Editorial

This collection of articles corresponds to the best papers of the *First International Workshop on Similarity Search and Applications (SISAP)*. This new conference is devoted to similarity searching, and in particular to metric space searching, that is, the case where only a black-box distance can be used in the search, either because the objects have no coordinates or because high dimensionality prevents the use of multidimensional access methods. In spite of the practical interest of this problem and its many applications, there was no special conference or forum bringing together theoreticians and practitioners around it.

In response to the lack of established benchmarks for experimentation and to the dispersion of the publications on the topic, *SISAP* is associated to a Web site, `www.sisap.org`, containing a set of standardized implementations of algorithms and data structures for metric space searching, datasets and benchmarking tools. It aims to include in the future up-to-date literature and benchmarks, so as to become the place to exchange real-world, challenging and exciting examples of applications, and to perform fair and reproducible performance comparisons.

SISAP 2008 was held in April 11-12 in Cancun, Mexico, as an *ICDE 2008* Satellite Workshop. It has enjoyed an enthusiastic reception from the top researchers in the field, as well as by uprising authors that have contributed their papers. Fifteen regular and two invited papers were presented at the conference, and appear in the proceedings published by the IEEE Computer Society Press. The regular papers were selected from 33 submissions, by a renowned Program Committee. We received submissions from all over the world (Argentina, Canada, Chile, China, Czech Republic, Finland, France, Japan, Mexico, Spain, Switzerland, Tunisia, Turkey, UK, and the USA). The acceptance rate was 45%. The papers were selected based on their originality, relevance, and technical strength.

Contributions to the conference were divided into three categories: (1) general indexing and searching methods that apply to arbitrary metric spaces, nonmetric or (dis)similarity spaces, or high-dimensional vector spaces; (2) methods that apply to specific similarity search problems; and (3) new spaces where searching is challenging, which uncover novel applications of the paradigm

Given the scope of the *Journal of Discrete Algorithms*, we favored the papers with more algorithmic content for this special issue. The selection was done by the Program Chairs considering the conference reviews and presentations, and filtered and confirmed by the whole Program Committee. The extended conference papers underwent two or three rounds

of additional reviewing, usually by the same conference reviewers.

In article "Parallel Query Processing on Distributed Clustering Indexes", Verónica Gil-Costa, Mauricio Marín and Nora Reyes study the metric indexing and search process under the perspective of a large search engine, which needs parallelization to achieve scalable performance. This is a very important topic, for example in order to integrate multimedia search by content in Web search engines.

Rodrigo Paredes and Nora Reyes present a novel solution to the largely unexplored problem of metric joins in article "Solving Similarity Joins and Range Queries in Metric Spaces with the List of Twin Clusters". This problem is essential to incorporate metric data types into the current database technology. This is an area still needing much research, and this paper contributes to it.

Article "Approximate Similarity Search: A Multi-Faceted Problem", by Marco Patella and Paolo Ciaccia, is one of those surveys that not only cover and classify a large part of the heterogeneous research on the topic, but also dig into the fundamentals of the problem and propose models to understand it. We believe this article will quickly become a must-read survey in the area.

Matthew Skala presents a beautiful mix of theory and practice contributing to the understanding of the structure of several metric spaces under a perspective that turns out to be useful for indexing purposes, and experimentally validates his conclusions, in article "Counting Distance Permutations".

The M-tree is arguably the best-known secondary memory data structure for metric space searching, and its dynamic construction is a must for large databases. Thus we expect the article "New Dynamic Construction Techniques for M-tree" by Tomáš Skopal and Jakub Lokoč, which introduces improved techniques for constructing the M-tree dynamically, to have a large practical impact on the community. With this last article we close the special issue.

We wish to thank Benjamin Bustos, Alfredo Ferro, Vlatislav Dohnal, Edwin Jacox, Daniel Miranker, David Novak, Marco Patella, Hanan Samet, Tomáš Skopal, and Pavel Zezula, for their excellent and timely job at refereeing the articles. We would also like to thank the Editors of the *Journal of Discrete Algorithms* for giving us the opportunity to present the challenging algorithmic aspects of similarity searching to a wide audience. We hope this will help attracting the attention of other researchers to this fascinating area.

<div align="right">

Edgar Chávez and Gonzalo Navarro
SISAP 2008 Program Committee Chairs

</div>