Preface

This volume contains refereed proceedings of the Sixth International conference on Analysis of Images, Social Networks, and Texts (AIST 2017)¹. The previous conferences during 2012–2016 attracted a significant number of students, researchers, academics, and engineers working on interdisciplinary data analysis of images, texts, and social networks.

The broad scope of AIST made it an event where researchers from different domains, such as image and text processing, exploiting various data analysis techniques, can meet and exchange ideas. We strongly believe that this may lead to cross fertilisation of ideas between researchers relying on modern data analysis machinery. Therefore, AIST brought together all kinds of applications of data mining and machine learning techniques. The conference allowed specialists from different fields to meet each other, present their work, and discuss both theoretical and practical aspects of their data analysis problems. Another important aim of the conference was to stimulate scientists and people from industry to benefit from the knowledge exchange and identify possible grounds for fruitful collaboration.

The conference was held during July 27–29, 2017. The conference was organised in Moscow, the capital of Russia, at the campus of Moscow Polytechnic University². This year, the key topics of AIST were grouped into six tracks:

- 1. General topics of data analysis chaired by Sergei Kuznetsov (Higher School of Economics, Russia) and Amedeo Napoli (LORIA, France)
- 2. Natural language processing chaired by Natalia Loukachevitch (Lomonosov Moscow State University, Russia) and Alexander Panchenko (University of Hamburg, Germany)
- 3. Social network analysis chaired by Stanley Wasserman (Indiana University, USA)
- 4. Analysis of images and video chaired by Victor Lempitsky (Skolkovo Institute of Science and Technology, Russia) and Andrey Savchenko (Higher School of Economics, Russia)
- Optimization problems on graphs and network structures chaired by Panos Pardalos (University of Florida, USA) and Mikhail Khachay (IMM UB RAS & Ural Federal University, Russia)
- 6. Analysis of dynamic behavior through event data Wil van der Aalst (Eindhoven University of Technology, The Netherlands) and Irina Lomazova (Higher School of Economics, Russia).

One of the novelties this year was the introduction of a new specialised track on Process Mining (Track 6).

¹ http://aistconf.org/

² http://mospolytech.ru/?eng

The Program Committee and the reviewers of the conference included 167 well-known experts in data mining and machine learning, natural language processing, image processing, social network analysis, and related areas from leading institutions of 30 countries including Argentina, Australia, Austria, Belgium, Brazil, Canada, China, Croatia, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Hungary, India, Ireland, Japan, Lithuania, Norway, Portugal, Qatar, Romania, Russia, Spain, Ukraine, The Netherlands, UK, and USA. This year we have received 127 submissions mostly from Russia but also from Algeria, Australia, Brazil, China, Finland, Germany, India, Iran, Kazakhstan, Latvia, Mexico, Netherlands, Norway, Turkey, USA, and Viet Nam.

Out of 127 submissions only 37 papers were accepted as regular oral papers. Thus, the acceptance rate of this volume was around 29%. In order to encourage young practitioners and researchers, we included 34 papers to the supplementary proceedings after their poster presentation at the conference. Each submission was reviewed by at least three reviewers, experts in their fields, in order to supply detailed and helpful comments.

The conference featured several invited talks and an industry session dedicated to current trends and challenges.

The keynote talk was presented by Andrzej Cichocki on "Bridge Between Tensor Networks and Deep Neural Networks: From Fundamentals to Real Applications".

The invited talks were:

- Stanley Wasserman (Indiana University), "Sensitivity analysis of p* and SAOM: The effects of missing data on parameter estimates".
- Dirk Fahland (TU Eindhoven, Netherlands), "Process Mining: Past, Present, and Open Challenges".

Sergey Nikolenko from the St. Petersburg Department of the Steklov Mathematical Institute presented a tutorial on "Deep Learning for Natural Language Processing".

