

Mieczysław A. Kłopotek,
Sławomir T. Wierzchoń,
Krzysztof Trojanowski
Editors

*Intelligent Information Processing and
Web Mining*

Proceedings of the International IIS:IIPWM'04 Conference

held in Zakopane, Poland, May 17-20, 2004

Springer

Berlin Heidelberg New York

Barcelona Budapest Hong Kong

London Milan Paris

Santa Clara Singapore Tokyo

Preface

This volume contains articles accepted for presentation during The Intelligent Information Processing and Web Mining Conference IIS:IIPWM'04 which was held in Zakopane, Poland, on May 17-20, 2004. This conference is a continuation of a series of 13 successful symposia and conferences on Intelligent Information Systems, organized by the Institute of Computer Science of Polish Academy of Sciences, devoted to new trends in (broadly understood) Artificial Intelligence.

The idea of organizing such meetings dates back to 1992. Our main intention guided the first, rather small-audience, workshop in the series was to resume the results gained in Polish scientific centers as well as contrast them with the research performed by Polish scientists working at the universities in Europe and USA and their foreign collaborators. This idea proved to be attractive enough that we decided to continue such meetings. As the years went by, the workshops has transformed into regular symposia devoted to such fields like Machine Learning, Knowledge Discovery, Natural Language Processing, Knowledge Based Systems and Reasoning, and Soft Computing (i.e. Fuzzy and Rough Sets, Bayesian Networks, Neural Networks and Evolutionary Algorithms). At present, about 50 papers prepared by researches from Poland and other countries are usually presented.

This year conference devotes much more attention to the newest developments in the area of Artificial Intelligence, related to broadly understood Web mining. In connection with these and related issues, contributions were accepted, concerning:

- recommenders and text classifiers
- natural language processing and understanding for search engines and other web applications
- computational linguistics
- information extraction and web mining
- data mining, machine learning and knowledge discovery technologies item knowledge representation
- web services and ontologies
- logics for artificial intelligence
- foundations of data mining
- medical and other applications of data mining

The above-mentioned topics were partially due to invited sessions organized by Erhard W. Hinrichs, Zbigniew W. Raś, Roman Świniarski, Henryk Rybiński, and Ryszard Tadeusiewicz.

Out of an immense flow of submissions, the Program Committee has selected only about 40 full papers for presentation and about a dozen of posters, constituting about 55% of the total number of submitted contributions.

On behalf of the Program Committee and of the Organizing Committee we would like to thank all participants: computer scientists mathematicians, engineers, logicians and other interested researchers who found excitement in advancing the area of intelligent systems. We hope that this volume of IIS:IIPWM:'04 Proceeding will be a valuable reference work in your further research.

We would like to thank the Programme Committee Members for their effort in evaluating contributions and in making valuable suggestions both concerning the scientific level and the organization of the Conference. we are also indebted to the scientists outside of the programme Committee that agreed to act as additional reviewers to cope with the extraordinary flow of papers.

We would like to thank Mr. M. Woliński for his immense effort in resolving technical issues connected with the preparation of this volume.

Zakopane, Poland,
May 2004

Mieczysław A. Kłopotek, Conference Co-Chair
Sławomir T. Wierzchoń, Conference Co-Chair
Krzysztof Trojanowski, Organizing Committee Chair

We would like to thank to the Programme Committee Members for their great job of evaluating the submissions

