

Lecture Notes in Computer Science

16134

Founding Editors

Gerhard Goos

Juris Hartmanis

Editorial Board Members

Elisa Bertino, *Purdue University, West Lafayette, IN, USA*

Wen Gao, *Peking University, Beijing, China*

Bernhard Steffen , *TU Dortmund University, Dortmund, Germany*

Moti Yung , *Columbia University, New York, NY, USA*

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.

LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Giuseppe Amato · Vladimir Mic · Agma Traina ·
Nicola Messina · Laurent Amsaleg ·
Gylfi Þór Guðmundsson · Björn Þór Jónsson ·
Lucia Vadicamo
Editors

Similarity Search and Applications

18th International Conference, SISAP 2025
Reykjavik, Iceland, October 1–3, 2025
Proceedings

Editors

Giuseppe Amato 
CNR-ISTI
Pisa, Italy

Agma Traina 
University of Sao Paulo
Sao Carlos, Brazil

Laurent Amsaleg 
IRISA-CNRS
Rennes, France

Björn Þór Jónsson 
Reykjavik University
Reykjavik, Iceland

Vladimir Mic 
Aarhus University
Aarhus, Denmark

Nicola Messina 
CNR-ISTI
Pisa, Italy

Gylfi Þór Guðmundsson 
Reykjavik University
Reykjavík, Iceland

Lucia Vadicamo 
CNR-ISTI
Pisa, Italy

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Computer Science

ISBN 978-3-032-06068-6

ISBN 978-3-032-06069-3 (eBook)

<https://doi.org/10.1007/978-3-032-06069-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2026

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

This volume contains the papers presented at the 18th International Conference on Similarity Search and Applications (SISAP). The conference was held in Reykjavík, Iceland, from October 1 to 3, 2025. Reykjavik University hosted SISAP 2025, continuing the tradition of offering an annual forum to researchers and application developers in the area of similarity data management.

Efficient data management is more crucial than ever for applications that process large-scale data, as digital data continues to grow. Complex data domains such as images, geometric shapes, biometric data, and texts in natural languages do not allow for natural ordering of data items that would preserve pairwise similarities. The similarity, as perceived by humans, must be modeled and treated considering particular data items as reference objects. That is, processed data items can only be clustered or sorted based on their similarities to a given reference data item.

The contemporary explosion of machine learning techniques causes increasing demands on data processing. Contemporary data sets, therefore, require innovative approaches to similarity management and processing to meet the growing user expectations. SISAP is the main scientific conference that focuses on similarity management mostly considering (1) the similarity foundations, i.e., similarity learning, modeling, similarity functions, local intrinsic dimensionality, and distance densities; (2) data processing based on similarities of data items, i.e., data organization, searching, retrieval, indexing, filtering, and clustering; and (3) applications of similarity-based approaches in various fields such as multimedia processing, biometrics processing and searching, computer vision, big data processing, and data mining.

SISAP 2025 hosted the third SISAP Indexing Challenge, the fifth Doctoral Symposium on Similarity Search and Applications, a demonstration track, and a special session titled “IRMC: Interactive Retrieval for Multimedia Collections”. It also featured three distinguished keynote speakers, Kai Uwe Barthel (HTW Berlin, Germany), Franco Maria Nardini (CNR-ISTI, Italy), and Ari Jónsson (Videntifier Technologies, Iceland), who provided diverse theoretical, practical, and applied insights. The abstracts of their invited talks are included in these proceedings.

This year, SISAP received 47 submissions for the SISAP main track and special session, 6 submissions for the SISAP Indexing Challenge, 3 demonstration proposals, and 2 submissions for the Doctoral Symposium. The authors of these 58 articles were affiliated with institutions in 24 countries. These numbers represent a significant increase, with the highest values seen at SISAP since the beginning of the COVID-19 pandemic in 2020. In total, 29 regular and special session papers were accepted, of which 18 were both submitted and accepted as full papers, 2 were submitted and accepted as short papers, and 9 were submitted as full papers but accepted as short papers. The acceptance rate among the full regular and special session papers was 43% (18 out of 42), and an additional 21% of full articles were accepted as short ones.¹

¹ Two papers were subsequently withdrawn by their authors due to travel and registration issues.