The business speakers also covered a wide variety of topics³. We list those invited talks below:

- Iosif Itkin (Exactpro Systems), Industrial day opening. Keynote on "Exactpro and Data Analysis in London Stock Exchange: Capital Markets, Post Trade and Information Services".
- Dmitry Bugaychenko (OK.ru), "Odnoklassniki Data Science Research Initiative".
- Olga Megorskaya (Yandex), "Yandex Toloka: Crowdsource Your Data".
- Artur Kuzin (Avito), "AvitoNet: computer vision service in Avito".
- Alexander Zhebrak, Arthur Kadurin, Daniil Polykovskiy (Insilico Medicine),
 "Artificial Intelligence for Drug Discovery".

³ the detailed program of AIST 2017 Business Day can be found at a separate website: http://aistconf.ru

We would like to thank the authors for submitting their papers and the members of the Program Committee for their efforts in providing exhaustive reviews. We would also like to express special gratitude to all the invited speakers and industry representatives.

We deeply thank all the partners and sponsors. Our golden sponsors is Exactpro. Exactpro, a fully owned subsidiary of the London Stock Exchange Group,
specialises in quality assurance for exchanges, investment banks, brokers, and
other financial sector organisations. Our special thanks goes to Springer for their
help, starting from the first conference call to the final version of the proceedings. Last but not least, we are grateful to all organisers, especially to Marina
Danshina, and the volunteers, whose endless energy saved us at the most critical
stages of the conference preparation.

Traditionally, we would like to mention the Russian word "aist" is more than just a simple abbreviation (in Cyrillic) - it means a "stork." Since it is a wonderful free bird, a symbol of happiness and peace, this stork brought us the inspiration to organise the AIST conference. So we believe that this young and rapidly growing conference will likewise be bringing inspiration to data scientists around the world!

October, 2017

Wil van der Aalst
Michael Khachay
Sergei Kuznetsov
Victor Lempitsky
Irina Lomazova
Natalia Loukachevitch
Amedeo Napoli
Alexander Panchenko
Panos Pardalos
Andrey Savchenko
Stanley Wasserman
Dmitry Ignatov

Organisation

Program Committee Chairs

Wil van der Aalst Technische Universiteit Eindhoven, The Netherlands

Mikhail Khachay Krasovskii Institute of Mathematics and Mechanics of RAS, Russia & Ural Federal University, Yekaterinburg,

Russia

Sergei Kuznetsov National Research University Higher School of Eco-

nomics, Moscow, Russia

Amedeo Napoli LORIA CNRS, University of Lorraine, and INRIA,

Nancy, France

Victor Lempitsky Skolkovo Institute of Science and Technology, Russia Irina Lomazova National Research University Higher School of Eco-

nomics, Moscow, Russia

Natalia Loukachevitch Computing Centre of Lomonosov Moscow State Univer-

sity, Russia

Alexander Panchenko Hamburg University, Germany & Université catholique

de Louvain, Belgium

Panos Pardalos University of Florida, US

Andrey Savchenko National Research University Higher School of Eco-

nomics, Nizhny Novgorod, Russia

Stanley Wasserman Indiana University, US

Proceedings Chair

Dmitry Ignatov National Research University Higher School of Eco-

nomics, Russia

Business Day Chair

Rostislav Yavorskiy National Research University Higher School of Eco-

nomics, Moscow

Steering Committee

Dmitry Ignatov National Research University Higher School of Eco-

nomics, Russia

Mikhail Khachay Krasovskii Institute of Mathematics and Mechanics of

RAS, Russia & Ural Federal University, Yekaterinburg,

Russia

Alexander Panchenko University of Hamburg, Germany

Rostislav Yavorskiy National Research University Higher School of Eco-

nomics, Russia

Program Committee

Mehwish Alam Université Paris 13, France Gabriela Arevalo Universidad Austral, Argentina

Artem Babenko Yandex, Russia

Jaume Baixeries Universitat Politècnica de Catalunya, Spain

Artem Baklanov International Institute for Applied Systems Analy-

sis, Austria

Sergey Bartunov National Research University Higher School of Eco-

nomics, Moscow, Russia & DeepMind, UK

Timo Baumann Universität Hamburg, Germany Darina Benikova University of Duisburg-Essen, Germany