- Peter J. Bentley (University College London, UK)
- Petr Berka (University of Economics, Czech Republic)
- Leonard Bolc (Polish Academy of Science, Poland)
- Piotr Dembiński (Polish Academy of Sciences, Poland)
- Włodzisław Duch (Nanyang Technological University, Singapore)
- Tapio Elomaa (University of Helsinki, Finland)
- Ursula Gather (University of Dortmund, Germany)
- Jerzy W. Grzymała-Busse (University of Kansas, USA)
- Mohand-Saïd Hacid (Université Claude Bernard Lyon 1, France)
- Mirsad Hadzikadic (University of North Carolina at Charlotte, USA)
- Jan Hajič (Charles University, Czech Republic)
- Ray J. Hickey (University of Ulster, UK)
- Erhard Hinrichs (University of Tuebingen, Germany)
- Olgierd Hryniewicz (Polish Academy of Sciences, Poland)
- Janusz Kacprzyk (Polish Academy of Sciences, Poland)
- Samuel Kaski (Helsinki University of Technology, Finland)
- Jan Komorowski (Norwegian Univ. of Science and Technology, Norway)
- Józef Korbicz (University of Zielona Góra, Poland)
- Jacek Koronacki (Polish Academy of Sciences, Poland)
- Witold Kosiński (Polish-Japanese Institute of Information Technologies, Poland)
- Stan Matwin (University of Ottawa, Canada)
- Detmar Meurers (Ohio State University, USA)
- Maciej Michalewicz (NuTech Solutions Polska, Poland)
- Zbigniew Michalewicz (NuTech Solutions, USA)
- Ryszard Michalski (George Mason University, USA)
- Fionn Murtagh (Queen's University Belfast, UK)
- Zdzisław Pawlak (Scientific Research Committee, Poland)
- James F. Peters (University of Manitoba, Canada)
- Adam Przepiórkowski (Polish Academy of Sciences, Poland)
- Zbigniew W. Raś (University of North Carolina at Charlotte, USA)
- Jan Rauch (University of Economics, Czech Republic)
- Henryk Rybiński (Warsaw University of Technology, Poland)
- Abdel-Badeeh M. Salem (Ain Shams University, Egypt)
- Kiril Simov (Bulgarian Academy of Science, Bulgaria)
- Andrzej Skowron (Warsaw University, Poland)
- V.S. Subrahmanian (University of Maryland, USA)
- Katia Sycara (Carnegie Mellon University, USA)
- Roman Świniarski (San Diego State University, USA)
- Ryszard Tadeusiewicz (University of Mining and Metallurgy, Poland)
- Jonathan Timmis (University of Kent, UK)
- Antony Unwin (University of Augsburg, Germany)

VIII

- Alicja Wakulicz-Deja (University of Silesia, Poland)
- Jan Węglarz (Poznan University of Technology, Poland)
- Stefan Węgrzyn (Polish Academy of Sciences, Poland)
- Krzysztof Zieliński (University of Mining and Metallurgy, Poland)
- Djamel A. Zighed (Lumière Lyon 2 University, France)
- Jana Zvarová (EuroMISE Centre, Czech Republic)

We would like also to thank additional reviewers

- Witold Abramowicz (Poznań University of Economics, Poland)
- Jerzy Cytowski (Warsaw University, Poland)
- Andrzej Czyżewski (Gdańsk University of Technology, Poland)
- Jirka Hana (The Ohio State University, USA)
- Bożena Kostek (Gdańsk University of Technology, Poland)
- Geert-Jan Kruijff (Universität des Saarlandes, Germany)
- Anna Kupś (Polish Academy of Sciences, Poland)
- Małgorzata Marciniak (Polish Academy of Sciences, Poland)
- Agnieszka Mykowiecka (Polish Academy of Sciences, Poland)
- Zygmunt Vetulani (Adam Mickiewicz University, Poland)

Table of contents

Part I. Regular Sessions: Machine Learning, Machine Discovery and Data Mining

Mining Spatial Association Rules	3
<i>Robert Bembeník, Grzegorz Protaziuk</i>	
Modeling of the Form of Heliosphere and its Role on the Annual Variations of Galactic Cosmic Rays	13
<i>Tengiz B. Botchorishvili, Marina Nebieridze</i>	
ADX Algorithm: a brief description of a rule based classifier ..	19
<i>Michał Dramiński</i>	
Query Selectivity Estimation via Data Mining	29
<i>Jarek Gryz, Dongming Liang</i>	
Exploiting sensitivity analysis in Bayesian networks for customer satisfaction study	39
<i>Waldemar Jaroński, Josée Bloemer, Koen Vanhoof, Geert Wets</i>	
On Admissible Substitutions in Classical and Intuitionistic Sequent Logics	49
<i>Alexander Lyaletski</i>	
On Learning of Sunspot Classification	59
<i>Trung Thanh Nguyen, Claire P. Willis, Derek J. Paddon, Hung Son Nguyen</i>	
Testing microarray analysis methods with semi-synthetic data	69
<i>Michał Okoniewski, Bart Naudts</i>	
A new algorithm to build consolidated trees: study of the error rate and steadiness	79
<i>Jesus Maria Pérez, Javier Muguerza, Olatz Arbelaitz, Ibai Gurrutxaga</i>	
Mining texts by association rules discovery in a technical corpus	89
<i>Mathieu Roche, Jérôme Azé, Oriane Matte-Tailliez, Yves Kodratoff</i>	
A Hybrid Approach to a Classification Problem	99
<i>Piotr Sapiecha, Henry Selvaraj, Jarosław Stańczak, Krzysztof Sęp, Tadeusz Łuba</i>	
Confusion Matrix Visualization	107
<i>Robert Susmaga</i>	