The Indexing Challenge was organized by Eric S. Tellez (CICESE, Mexico), Edgar Chavez (CICESE, Mexico), Martin Aumüller (ITU Copenhagen, Denmark), and Vladimir Mic (Aarhus University, Denmark). It contained two different tasks to promote the dissemination of best practices throughout the community and allow an honest comparison of competitors in a predefined environment. Similarly to previous editions, the Indexing Challenge offered public datasets and allowed full reproducibility via GitHub actions. This year, it welcomed 7 participating teams. The authors submitted 6 explanatory articles, and 5 of them were accepted for publication in these proceedings. In addition, the proceedings feature an invited overview paper by the organizers, providing a detailed summary of the Indexing Challenge.

The Doctoral Symposium provided a forum for PhD students to present their research ideas and receive feedback from senior members of the research community. The goal of the Doctoral Symposium was to foster a collaborative environment and encourage students to participate in constructive discussions to share their ideas. SISAP 2025 received 2 submissions to the Doctoral Symposium, and both were accepted.

The Demonstration Track received 3 high-quality submissions, all of which were accepted. The evaluation took into account both academic aspects, such as novelty and technical challenges, and practical considerations, including how engaging and impactful the demonstrations would be for attendees during the conference. At SISAP 2025, these demonstrations were presented in a dedicated session, integrated with the conference reception, creating an interesting and engaging social atmosphere.

The IRMC special session on Interactive Retrieval for Multimedia Collections, organized by Omar Shahbaz Khan (Reykjavik University, Iceland), Rahel Arnold (University of Basel, Switzerland), Fabio Carrara (CNR-ISTI, Italy), and Ujjwal Sharma (University of Amsterdam, Netherlands), attracted 3 submissions. Two of these submissions were accepted and presented in a dedicated conference session, complemented by a lively panel discussion.

The SISAP 2025 Program Committee had 56 members from 20 countries. A high commitment to the conference is illustrated by the fact that all PC members provided all the reviews requested. Consequently, each regular paper received at least 3 reviews, and articles submitted to the special sessions received at least 4 reviews. For special session submissions, two of the four reviews were proposed by the special session organizers, and two were from the ordinary PC committee. All reviews of regular and special session papers were conducted in a double-blind manner. Additionally, reviewers were unaware of the identity of the other reviewers evaluating the same article. Doctoral Symposium, Demonstration, and Indexing Challenge papers, on the other hand, were reviewed in a single-blind fashion, as most appropriate for these categories of papers.

We express our deep appreciation to all members of the Program Committee for their commitment, their detailed interest in the submitted articles, and their support of the SISAP conference. We believe that the detailed and high-quality reviews of the articles maintained the high scientific and academic standards of the SISAP conference.

The proceedings of SISAP 2025 are published by Springer in the Lecture Notes in Computer Science (LNCS) series. The editors of these proceedings express their gratitude to Springer for continued support and for sponsoring the SISAP Awards. Following

tradition, authors of selected outstanding papers were invited to submit extended versions of their work for publication in a special issue of the Information Systems journal published by Elsevier.

We would like to thank all the authors who submitted papers to SISAP 2025. We also extend our gratitude to the members of the Organizing Committee for their work and pleasant collaboration, and to our sponsors and supporters for their generosity.