Malay Bhattacharyya Indian Institute of Engineering Science and Technol-

ogy, India

Chris Biemann Technische Universität Darmstadt, Germany

Elena Bolshakova Moscow State Lomonosov University, Russia & Na-

tional Research University Higher School of Eco-

nomics, Russia

Anastasia Bonch- National Research University Higher School of Eco-

Osmolovskaya nomics, Russia

Aurélien Bossard Université Paris 8, France

Jean-Leon Bouraoui CENTAL (Université Catholique de Louvain), Bel-

gium

Joos Buijs Eindhoven University of Technology, The Nether-

lands

Andrea Burattin Technical University of Denmark, Denmark

Evgeny Burnaev Institute for Information Transmission Problems of

RAS, Russia

Aleksey Buzmakov INRIA-LORIA(CNRS-Université de Lorraine),

Nancy, France & National Research University

Higher School of Economics, Perm, Russia

Ignacio Cassol Universidad Austral, Argentina

Artem Chernodub Institute of Mathematical Machines and Systems of

NASU, Ukraine

Vladimir Chernov Institute for Image Processing of RAS, Russia Ekaterina Chernyak National Research University Higher School of Eco-

nomics, Moscow, Russia

Marina Chicheva Samara State Aerospace University, Russia Bonaventura Coppola Technische Universität Darmstadt, Germany

Hernani Costa University of Malaga, Spain

Massimiliano de Leoni Eindhoven University of Technology, The Nether-

lands

Boris Dobrov Lomonosov Moscow State University, Russia

Sofia Dokuka National Research University Higher School of Eco-

nomics, Moscow, Russia

Florent Domenach Akita International University, Japan

Alexey Drutsa Lomonosov Moscow State University & Yandex,

Russia

Mirela-Stefania Duma University of Hamburg, Germany

Richard Eckart de Castilho Technische Universität Darmstadt, Germany

Judith Eckle-Kohler UKP Lab, Technische Universität Darmstadt, Ger-

nany

Maria Eskevich Radboud University Nijmegen, The Netherlands Dirk Fahland Technische Universiteit Eindhoven, The Netherlands

Stefano Faralli University of Mannheim, Germany

Victor Fedoseev Samara National Research University, Russia Michael Figurnov Skolkovo Institute of Science and Technology, Russia

Elena Filatova City University of New York, US

Kerstin Fischer University of Southern Denmark, Denmark

Fedor Fomin University of Bergen, Norway

Thomas François Université catholique de Louvain, Belgium

Oleksandr Frei Universitetet i Oslo, Norway

Edward Kh. Gimadi Sobolev Institute of Mathematics of RAS, Russia Ivan Gostev National Research University Higher School of Eco-

nomics, Moscow, Russia

Natalia Grabar Université Lille 3 & CNRS, France

Dmitry Granovsky Yandex, Russia Alexey Gruzdev Intel, Russia

Ivan Habernal Technische Universität Darmstadt, Germany Mena Habib Maastricht University, The Netherlands Marianne Huchard Université Montpellier 2 & CNRS, France

Dmitry Ignatov National Research University Higher School of Eco-

nomics, Moscow, Russia

Dmitry Ilvovsky National Research University Higher School of Eco-

nomics, Moscow, Russia

Vladimir Ivanov Innopolis University, Russia

Pei Jun Hefei University of Technology, China

Anna Kalenkova National Research University Higher School of Eco-

nomics, Moscow, Russia

Nikolay Karpov National Research University Higher School of Eco-

nomics, Nizhniy Novgorod, Russia

Egor Kashkin V. V. Vinogradov Russian Language Institute of

RAS, Russia

Mehdi Kaytoue LIRIS - INSA de Lyon, France

Alexander Kelmanov Sobolev Institute of Mathematics of RAS, Russia Oleg Khamisov Melentiev Institute of Energy Systems of RAS, Russia