Independent Component Analysis for Filtration in Data Mining	117
<i>Ryszard Szupiluk, Piotr Wojewnik, Tomasz Ząbkowski</i>	

Part II. Regular Sessions: Biologically Motivated Algorithms and Systems

State Assignment by Evolutionary Algorithms	129
<i>Mariusz Chyży, Witold Kosiński</i>	

Coevolutionary feature construction for transformation of representation of machine learners	139
<i>Krzysztof Krawiec, Leszek Włodarski</i>	

Theory of Collective Intelligence provides formal relations between Intelligence, Life, and Evolution	151
<i>Tadeusz Szuba, Marcin Szpyrka</i>	

Application of evolutionary computing to conformational analysis	161
<i>Aleksander Wawer, Franciszek Seredyński, Pascal Bouvry</i>	

A Novel Genetic Algorithm Based on Multi-Agent Systems ...	169
<i>Weicai Zhong, Jing Liu, Licheng Jiao</i>	

Part III. Regular Sessions: Natural Language Processing

Learning reversible categorial grammars from structures	181
<i>Jérôme Besombes, Jean-Yves Marion</i>	

An Approach to Computational Morphology	191
<i>Janusz S. Bień</i>	

Semi-incremental Addition of Strings to a Cyclic Finite Automaton	201
<i>Jan Daciuk</i>	

Semi-automatic stemmer evaluation	209
<i>Leo Galambos</i>	

Development of a Pediatric Text-Corpus for Part-of-Speech Tagging	219
<i>John Pestian, Łukasz Itert, Włodzisław Duch</i>	

Information Extraction for Polish Using the SProUT Platform	227
<i>Jakub Piškorski, Peter Homola, Małgorzata Marciniak, Agnieszka Mykowiecka, Adam Przepiórkowski, Marcin Woliński</i>	
Towards the Design of a Syntactico-Semantic Lexicon for Polish	237
<i>Adam Przepiórkowski</i>	
Automatic utterance boundaries recognition in large Polish text corpora	247
<i>Michał Rudolf, Marek Świdziński</i>	
A Flexible Rule Compiler for Speech Synthesis	257
<i>Wojciech Skut, Stefan Ulrich, Kathrine Hammervold</i>	
A Practical Application of Morphological Analysis of Polish ..	267
<i>Krzysztof Szafran</i>	
<hr/>	
Part IV. Regular Sessions: Web Mining, Web Technologies and Information Retrieval	
<hr/>	
On-line Thematic and Metadata Analysis of Document Collection	279
<i>Mikhail Ageev, Boris Dobrov, Nikolai Makarov-Zemlyanski</i>	
Analysis of priority and partitioning effects on web crawling performance	287
<i>Ali Mohammad Zareh Bidoki, Mazeiar Salehie, Mohammad Azadnia</i>	
Dynamization in IR Systems	297
<i>Leo Galambos</i>	
An Enhanced Semantic Indexing Implementation for Conceptual Information Retrieval	311
<i>Eric Jiang</i>	
Towards a More Realistic Web Graph Model	321
<i>Mieczysław A. Kłopotek, Marcin Sydow</i>	
Ontology As a Definition of Meaning	331
<i>Marek Łabuzek</i>	
Text Categorization using Learning Vector Quantization Algorithm	341
<i>M. Teresa Martín-Valdivia, Manuel García-Vega, Miguel A. García-Cumbreras, L. Alfonso Ureña López</i>	

Semantic Web Services in L2L	349
<i>Marcin Okraszewski, Henryk Krawczyk</i>	
Lingo: Search Results Clustering Algorithm Based on Singular Value Decomposition	359
<i>Stanisław Osiński, Jerzy Stefanowski, Dawid Weiss</i>	
Conceptual Clustering Using Lingo Algorithm: Evaluation on Open Directory Project Data	369
<i>Stanisław Osiński, Dawid Weiss</i>	
What Can Be Learned from Previously Answered Questions? A Corpus-Based Approach to Question Answering	379
<i>Marcin Skowron, Kenji Araki</i>	
Extensions of PageRank. The RBS Algorithm	389
<i>Marcin Sydow</i>	

