

# Preface

Collecting, processing, and analyzing data have become important branches of computer science. Many areas of our existence generate a wealth of information that must be stored in a structured manner and processed appropriately in order to gain the knowledge from the inside. Databases have become a ubiquitous way of collecting and storing data. They are used to hold data describing many areas of human life and activity, and as a consequence, they are also present in almost every IT system. Today's databases have to face the problem of data proliferation and growing variety. More efficient methods for data processing are needed more than ever. New areas of interests that deliver data require innovative algorithms for data analysis.

Beyond Databases, Architectures and Structures (BDAS) is a series of conferences located in Central Europe and very important for this geographic region. The conference intends to give the state of the art of the research that satisfies the needs of modern, widely understood database systems, architectures, models, structures, and algorithms focused on processing various types of data. The aim of the conference is to reflect the most recent developments of databases and allied techniques used for solving problems in a variety of areas related to database systems, or even go one step forward — beyond the horizon of existing databases, architectures, and data structures.

The 13th International BDAS Scientific Conference (BDAS 2017), held in Ustroń, Poland, from May 30 to June 2, 2017, was a continuation of the highly successful BDAS conference series started in 2005. For many years BDAS has been attracting hundreds or even thousands of researchers and professionals working in the field of databases. Among attendees of our conference were scientists and representatives of IT companies. Several editions of BDAS were supported by our commercial, world-renowned partners, developing solutions for the database domain, such as IBM, Microsoft, Sybase, Oracle, and others. BDAS annual meetings have become an arena for exchanging information on the widely understood database systems and data-processing algorithms.

BDAS 2017 was the 13th edition of the conference, organized under the technical co-sponsorship of the IEEE Poland Section. We also continued our successful cooperation with Springer, which resulted in the publication of this book. The conference attracted more than a hundred participants from 15 countries, who made this conference a successful and memorable event. There were three keynote talks and one tutorial given by leading scientists: Prof. Jens Allmer from the Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir, gave an excellent keynote talk entitled “Database Integration Facilitating the Merging of MicorRNA and Gene Regulatory Pathways in ALS.” Prof. Dirk Labudde from the Bioinformatics group Mittweida (bigM) and Forensic Science Investigation Lab (FoSIL), University of Applied Sciences, Mittweida, Germany, honored us with a presentation entitled “3D Crime Scene and Disaster Site Reconstruction using Open Source Software.” Dr. Dominik Szczerba from Future Processing, Gliwice, Poland, gave a talk on “Computational Physiology.” Prof. Jean-Charles Lamirel from SYNALP team, LORIA,

Vandœuvre-lès-Nancy, France, prepared a tutorial on “Text Mining in the Big Data Context: Existing Approaches and Challenges.” The keynote speeches, tutorials, and plenary sessions allowed participants to gain insight into new areas of data analysis and data processing.

BDAS is focused on all aspects of databases. It is intended to have a broad scope, including different kinds of data acquisition, processing, and storing, and this book reflects fairly well the large span of research presented at BDAS 2017. This volume consists of 44 carefully selected papers that are assigned to seven thematic groups:

- Big data and cloud computing
- Artificial intelligence, data mining, and knowledge discovery
- Architectures, structures, and algorithms for efficient data processing
- Text mining, natural language processing, ontologies, and Semantic Web
- Bioinformatics and biological data analysis
- Industrial applications
- Data mining tools, optimization, and compression

The first group, containing four papers, is devoted to big data and cloud computing. Papers in this group discuss hot topics of stream processing with MapReduce, a tensor-based approach to temporal features modeling with application in big data, querying XML documents with SparkSQL, and automatic scaling computing infrastructure of the cloud. The second group contains six papers devoted to various methods used in data mining, knowledge discovery, and knowledge representation. Papers assembled in this group show a wide spectrum of applications of various exploration techniques, including decision rules, knowledge-based systems, clustering and classification algorithms, and rough sets, to solve many real-world problems.

The third group contains nine papers devoted to various database architectures and models, data structures, and algorithms used for efficient data processing. Papers in this group discuss the effectiveness of query execution, performance, and consistency of various database systems, including relational and NoSQL databases, indexing structures, sorting algorithms, and distributed data processing. The fourth group consists of nine papers devoted to natural language processing, text mining, ontologies, and the Semantic Web. These papers discuss problems of building recommendation systems with the use ontologies, extending expressiveness of knowledge description, ontology reuse for fast prototyping of new concepts, processing natural language instructions by robots, data integration in NLP, authorship attribution for texts, plagiarism detection, and RDF validation.

The research devoted to bioinformatics and biological data analysis is presented in six papers gathered in the fifth group. The papers cover problems connected with gene expression and chromatography but also medical diagnosing as well as face and emotion recognition. The sixth group includes four papers describing various applications of data mining — especially in coal mining and automotive industries. The last group includes six papers presenting various data-mining tools, performance optimization techniques, and a compression algorithm.

We hope that the broad scope of topics related to databases covered in this proceedings volume will help the reader to understand that databases have become an important element of nearly every branch of computer science.

We would like to thank all Program Committee members and additional reviewers for their effort in reviewing the papers. Special thanks to Piotr Kuźniacki — builder and for 12 years administrator of our website [bdas.polsl.pl](http://bdas.polsl.pl). The conference organization would not have been possible without the technical staff: Dorota Huget and Jacek Pietraszuk.

March 2017

Stanisław Kozielski  
Dariusz Mrozek  
Paweł Kasprowski  
Bożena Małysiak-Mrozek  
Daniel Kostrzewa

Beyond Databases, Architectures and Structures.  
Towards Efficient Solutions for Data Analysis and  
Knowledge Representation

13th International Conference, BDAS 2017, Ustroń,  
Poland, May 30 - June 2, 2017, Proceedings

KOZIELSKI, S.; Mrozek, D.; Kasprowski, P.;

Małysiak-Mrozek, B.; Kostrzewa, D. (Eds.)

2017, XVIII, 578 p. 184 illus., Softcover

ISBN: 978-3-319-58273-3