October 2025

Giuseppe Amato
Vladimir Mic
Agma J. M. Traina
Nicola Messina
Laurent Amsaleg
Gylfi Þór Guðmundsson
Björn Þór Jónsson
Lucia Vadicamo

Organization

General Chairs

Björn Þór Jónsson
Lucia Vadicamo

Reykjavik University, Iceland
CNR-ISTI, Italy

Program Committee Chairs

Giuseppe Amato
Vladimir Mic
Agma J. M. Traina

CNR-ISTI, Italy
Aarhus University, Denmark
University of São Paulo, Brazil

Demonstration Chair

Gylfi Þór Guðmundsson

Reykjavik University, Iceland

Doctoral Symposium Chair

Laurent Amsaleg

IRISA-CNRS, France

SISAP Indexing Challenge Chairs

Eric S. Tellez
Martin Aumüller
Vladimir Mic
Edgar Chávez

CICESE-INFOTEC-CONACyT, Mexico
ITU Copenhagen, Denmark
Aarhus University, Denmark
CICESE, México

Publication Chair

Nicola Messina

CNR-ISTI, Italy

Publicity Chairs

Luca Ciampi
Conrado Martinez

CNR-ISTI, Italy
Universitat Politècnica de Catalunya, Spain

Special Sessions Organizers

Omar Shahbaz Khan
Rahel Arnold
Fabio Carrara
Ujjwal Sharma
Dominik Bönisch

Reykjavik University, Iceland
University of Basel, Switzerland
CNR-ISTI, Italy
University of Amsterdam, Netherlands
University of Applied Sciences Düsseldorf,
Germany

Fabrizio Falchi
Peter Fornaro
Luca Rossetto
Heiko Schuldt

CNR-ISTI, Italy
University of Basel, Switzerland
Dublin City University, Ireland
University of Basel, Switzerland

Steering Committee

Giuseppe Amato
Edgar Chávez
Stéphane Marchand-Maillet
Marco Patella
Benjamin Kimia
Oscar Pedreira

CNR-ISTI, Italy
CICESE, México
University of Geneva, Switzerland
University of Bologna, Italy
Brown University, USA
University of A Coruña, Spain

Program Committee

Laurent Amsaleg
Fabrizio Angiulli
Lakshit Arora
Diego Arroyuelo
Werner Bailer
Kai Uwe Barthel
Ilaria Bartolini
Michal Batko
Marcos Bedo
Petra Budikova

CNRS-IRISA, France
University of Calabria, Italy
Google, USA
Pontificia Universidad Católica de Chile, Chile
JOANNEUM RESEARCH, Austria
HTW Berlin, Germany
University of Bologna, Italy
Masaryk University, Czech Republic
Fluminense Federal University, Brazil
Masaryk University, Czech Republic

Benjamin Bustos	University of Chile, Chile
K. Selçuk Candan	Arizona State University, USA
Mario G. C. A. Cimino	University of Pisa, Italy
Edgar Chávez	CICESE, Mexico
Richard Connor	University of St. Andrews, UK
Alan Dearle	University of St. Andrews, UK
Vlastislav Dohnal	Masaryk University, Czech Republic
Amalia Duch	Universitat Politècnica de Catalunya, Spain
Vladimir Estivill-Castro	Universitat Pompeu Fabra, Canada
Rolf Fagerberg	University of Southern Denmark, Denmark
Fabrizio Falchi	CNR-ISTI, Italy
Karina Figueroa	Universidad Michoacana de San Nicolás de Hidalgo, Mexico
Claudio Gennaro	CNR-ISTI, Italy
Magnus Lie Hetland	Norwegian University of Science and Technology, Norway
Michael Houle	New Jersey Institute of Technology, USA
Daniyal Kazempour	Christian-Albrechts-Universität zu Kiel, Germany
Daniel Keim	University of Konstanz, Germany
Benjamin Kimia	Brown University, USA
Jakub Lokoč	Charles University in Prague, Czech Republic
Maria Luisa Sapino	Università di Torino, Italy
Stéphane Marchand-Maillet	University of Geneva, Switzerland
Conrado Martínez	Universitat Politècnica de Catalunya, Spain
Vasileios Mezaris	Information Technologies Institute - Centre for Research and Technology Hellas, Greece
Lia Morra	Politecnico di Torino, Italy
Vincent Oria	New Jersey Institute of Technology, USA
Marco Patella	University of Bologna, Italy
Oscar Pedreira	Universidade da Coruña, Spain
Miloš Radovanović	University of Novi Sad, Serbia
Nora Reyes	Universidad Nacional de San Luis, Argentina
Marcela Ribeiro	Federal University of São Carlos, Brazil
Kunihiko Sadakane	University of Tokyo, Japan
Shin'ichi Satoh	National Institute of Informatics, Japan
Klaus Schöffmann	Klagenfurt University, Austria
Erich Schubert	TU Dortmund University, Germany
Heiko Schuldt	University of Basel, Switzerland
Jan Sedmidubský	Masaryk University, Czech Republic
Tetsuo Shibuya	University of Tokyo, Japan
Yasin N. Silva	Loyola University Chicago, USA
Tomáš Skopal	Charles University in Prague, Czech Republic

