Proceedings

sisap 2008

Table of Contents

Invited Papers

Searching in Specific Metric Spaces

Spaces and Similarities

Distributed Indexes

Index (Re)Organization

Advanced Similarity Problems

Foreword

The International Workshop on Similarity Search and Applications (SISAP) is a new conference devoted to similarity searching, with emphasis on metric space searching. It aims to fill 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.

SISAP aims to become an ideal forum to exchange real-world, challenging and exciting examples of applications, new indexing techniques, common testbeds and benchmarks, source code, and up-to-date literature through a Web page serving the similarity searching community. To encourage uniform, fair, and reproducible comparisons, SISAP authors are expected to use the testbeds and code from the SISAP Metric Library (http://sisap.org) for comparing new applications, databases, indexes, and algorithms in their papers. They are also welcome to enrich the library with their own contributions.

Contributions to the conference are welcome in 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.

SISAP 2008 is the first edition of this conference. It has enjoyed an enthusiastic reception from the top researchers in the field as well as by uprising authors who have contributed their papers. We expect that, in the years to come, SISAP will become the main reference and meeting point of the similarity search community, which for many years has been spread along several venues.

Preface

This volume contains 15 regular and two invited papers presented at SISAP 2008 conference, held in Cancún, Mexico, April 11-12, 2008. 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.

The submissions cover the spectrum of similarity searching, from theory to practice. For example, we find novel ways to regard the intrinsic properties of metric spaces, handle distributed indexes, maintain and optimize dynamic index structures, improve performance on particular cases like high-dimensional vector spaces and biological sequences, and face advanced problems such as metric joins and incremental nearest neighbor algorithms, among others. In addition, two novel metric space search challenges are presented, which are relevant in applications.

Two invited papers were presented. First, Daniel Miranker surveys the current state of the art and challenges in the application of similarity searching to computational biology problems. Second, Marco Patella and Paolo Ciaccia review approximate similarity searching, a very promising field not yet fully explored. Those surveys by leading researchers in each of the fields will surely contribute to an in-depth and up-to-date understanding of those fascinating fields and stimulate further research.

A panel to discuss new trends and open problems was held in SISAP; the main conclusions can be read in the SISAP Web page. A special issue with the extended versions of the best SISAP 2008 papers will appear in the *Journal of Discrete Algorithms* (Elsevier).

We wish to thank the Program Committee members, the organizers, and the external reviewers for their timely, committed, and high-quality work, as well as Karina Figueroa for setting up and maintaining the Metric Library, and those who contributed to it. We also thank the IEEE ICDE conference for hosting this workshop as well as Universidad de Chile and the Millennium Nucleus Center for Web Research (Chile), and Universidad Michoacana (Mexico), for providing support for this conference. Paper reviewing was done using the Easychair conference system, for which we specially thank Andrei Voronkov. Finally, we thank Olga Rodionova for not paying much attention to our suggestions, thereby enabling her to design an excellent cover.

Edgar Chavez and **Gonzalo Navarro**
Cancun, Mexico
12 April 2008

Organizing and Program Committees

Organizing Committee

Edgar Chávez

Karina Figueroa *(SISAP Library)*

Gonzalo Navarro

Marco Patella *(Publicity)*

Program Committee

Edgar Chávez, Co-Chair
Universidad Michoacana, México

Paolo Ciaccia
Universitá di Bologna, Italy

Alfredo Ferro
Universitá di Catania, Italy

Daniel Keim
Universitat Konstanz, Germany

Daniel Miranker
University of Texas at Austin, USA

Gonzalo Navarro, Co-Chair
Universidad de Chile, Chile

Marco Patella
Universitá di Bologna, Italy

Hanan Samet
University of Maryland, USA

Tomás Skopal
Charles University in Prague, Czech Republic

Pavel Zezula
Masaryk University, Czech Republic

External Reviewers

Stanislav Barton

Michal Batko

Benjamin Bustos

Vlastislav Dohnal

Rosalba Giugno

Aaron Harwood

Michael Houle

Edwin Jacox

Florian Mansmann

David Novak

Daniela Oelke

Rodrigo Paredes

Alfredo Pulvirenti

Jagan Sankaranarayanan

Jan Sedmidubsky

Dave Tahmoush

Weijia Xu

Hartmut Ziegler