2009 Second International Workshop on Similarity Search and Applications

(SISAP 2009)

Prague, Czech Republic
29 - 30 August 2009
# 2009 Second International Workshop on Similarity Search and Applications

## SISAP 2009

### Table of Contents

- Foreword
- Preface
- Organizing and Program Committees
- External Reviewers

---

### Invited Papers

- Analyzing Metric Space Indexes: What For?  
  Gonzalo Navarro  
  Page 3

- Combinatorial Framework for Similarity Search  
  Yury Lifshits  
  Page 11

### Models of Similarity, Data Analysis and Transformations

- Structural Entropic Difference: A Bounded Distance Metric for Unordered Trees  
  Richard Connor, Fabio Simeoni, and Michael Iakovos  
  Page 21

- Efficient Similarity Search by Reducing I/O with Compressed Sketches  
  Arnoldo José Müller-Molina and Takeshi Shinohara  
  Page 30

- Curse of Dimensionality in Pivot Based Indexes  
  Ilya Volnyansky and Vladimir Pestov  
  Page 39

- CoPhIR Image Collection under the Microscope  
  Michal Batko, Petra Kohoutkova, and David Novak  
  Page 47

### Indexing Techniques

  Roberto Uribe Paredes and Gonzalo Navarro  
  Page 57

- Metric Index: An Efficient and Scalable Solution for Similarity Search  
  David Novak and Michal Batko  
  Page 65

- Optimal Pivots to Minimize the Index Size for Metric Access Methods  
  Luis G. Ares, Nieves R. Brisaboa, María F. Esteller, Oscar Pedreira, and Ángeles S. Places  
  Page 74

- Dynamic Spatial Approximation Trees for Massive Data  
  Gonzalo Navarro and Nora Reyes  
  Page 81
Similarity Queries, Query-Evaluation Algorithms
Approximate Direct and Reverse Nearest Neighbor Queries, and the k-nearest Neighbor Graph ................................................................. 91
Karina Figueroa and Rodrigo Paredes
Principles of Information Filtering in Metric Spaces .................................................. 99
Paolo Ciaccia and Marco Patella
Speeding Up Permutation Based Indexing with Indexing ........................................ 107
Karina Figueroa and Kimmo Frediksson

Distributed and Parallel Index Structures
Parallel Dynamic Batch Loading in the M-tree ......................................................... 117
Jakub Lokoč
Dynamic P2P Indexing and Search Based on Compact Clustering ........................ 124
Mauricio Marin, Veronica Gil-Costa, and Cecilia Hernandez
Query Routing Mechanisms in Self-Organizing Search Systems ................................ 132
Vlastislav Dohnal and Jan Sedmidubský

Demonstrations
OBSearch: A High Performance Similarity Search Engine for Java .......................... 143
Arnoldo José Müller-Molina
MiPai: Using the PP-Index to Build an Efficient and Scalable Similarity Search System ................................................................. 146
Andrea Esuli
Searching by Similarity and Classifying Images on a Very Large Scale ................ 149
Giuseppe Amato and Pasquale Savino
Img(Rummager): An Interactive Content Based Image Retrieval System ............... 151
Savvas A. Chatzichristofis, Yiannis S. Boutilis, and Mathias Lux
img(Anaktisi): A Web Content Based Image Retrieval System ............................... 154
Konstantinos Zagoris, Savvas A. Chatzichristofis, Nikos Papamarkos, and Yiannis S. Boutilis
Text-Based and Content-Based Image Retrieval on Flickr: DEMO ......................... 156
Juan Manuel Barrios, Diego Diaz-Espinoza, and Benjamin Bustos
MUFIN: A Multi-feature Indexing Network ............................................................. 158
Michal Batko, Vlastislav Dohnal, David Novak, and Jan Sedmidubsky

Posters
Using Tuneable Fuzzy Similarity in Non-metric Search .......................................... 163
Peter Vojtáš and Alan Eckhardt
On the Effectiveness of Distances Measuring Protein Structure Similarity ................ 165
Jakub Galgonek and David Hokzsa

Author Index