Eric S. Téllez
Caetano Traina
Goce Trajcevski
Stefanos Vrochidis

Pavel Zezula
Kaiping Zheng
Arthur Zimek

CONACyT - INFOTEC, Mexico
Universidade de São Paulo, Brazil
Iowa State University, USA
Information Technologies Institute - Centre for
Research and Technology Hellas, Greece
Masaryk University, Czech Republic
National University of Singapore, Singapore
University of Southern Denmark, Denmark

Additional Reviewers

Rahel Arnold
Fabio Carrara
Ben Claydon
Fabio Fassetti
Bastian Jäckl
Rodrigo Paredes
Muhammad Rajabinasab
Ujjwal Sharma
Ourania Theodosiadou

Athina Tsanousa

University of Basel, Switzerland
CNR-ISTI, Italy
University of St. Andrews, UK
University of Calabria, Italy
University of Konstanz, Germany
Universidad Finis Terrae, Chile
University of Southern Denmark, Denmark
University of Amsterdam, The Netherlands
Information Technologies Institute, CERTH,
Greece
Information Technologies Institute, CERTH,
Greece

Sponsors

Videntifier



Reykjavik University



CNR-ISTI



Springer



Information Systems, Elsevier



Keynotes

Scalable Methods for Visual Data Representation, Search, and Visualization

Kai Uwe Barthel

Institute for Media and Computing, HTW Berlin, Germany
kai-uwe.barthel@htw-berlin.de

Abstract. Large visual datasets are increasingly common, demanding efficient systems for retrieval and exploration. This talk will present recent methods for handling such data, including leveraging large visual encoders to generate compact, high-quality descriptors. It will cover strategies to enhance cross-modal retrieval and explore approximate nearest neighbor search techniques, focusing on graph-based methods for dynamic datasets. Visualization approaches for high-dimensional data, like grid-based sorting, will also be demonstrated for intuitive navigation. Grid sorting requires efficient methods for such large datasets. Therefore, we introduce a parameter-efficient approach to learning permutations, requiring only N parameters for N objects. This method iteratively refines permutations, offering scalability for applications such as image sorting, graph optimization, or combinatorial tasks like solving Sudokus.

Efficient Retrieval Over Learned Representations of Text: Challenges and Opportunities

Franco Maria Nardini

CNR-ISTI, Italy

`francomaria.nardini@isti.cnr.it`

Abstract. In recent years, transformer-based large language models (LLMs) have significantly reshaped how large textual collections are indexed and retrieved. A key development driving this transformation is the use of LLMs to learn high-dimensional, contextual representations—either dense or sparse—of input text. These learned representations enable the identification of semantically similar items through nearest neighbor search techniques. This paradigm has had a profound impact not only on core information retrieval tasks, but also on other areas of research such as recommender systems, where learned embeddings improve the modeling of user preferences and item relevance, and retrieval-augmented generation, where external knowledge retrieved via these representations is used to ground generative models. In this talk, we will review recent advancements that enable efficient indexing and retrieval based on such representations. We will discuss the current limitations and emerging challenges. We will then propose some research opportunities that arise from this very dynamic area of research.

