based on entropy correlations. From the two demos, one presents an image meta-search engine, and the other introduces a tool for identifying protein and peptide sequences from tandem mass spectra.

Overall, the articles formed an extremely stimulating set of contributions to many of the most relevant aspects of similarity searching. Two invited presentations and papers from prominent researchers further enriched this year's SISAP. The first one, "Effective Principal Component Analysis", by Santosh Vempala,

is about the success and challenges around this technique, of wide relevance in similarity search and various other fields. The second one, "Future Trends in Similarity Searching", by Pavel Zezula, is a revealing survey and analysis of where the discipline is expected to head in the forthcoming years.

This year the Proceedings of SISAP were published by Springer-Verlag, in the *Lecture Notes in Computer Science* series. A selection of the best papers was recommended for inclusion in a special issue of *Information Systems* journal dedicated to this conference. Those were chosen by the PC chairs based on the original reviews of the articles and their oral presentation during the conference, as well as appropriateness to the journal.

The subject matter of the SISAP conferences, although primarily a Computer Science topic, uses a great deal of advanced mathematical methods, such as those of geometric functional analysis and statistical machine learning. The conference is a perfect platform for interactions between computer scientists and mathematicians, and the stimulating research ambiance of the Fields institute gave fresh impetus to such interactions. We thank the Fields institute for the hosting of SISAP 2012 conference.

Last, but not least, we acknowledge the generous financial support from (again) the Fields Institute for Research in Mathematical Sciences, Canada; the Canadian Network of Excellence in Mathematics of Information Technology and Complex Systems (MITACS); and Natural Sciences and Engineering Research Council of Canada (NSERC) research grant "New Set-theoretic Tools for Statistical Learning". All the submission, reviewing, and Proceedings generation processes were handled through the Easychair platform.

August 2012

Gonzalo Navarro, Santiago, Chile
Vladimir Pestov, Ottawa, Canada
Program Chairs
SISAP 2012

Organization

Committees

Steering Committee:	Edgar Chávez (Universidad Michoacana, Mexico) Gonzalo Navarro (University of Chile, Chile)
PC Chairs:	Gonzalo Navarro (University of Chile, Chile) Vladimir Pestov (Université d'Ottawa, Canada)
PC members:	Edgar Chávez (Universidad Michoacana, Mexico) Paolo Ciaccia (Università di Bologna, Italy) Alfredo Ferro (Università di Catania, Italy) Daniel Keim (Universität Konstanz, Germany) Daniel Miranker (University of Texas at Austin, USA) Marco Patella (Università di Bologna, Italy) Hanan Samet (University of Maryland, USA) Tomáš Skopal (Charles University in Prague, Czech Republic) Aleksandar Stojmirović (NCBI/NLM/NIH, USA) Agma Traina (Universidade de São Paulo – São Carlos, Brazil) Pavel Zezula (Masaryk University, Czech Republic)
Organization Chair:	Vladimir Pestov (Université d'Ottawa, Canada)
Publicity Chair:	Tomáš Skopal (Charles University in Prague, Czech Republic)

Additional Reviewers

Marco Adelfio	David Hoksza	Luís M. Silveira Russo
Gelio Alves	Eamonn Keogh	Jan Sedmidubsky
Michal Batko	Jakub Lokoč	John Spouge
Petra Budikova	Jiří Novák	Eric Sadit Téllez Avila
Benjamin Bustos	Sarana Nutanong	Lee Thompson
Carlos Castillo	Ives Pola	German Tischler
Vlastislav Dohnal	Mônica R. Porto Ferreira	Kesheng Wu
Magnus Lie Hetland	Nora Reyes	

Sponsoring Institutions

Fields Institute for Research in Mathematical Sciences, Canada.
Canadian Network of Excellence in Mathematics of Information Technology and Complex Systems (MITACS).
Natural Sciences and Engineering Research Council of Canada (NSERC).

Table of Contents

Searching in Specific Spaces

New Similarity Spaces

Demo Papers