Proceedings

Third International Conference on SImilarity Search and APplications

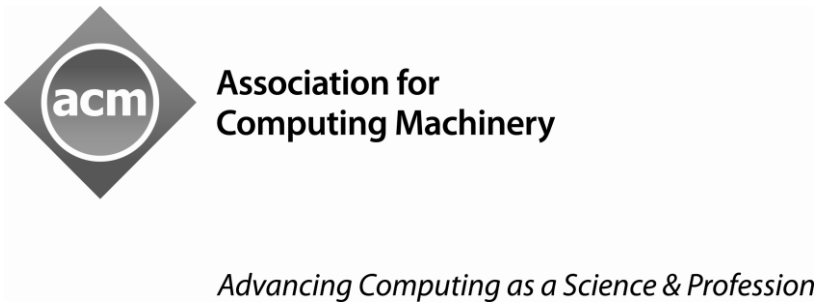
2010

Proceedings

Third International Conference on SImilarity Search and APplications

18-19 September 2010
Istanbul, Turkey

Editors:
Paolo Ciaccia and Marco Patella



The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York New York 10121-0701

ACM COPYRIGHT NOTICE. Copyright © 2010 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept., ACM, Inc., fax +1 (212) 869-0481, or permissions@acm.org.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, +1-978-750-8400, +1-978-750-4470 (fax).

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that was previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ACM ISBN: 978-1-4503-0420-7

Table of Contents

Foreword vii

Preface viii

Conference Organization ix

Invited Papers

Indexability, Concentration, and VC Theory 3
Vladimir Pestov

Where are you heading, metric access methods? A provocative survey 13
Tomáš Skopal

Indexing

Dimension Reduction for Distance-Based Indexing 25
Rui Mao, Willard Miranker, Daniel Miranker

On the Asymptotic Behavior of Nearest Neighbor Search Using Pivot-Based Indexes 33
Benjamin Bustos, Nelson Morales

Enlarging Nodes to Improve Dynamic Spatial Approximation Trees 41
Marcelo Barroso, Nora Reyes, Rodrigo Paredes

Approximate Search

Indexing Inexact Proximity Search with Distance Regression in Pivot Space 51
Ole Edsberg, Magnus Lie Hetland

On Locality-sensitive Indexing in Generic Metric Spaces 59
David Novak, Martin Kyselak, Pavel Zezula

On Locality Sensitive Hashing for Metric Spaces 67
Eric Sadit Tellez, Edgar Chavez

CP-Index: Using Clustering and Pivots for Indexing Non-Metric Spaces 75
Victor Sepulveda, Benjamin Bustos

Applications

Improving the Similarity Search of Tandem Mass Spectra using Metric Access Methods 85
Jiří Novak, Tomáš Skopal, David Hoksza, Jakub Lokoč

Case Study: An Inverted Index for Mass Spectra Similarity Query and Comparison with a Metric-space Method 93
Rui Mao, Smriti Ramakrishnan, Glen Nuckolls, Daniel Miranker

kNN Based Image Classification Relying on Local Feature Similarity 101
Giuseppe Amato, Fabrizio Falchi

Similarity Matrix Compression for Efficient Signature Quadratic Form Distance Computation . 109
Christian Beecks, Merih Seran Uysal, Thomas Seidl

Demonstrations

www.MMRetrieval.net: A Multimodal Search Engine 117
Konstantinos Zagoris, Avi Arampatzis, Savvas Chatzichristofis

SHIATSU: Annotating Your Videos the Easy Way! 119
Ilaria Bartolini, Corrado Romani

Audio Similarity Retrieval Engine 121
Pavel Jurkas, Milan Štefina, David Novak, Michal Batko

Improving the Image Retrieval System by Ranking 123
Petra Budikova, Michal Batko, Pavel Zezula

Visual Video Retrieval System Using MPEG-7 Descriptors 125
Vojtěch Zavřel, Michal Batko, Pavel Zezula

Sub-image Searching Through Intersection of Local Descriptors 127
Tomas Homola, Vlastislav Dohnal, Pavel Zezula

Posters

On Applications of Parameterized Hyperplane Partitioning 131
Jakub Lokoč, Tomáš Skopal

Efficient and Effective Similarity-based Video Retrieval 133
Ilaria Bartolini, Corrado Romani

Author Index 135

Foreword

The *International Conference on Similarity Search and Applications (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.