Videntifier: Applications and Challenges of Scalable Similarity Search

Ari Jónsson

Videntifier Technologies, Iceland
ari@videntifier.com

Abstract. Recognizing images and videos automatically and efficiently is an important application of similarity search across a wide range of domains: Law enforcement uses it to recognize and investigate illegal content; Social media platforms use it to both keep out unwanted content and to manage monetization of user-generated content; Advertisers use it to track the use of visual advertising content by partners, online advertisers, influencers, competitors, etc; Artists, studios and other content creators use it to detect the use of their copyrighted content; and so on. In this talk, we present the technology developed by Videntifier to make visual recognition accurate, robust, fast, and scalable. We discuss a number of different application cases that highlight real-world use of similarity search in multiple markets and at a wide range of scales. Finally, we discuss some of the exciting opportunities and challenges that emerging technologies are bringing.

Contents

Applications

ViSketch-GPT: Collaborative Multi-scale Feature Extraction For Hand-Drawn Sketch Retrieval	3
<i>Giulio Federico, Fabio Carrara, Claudio Gennaro, and Marco Di Benedetto</i>	
FLAME_GAN: High-Fidelity 3D Face Reconstruction via Disentangled Representations and Explicit Geometric Control	14
<i>Grina Wiem and Douik Ali</i>	
SCAPE Semantically Context-Aware Password Generation Using Word Embeddings	29
<i>Nadine-Sarah Schöler, Maximilian von Zastrow, Tobias Vent, and Michael Eichberg</i>	

Foundations

Efficient Local Intrinsic Dimensionality Estimation in Evolving Deep Representations	41
<i>Michael E. Houle, Vincent Oria, and Hao Xu</i>	
CF-OOD: Concentration-Free Density Estimation for Reliable Out-of-Distribution Detection	56
<i>Fabrizio Angiulli, Fabio Fasseti, and Maria Pia Zupì</i>	
Intrinsic Dimension, Degrees of Freedom, Odds and Uniformity: A Unified Perspective	71
<i>James Bailey, Ricardo J. G. B. Campello, and Michael E. Houle</i>	
Assessing the Quality of Dimensionality Reduction Methods Based on Fuzzy Simplicial Sets	85
<i>Victor Reyes and Stephane Marchand-Maillet</i>	
Variance-Based Pivot Selection for Metric Spaces	94
<i>Yan Ruan, Detao Ji, Kun Luo, Yuhang Lou, Minhua Lu, and Rui Mao</i>	

Search and Retrieval

Dynamic Sub-region Search In Homogeneous Collections Using CLIP 105
Bastian Jäckl, Vojtěch Kloda, Daniel A. Keim, and Jakub Lokoč

Lightweight Semantic Search For Low-Resource Languages: A Case
Study In Vietnamese Information Retrieval 119
*Chi Mai Nguyen, Hoang Anh Phi Tran, Phat Trien Thai,
Huu Doanh Nguyen, and Van Tuan Nguyen*

On the Theoretical Advantages of Bilinear Similarities in Dense Retrieval 132
Shubham Chatterjee

Experimental Evaluation Of Static Image Sub-Region-Based Search
Models Using CLIP 140
Bastian Jäckl, Vojtěch Kloda, Daniel A. Keim, and Jakub Lokoč

Integrating Relevance Feedback for Effective Personalisation in Vector
Search 154
Matúš Šikyňa and Pavel Zezula

Similarity Measures

Similarity Based on Resample Exposure 165
*Anton Danholt Lautrup, Hafiz Saud Arshad, Tobias Hyrup,
Muhammad Rajabinasab, Arthur Zimek, and Peter Schneider-Kamp*

Metric Matters: A Formal Evaluation of Similarity Measures in Active
Learning for Cyber Threat Intelligence 180
Sidahmed Benabderrahmane and Talal Rawhan

Quantification of Similarities: Embeddings Into Euclidean Spaces
and Taxonomies 195
Evgeny Dantsin and Alexander Wolpert