sia

Andrey Kibzun Moscow Aviation Institute, Russia

Edward Klyshinsky the HSE Moscow Institute of Electronics and Math-

ematics, Russia

Yury Kochetov Sobolev Institute of Mathematics of RAS, Russia

Ekaterina Kochmar University of Cambridge, UK

Sergei Koltcov National Research University Higher School of Eco-

nomics, St. Petersburg, Russia

Olessia Koltsova National Research University Higher School of Eco-

nomics, St. Petersburg, Russia

Jan Konecny Palacky University, Czech Republic

Daniil Kononenko Skolkovo Institute of Science and Technology, Russia

Natalia Konstantinova University of Wolverhampton, UK Andrey Kopylov Tula State University, Russia Mikhail Korobov ScrapingHub Inc., Ireland

Anton Korshunov Institute for System Programming of RAS, Russia

Evgeny Kotelnikov Vyatka State University, Russia Ilias Kotsireas Wilfrid Laurier University, Canada

Olga Krasotkina Lomonosov Moscow State University, Russia Tomas Krilavicius Vytautas Magnus University, Lithuania

Victor Kulikov Institute of Automation and Electrometry of RAS,

Russia

Valentina Kuskova National Research University Higher School of Eco-

nomics, Moscow, Russia

Andrey Kutuzov Universitetet i Oslo, Norway

Andrey Kuzmin Skolkovo Institute of Science and Technology, Russia

Andrey Kuznetsov Samara State Aerospace University, Russia

Sergei Kuznetsov National Research University Higher School of Eco-

nomics, Moscow, Russia

Alexander Lazarev Institute of Control Sciences of RAS, Russia

Florence Le Ber Université de Strasbourg, France

Vadim Lebedev Skolkovo Institute of Science and Technology, Russia Victor Lempitsky Skolkovo Institute of Science and Technology, Russia Alexander Lepskiy National Research University Higher School of Eco-

nomics, Moscow, Russia

Benjamin Lind Anglo-American School of St. Petersburg, Russia Irina Lomazova National Research University Higher School of Eco-

nomics, Moscow, Russia

Natalia Loukachevitch Lomonosov Moscow State University, Russia

Olga Lyashevskaya National Research University Higher School of Eco-

nomics, Moscow, Russia

Yury Malkov Institute of Applied Physics of RAS, Russia

Luis Marujo Carnegie Mellon University, US & Instituto Superior

Técnico, Portugal

Sérgio Matos Universidade de Aveiro, Portugal

Yelena Mejova Qatar Computing Research Institute, Qatar Nizar Messai Université François Rabelais Tours, France

Tristan Miller Technische Universität Darmstadt, Germany
Olga Mitrofanova St. Petersburg State University, Russia
Evgeny Myasnikov Samara National Research University, Russia
Sergey Nikolenko Steklov Mathematical Institute, St. Petersburg,

Russia

Vassilina Nikoulina Xerox Research Center Europe, France

Damien Nouvel Université Sorbonne, France

Dimitri Nowicki Institute of Mathematical Machines and Systems of

NASU, Ukraine

Panos Pardalos University of Florida, US

Georgios Petasis National Centre for Scientific Research "Demokri-

tos", Greece

Stefan Pickl Universität der Bundeswehr München, Germany

Lidia Pivovarova University of Helsinki, Finland

Vladimir Pleshko RCO, Russia

Hernan Ponce-De-Leon fortiss GmbH, Germany

Alexander Porshnev National Research University Higher School of Eco-

nomics, Nizhniy Novgorod, Russia

Alexey Potapov AIDEUS, Russia

Surya Prasath University of Missouri-Columbia, US

Uta Priss Ostfalia University of Applied Sciences, Germany

Oleg Prokopyev University of Pittsburgh, US

Artem Pyatkin Novosibirsk State University & Sobolev Institute of

Mathematics, Russia

Carlos Ramisch Aix Marseille University, France

Alexandr Rassadin KPMG, Russia & National Research University

Higher School of Economics, Nizhniy Novgorod,

Russia

Artem Revenko Semantic Web Company GmbH, Austria

Evgeniy Riabenko National Research University Higher School of Eco-

nomics, Moscow, Russia

Martin Riedl University of Hamburg, Germany
Alexey Romanov University of Massachusetts Lowell, US