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. 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.

After the very successful first events in Cancun, Mexico (2008) and Prague, Czech Republic (2009), this year SISAP conference has been held in Istanbul, Turkey, on September 18-19, 2010.

Preface

This volume contains 2 invited papers, 11 regular papers, 2 posters and 6 demos, presented at SISAP 2010 conference, held in Istanbul, Turkey, on September 18-19, 2010. The papers were selected from 29 submissions, by a renowned program committee. We received submissions from all over the world (Argentina, Chile, China, Cuba, Czech Republic, Germany, Greece, Italy, Japan, Mexico, Norway, Spain, and USA). In this third volume of the SISAP series, the acceptance ratio of 66% was achieved as a result of the increasing quality of submissions established last year. The papers were selected based on their originality, relevance, and technical strength.

The submissions cover the spectrum of similarity searching topics from novel indexing techniques, to approximate methods for speeding up similarity search in metric/non-metric spaces, to applications in computerized biology, image classifications, and content-based image retrieval. In the demonstrations section a large number of interesting prototypes of similarity search engines, related to content- and semantics-based retrieval of multimedia data (audio, images, and video), was presented. Finally, the posters present an approach for balancing splits in indexing techniques based on generalized hyperplane partitioning and a system for automatic video annotation based on similarity among detected shots.

Two invited speakers gave excellent and thought-provoking talks. Prof. Vladimir Pestov provided interesting connections between similarity searching in metric spaces, the phenomenon of concentration of measure on high-dimensional structures, and the Vapnik-Chervonenkis theory of statistical learning, deriving bounds on performance of well-know indexing structures and showing how statistical learning methods can be used to build indexing schemes in datasets drawn from distributions of lower intrinsic dimension. Prof. Tomáš Skopal discussed the impact of the metric indexing paradigm on real-world applications, analyzing the world of similarity searching in metric spaces from the point of view of practitioners, transforming his conclusions into research challenges and perspectives of the similarity search methods. These contributions given by leading researchers definitely contributed to an in-depth and up-to-date understanding of fascinating fields and will surely stimulate further research.

Extended versions of the best SISAP 2010 papers were collected and will be published in a special issue of Journal of Discrete Algorithms (Elsevier).

We are grateful to the program committee members, the local organizers, and the external reviewers for their committed and high-quality work. We highly appreciate the work done by Karina Figueroa for maintaining the Metric Library, and those who contributed to it. Reviewing of papers and preparation of the proceedings were done using the EasyChair conference system (`www.easychair.org`), for which we specially thank Andrei Voronkov. Last but not least, we thank Adrienne Griscti (ACM program manager for SIG publications) for her hard work in helping us publishing these proceedings.

Paolo Ciaccia and **Marco Patella**
Istanbul, Turkey
September 19th, 2010

Conference Organization

Program Chairs

Paolo Ciaccia, *Università di Bologna*
Marco Patella, *Università di Bologna*

Program Committee

Edgar Chávez, *Universidad Michoacana - CICESE*
Alfredo Ferro, *Università degli Studi di Catania*
Daniel Keim, *Universität Konstanz*
Daniel Miranker, *University of Texas at Austin*
Gonzalo Navarro, *Universidad de Chile*
Hanan Samet, *University of Maryland*
Tomáš Skopal, *Univerzita Karlova v Praze*
Agma Juci Machado Traina, *Universidade de São Paulo*
Pavel Zezula, *Masarykova Univerzita*

Local Organization

Cengiz Celik, *Bilkent University*

Referees

Maria Camila Nardini Barioni
Benjamin Bustos
Geoff Ellis
Karina Figueroa
David Hoksza
Curran Kelleher
Milos Krstajic
Jakub Lokoc
Mauricio Marin
Jiri Novak
Daniela Oelke
Oscar Pedreira
Vladimir Pestov
Giuseppe Pigola
Ives Pola
Alfredo Pulvirenti
Smriti Ramakrishnan
Humberto Razente
Nora Reyes
Matthias Schaefer
Andreas Stoffel
Andrada Tatu
Marcos Vieira
Leishi Zhang