PUFFME: Probabilistic Unkeyed Feature Fusion for Matching Entities 209
Andrea Leoni, Andrea Molinari, Filippo Costamagna, and Simone Sandri

Relation Similarity Based Thought Chain Extraction Mechanism
from Value Belief Network 223
JeongYon Shim

Clustering

Approximate Single-Linkage Clustering Using Graph-Based Indexes: MST-Based Approaches and Incremental Searchers	233
<i>Camilla Birch Okkels, Erik Thordsen, Martin Aumüller, Arthur Zimek, and Erich Schubert</i>	
Fast and Accurate Estimates for External Clustering Validation Measures	248
<i>Hugo Sanz-González and Conrado Martínez</i>	
Hierarchical Density-Based Clustering Using Incremental Similarity Search	263
<i>Erich Schubert</i>	

Indexing

Theoretical and Practical Insights Into Graph-Based Indexing	273
<i>Erik Thordsen and Erich Schubert</i>	
Equi-Voronoi Polytopes: A Geometric Basis for 2-bit Quantisation	288
<i>Richard Connor, Alan Dearle, and Ben Claydon</i>	
The Curious Case of High-Dimensional Indexing as a File Structure: A Case Study of eCP-FS	303
<i>Omar Shahbaz Khan, Gylfi Þór Guðmundsson, and Björn Þór Jónsson</i>	

Learning-Based Methods

Dynamic Graph Embedding Through Hub-Aware Random Walks	315
<i>Aleksandar Tomčić, Miloš Savić, Dušan Simić, and Miloš Radovanović</i>	
Towards Semi-supervised Subspace Learning for Outlier Detection in Big Data	330
<i>Muhammad Rajabinasab, Anton D. Lautrup, Peter Schneider-Kamp, and Arthur Zimek</i>	
Learning to Find Good Hash Functions for Embeddings	345
<i>Ben Claydon, Richard Connor, and Alan Dearle</i>	

Doctoral Symposium

Explaining Hubness by the Expected k -Occurrences	357
<i>Victor Reyes</i>	

Combining Dissimilarity Spaces to Improve ANN Search 364
Elena García-Morato

Demonstration Papers

A Comparative Demonstration of Relevance Feedback Methods for Image
Retrieval 375
Francesca Scotti, Lucia Vadicamo, Giuseppe Amato, and Fabio Carrara

VISAnt: Unsupervised Data Exploration with Chernoff Faces 384
*Ivana Sixtova, Ladislav Peska, Jakub Lokoč, David Bernhauer,
and Tomas Skopal*

Eddy Hunter: A Data Mining System for High-Resolution Eddy Signals,
Leveraging Spatio-Temporal Similarities in the SWOT Satellite Data 392
Federico Scarscelli, Claudius Zelenka, Peer Kröger, and Florian Schütte

SISAP Indexing Challenge

Overview of the SISAP 2025 Indexing Challenge 403
Eric S. Tellez, Edgar Chavez, Martin Aumüller, and Vladimir Mic

Memory-Constrained DiskANN: Efficient Approximate Nearest Neighbor
Search Under Resource Constraints 415
*Yuhang Lou, Linyun Ma, Kun Luo, Yan Ruan, Huijia Wu, Minhua Lu,
and Rui Mao*

Double Filtering Using Short and Long Quantized Projections 424
*Naoya Higuchi, Yasunobu Imamura, Takeshi Shinohara,
Kouichi Hirata, and Tetsuji Kuboyama*

SISAP Indexing Challenge 2025 – Solution for Task 2 Using Root Join 433
Benjamin Bustos and Jiale Chen

Fast, Compact NN-Table Build Using Equi-Voronoi Polytopes 441
Alan Dearle, Richard Connor, Ben Claydon, and Ferdia McKeogh

Refinement-Based Graph Construction for Search in Low-Memory
Systems 450
Cole Foster, Scarlett Magdaleno-Gatica, and Benjamin Kimia

Author Index 459