Andrey Ronzhin St. Petersburg Institute for Informatics and Automa-

tion of Russian Academy of Sciences, Russia

Alexandra Roshchina Institute of Technology Tallaght, Ireland Eugen Ruppert Technische Universität Darmstadt, Germany

Christian Sacarea Babes-Bolyai University, Hungary Mohammed Abdel-Mgeed Ain Shams University, Cairo

M. Salem

Sheikh Muhammad Sarwar University of Massachusetts Amherst, US

Andrey Savchenko National Research University Higher School of Eco-

nomics, Nizhniy Novgorod, Russia

Friedhelm Schwenker Ulm University, Germany

Alexander Semenov National Research University Higher School of Eco-

nomics, Moscow, Russia

Oleg Seredin Tula State University, Russia Andrey Shcherbakov University of Melbourne, Australia

Oleg Slavin Institute for Systems Analysis of Russian Academy

of Sciences

Jan Snajder University of Zagreb, Croatia Henry Soldano Université Paris 13, France Tobias Staron Universität Hamburg, Germany

Dmitry Stepanov Program System Institute of RAS, Russia Vadim Strijov Computing Center of RAS, Russia Maria Sukhareva Goethe University Frankfurt, Germany

Diana Sungatullina Skolkovo Institute of Science and Technology, Russia

Laszlo Szathmary University of Debrecen, Hungary

Irina Temnikova Qatar Computing Research Institute, Qatar

Diana Troanca Babes-Bolyai University, Hungary Christos Tryfonopoulos University of Peloponnese, Greece

Denis Turdakov Institute for System Programming of RAS, Russia
Dmitry Ulyanov Skolkovo Institute of Science and Technology, Russia
Dmitry Ustalov Krasovskii Institute of Mathematics and Mechanics
of RAS, Russia & Ural Federal University, Yekater-

inburg, Russia

Evgeniya Ustinova Skolkovo Institute of Science and Technology, Russia

Alexander Vakhitov St.Petersburg State University, Russia

Wil van der Aalst Eindhoven University of Technology, The Nether-

lands

Natalia Vassilieva Hewlett Packard Enterprise, US

Dmitry Vetrov Moscow State University & National Research Uni-

versity Higher School of Economics, Moscow, Russia

Renato Vimieiro Universidade Federal de Pernambuco Ekaterina Vylomova The University of Melbourne, Australia

Roman Yangarber University of Helsinki, Finland

Rostislav Yavorsky National Research University Higher School of Eco-

nomics, Moscow, Russia

Marcos Zampieri University of Wolverhampton, UK Nikolai Zolotykh University of Nizhniy Novgorod, Russia

Olga Zvereva Ural Federal University, Russia

Additional Reviewers

Sujoy Chatterjee Vladimir Khandeev
Anton Eremeev Sofya Kulikova
Aleksey Glebov Abhishek Kumar
Alexei Glebov Alexander Plyasunov
Sergey Khamidullin Vladimir Servakh

Organising Committee

Rostislav Yavorskiy, Conference Chair National Research University Higher

School of Economics, Russia

Andrey Novikov, Head of Organization National Research University Higher

School of Economics, Russia

Marina Danshina, Venue Organization Moscow Polytechnic University, Russia

& Management

Anna Ukhanaeva, Information Partners National Research University Higher

and Communications School of Economics, Russia

Anna Kalenkova, Visa Support and National Research University Higher

International Communications School of Economics, Russia

Alexander Gnevshev, Venue Moscow Polytechnic University, Russia

Organization & Management

Volunteers

Daniil Bannyh Moscow Polytechnic University, Russia Ksenia Belkova Moscow Polytechnic University, Russia Tatiana Mishina Moscow Polytechnic University, Russia Zahar Kuhtenkov Moscow Polytechnic University, Russia

