

## Introductory Remarks to the Special Issue Devoted to DAMDID/RCDL-2023

This special issue of *Automation and Remote Control* contains some scientific results presented at the 25th International Conference on Data Analytics and Management in Data-Intensive Domains (DAMDID/RCDL-2023). The conference was held on October 24–27, 2023, at National Research University Higher School of Economics (HSE) in Moscow, Russia. The reader is offered the full texts of selected DAMDID/RCDL-2023 papers relevant to the journal's topics.

The DAMDID conference was formed in 2015 by transforming the RCDL conference (Digital Libraries: Advanced Methods and Technologies, Digital Collections) to create a forum reflecting the urgent challenges of data analysis and management during research in various data-intensive domains. The transformation was due to expanding the scope of electronic libraries, in particular, with their active use in scientific research and the current demand for technologies for the efficient storage, processing, and analysis of scientific data as well as for the integration of heterogeneous data from multiple sources. The transformed conference preserved the continuity with RCDL and, moreover, inherited the RCDL community established during the sixteen years of its successful work.

The DAMDID conference is known as a multidisciplinary forum of researchers and practitioners from various fields of science and industry that promotes cooperation and exchange of ideas and methods in the area of data analysis and management being developed in specific data-intensive domains of  $X$ -informatics, where  $X \in$  astro, bio, geo, neuro, medical, physical, chemical, material science, socio, business, financial, ....

The conference topics include various tracks for data analysis, problem solving, and experiment organization:

- problem statement and solving in data-intensive domains (methods, processes, and tools);
- organization of experiments (theories, hypotheses, models, simulations, infrastructures, and workflows);
- advanced data-intensive analysis methods and procedures (exploratory analysis, statistics and machine learning, meta-analysis, and efficiency and scalability of methods);
- conceptual modeling of subject domains and knowledge representation (semantics, ontologies, database schemas, conceptualization of algorithms and workflows, and semantic interoperability);
- research support in data infrastructures, data-intensive use cases (the functions and architectures of virtual laboratories and observatories, data sharing in interdisciplinary research).

Also, the conference program traditionally includes tracks for data management:

- methods, tools and infrastructures for data acquisition, storage, and processing (data and metadata: semantics, quality and provenance, cleansing, and anomaly detection);
- data integration (data models, schemas, entity resolution and data merging, virtual and material integration, data warehouses, ETL processes, multidimensional data models, and multi-structured data);
- information extraction from observational data (complete and up-to-date information from data in astronomy, spectroscopy, material science, medicine, etc.);

- information extraction from texts;
- research data infrastructures and their applications (clouds, grids, distributed clusters, supercomputers; implementation, scaling, and performance assessment of infrastructures; new programming models and virtualization);
- semantic Web role in data-intensive domains.

The program included three keynote and three invited talks: four scientific and two industrial. The papers were distributed into the following sections and areas:

- algorithms for solving problems;
- data analysis in medicine and cognitive sciences;
- image analysis;
- social network analysis;
- data analysis in astronomy;
- databases in material science;
- extracting information from texts;
- artificial intelligence and machine learning in material science;
- conceptual models, ontologies and semantic Web;
- machine learning methods and their applications;
- modeling in geosciences.

The Program Committee of the conference, composed of researchers from the UK, Germany, India, Italy, Spain, China, Latvia, Russia, Serbia, USA, and Japan, considered 79 submissions. After the two-round single-blind peer-reviewing, including discussion of the review results by members of the Program Committee, 53 submissions were accepted as full papers and 15 as short papers whereas 11 submissions were rejected or withdrawn. Over 120 participants attended the conference. Three plenary sessions and 18 sectional sessions were held with six plenary talks and 61 sectional papers. The participants represented research centers and universities from 11 regions of Russia (Kazan, Moscow, Murom, Novosibirsk, Obninsk, Samara, St. Petersburg, Tomsk, Tyumen, and Chelyabinsk) as well as foreign countries (Armenia, UK, Germany, India, Iran, Canada, China, Pakistan, UAE, USA, and Japan).

As in previous years, the papers not included in special journal issues will be published in Springer's Communications in Computer and Information Science. Based on the recommendations of the Program Committee and the review results, eight papers were selected for publication in this special issue in accordance with journal's topics and the authors' preferences:

- "Self-Adjusted Consensus Clustering with Agglomerate Algorithms" (B.G. Mirkin and A.A. Parinov);
- "Genetic Engineering Algorithm (GEA): An Efficient Metaheuristic Algorithm for Solving Combinatorial Optimization Problems" (M. Sohrabi, A. Fathollahi-Fard, and V.A. Gromov);
- "Attacks on Machine Learning Models Based on the PyTorch Framework" (D.E. Namiot and T.M. Bidzhiev);
- "Interestingness Indices for Building Neural Networks Based on Concept Lattices" (M.M. Zueva and S.O. Kuznetsov);
- "On the Use of Digital Twin Data in Models Related to Considering the Environment Impact on Enterprises" (E.D. Viazilov, D.A. Melnikov, and O.A. Minkov);
- "Plausible Reasoning in an Algorithm for Generation of Good Classification Tests" (X.A. Naidenova, V.A. Parkhomenko, T.A. Martirova, and A.V. Schukin);
- "Transformer-Based Classification of User Queries for Medical Consultancy" (D.A. Lyutkin, D.V. Pozdnyakov, A.A. Soloviev, D.V. Zhukov, M.S.I. Malik, and D.I. Ignatov);

- “Smart Routes: A System for Development and Comparison of Algorithms for Solving Vehicle Routing Problems with Realistic Constraints” (A.G. Soroka, G.V. Mikhelson, A.V. Mescheryakov, and S.V. Gerasimov).

These papers focus on machine learning and data mining methods, plausible reasoning and neural networks, control and optimization algorithms, and text mining and digital twin modeling.

*J. Baixeries, Barcelona*

*D.I. Ignatov, Moscow*

*S.O. Kuznetsov, Moscow*

*S.A. Stupnikov, Moscow*