Maxim Pasynkov Krasovsky Institute of Mathematics and Mechanics

of RAS, Russia, Yekaterinburg

Ivan Poylov Moscow Polytechnic University, Russia

Sponsors

Golden sponsor: Exactpro Bronze sponsor: Springer

Table of Contents

Statistical Description of Russian Texts: Parameters and Factors	1
Baselines and Symbol N-Grams: Simple Part-Of-Speech Tagging of Russian	9
A Uniform Morphological Analyzer for the Kazakh and Turkish Languages Gulmira Bekmanova, Altynbek Sharipbay, Gulila Altenbek, Esref Adali, Lena Zhetkenbay, Unzila Kamanur and Altanbek Zulkhazhav	20
Extraction of Named Entities from Semi-Structured Texts for Medical Domain	31
Natalia Zhukova, Maksim Berezov, Sergey Lebedev and Ekaterina Zavad- skaya	91
Architecture and Implementation of an Intelligent News Analysis System Yury Gapanyuk, Sergey Chernobrovkin, Igor Latkin, Aleksey Leontiev, Grigory Ozhegov, Alexander Opryshko and Maxim Myalkin	41
Automatic Compilation of Person's Information Portraits as an Instrument of Historical Research	56
Optimization Task in Equivalent to word2vec Matrix Factorization \dots <i>Victor Kantor</i>	63
Development of a Model to Predict Intention Using Deep Learning Nikolay Karpov, Alexander Demidovskij and Alexey Malafeev	69
Comparison of Vector Space Representations of Documents for the Task of Matching Contents of Educational Course Programmes Julius Klenin and Dmitry Botov	79
Applicability of Automatically Generated Thesauri to Text Classification in Specific Domains	91
Ontology-Based Approach to Academic Style Marker Identification Viacheslav Lanin and Sofia Philipson	99

Author Gender Prediction in Russian Social Media Texts	105
Generalizable Architecture for Robust Word Vectors Tested by Noisy Paraphrases	111
Automatization of Scientific Articles Classification According to UDC Aleksandr Romanov, Konstantin Lomotin and Ekaterina Kozlova	122
Linguistic Processor Based on Object-Attribute Grammar Sergey Salibekyan and Petr Panfilov	134
Active Learning Strategy for Text Categorization Based on Support Vectors Relative Positioning	146
Artificial Intelligence User Support System for SAP Enterprise Resource Planning	156
Stance Detection in Russian: a Feature Selection and Machine Learning Based Approach	166
Bank Licenses Revocation Modeling	180
Formation of Coalition Structures as a Non-Cooperative Game	188
Adaptation Algorithm of the Computer Text Font Size for Optimal Perception	198
Multichannel Image Segmentation Algorithm for Agricultural Crop Detection	209
An Algorithm for Image Time Series Forgery Detection Based on the Anomalies Detection	217
Feature Engineering for Tree Leaf Classifier	224

Learning to Play Pong Video Game via Deep Reinforcement Learning Ilya Makarov, Andrej Kashin and Alisa Korinevskaya	236
Fast Multigrid Pattern Search for Motion Estimation in Hybrid Compression Systems	242
A Method for Describing and Recognizing Fragments of Visual Scenes Based on Analysis of Descriptors Trajectories in the Feature Space Mikhail Petrushan, Mikhail Kopeliovich, Aik Kurdoglyan and Anatoly Samarin	254
Automated Area Assessment of Objects Using Deep Learning Approach and Satellite Imagery Data	263
Similar Product Clustering for Long-Tail Cross-Sell Recommendations Vladislav Grozin and Alla Levina	273
An Algorithm for Packing Circles of Two Types in a Fixed Size Container with Non-Euclidean Metric	281
On a Typology of Nodes and Its Applications in Network Analysis Vladimir Matveenko	293
Process Model Repair by Detecting Unfitting Fragments	301
A Survey of Smart Cards Data Mining	314
Structural Patterns for System Sustainability: Experimental Study Olga Zvereva, Dmitry B. Berg and Andrey Kondratyev	326