Part V. Poster Session

Visualizing large data by the SOM and GTM methods – what are we obtaining?	399
<i>Anna Bartkowiak</i>	
AntSearch — a new search algorithm	404
<i>Urszula Boryczka</i>	
Trigram morphosyntactic tagger for Polish	409
<i>Łukasz Dębowski</i>	
Natural Language Processing Using Generative Indirect Dependency Grammar	414
<i>Stefan Diaconescu</i>	
Automatic Text Clustering in the Polish Language	419
<i>Marek Gajęcki</i>	
The impact of NLP techniques in the multilabel text classification problem	424
<i>Teresa Gonçalves, Paulo Quaresma</i>	
Cluster Analysis of Named Entities	429
<i>Zornitsa Kozareva, Joaquim Silva, Pablo Gamallo, Gabriel Lopes</i>	
Indexing Object Propagation in Large Systems	434
<i>Dariusz Król, Artur Moźdzynski</i>	
Intelligent Resources Search in Virtual Libraries	439
<i>Sebastian Ryszard Kruk, Henryk Krawczyk</i>	

A syntactical analysis of anaphora in the Polsyn parser	444
<i>Stawomir Kulików, Julia Romaniuk, Nina Suszczańska</i>	
Genetic Algorithm as an Attributes Selection Tool for Learning Algorithms	449
<i>Halina Kwasnicka, Piotr Orski</i>	
IBIS: A New Evolutionary Algorithm for Timetable Problem .	454
<i>Paweł B. Myszowski, Halina Kwasnicka</i>	
Tree-based Algorithm for Discovering Extended Action-Rules (System DEAR2)	459
<i>Li-Shiang Tsay, Zbigniew W. Ras, Alicja Wieczorkowska</i>	
Semantic Representation of English Phrasal Verbs	465
<i>Julija Telewnaja, Krista Bennett, Christian Hempelmann, Katrina E. Triezenberg</i>	
An algorithm of knowledge extraction from trained neural networks	470
<i>Tadeusz Wieczorek, Stawomir Golak</i>	
Document Ontology Discovery Tool	475
<i>Janusz Wnek</i>	

Part VI. Invited Session: Machine Learning Techniques for Computational Linguistics

Rule-based and Statistical Approaches to Morpho-syntactic Tagging of German	483
<i>Erhard W. Hinrichs, Julia S. Trushkina</i>	
Phonotactics in Inductive Logic Programming	493
<i>John Nerbonne, Stasinios Konstantopoulos</i>	

Part VII. Invited Session: Knowledge Discovery

A Hybrid Approach to Semantic Web Services Discovery	505
<i>Salima Benbernou, Etienne Candaud, Mohand-Said Hacid, Farouk Toumani</i>	
Discovering Logical Structures in Digital Documents	513
<i>Floriana Esposito, Stefano Ferilli, Teresa M.A. Basile, Nicola Di Mauro</i>	
Email Classification with Temporal Features	523
<i>Svetlana Kiritchenko, Stan Matwin, Suhayya Abu-Hakima</i>	

Rule Discovery by Axes-Driven Hyperplanes Construction . . .	535
<i>Zbigniew W. Ras, Agnieszka Dardzinska, Xingzhen Liu</i>	

Part VIII. Invited Session: Selected issues of knowledge discovery and representation in information systems

Scalable Methods of Discovering Rough Sets Reducts	547
<i>Marzena Kryszkiewicz, Katarzyna Cichon</i>	

Incorporation of quantitative knowledge into genetic information systems	557
<i>Michał Okoniewski, Koen Van Leemput, Piotr Gawrysiak</i>	

Exchange of Knowledge and Learning in Default Reasoning Based Agent Systems	565
<i>Dominik Ryzko, Henryk Rybiński</i>	

Part IX. Invited Session: Automatic Understanding of Medical Images and Other Signals

Automatic Understanding of Signals	577
<i>Ryszard Tadeusiewicz</i>	

Automatic understanding of ECG signal	591
<i>Piotr Augustyniak</i>	

Picture Languages in Cognitive Approach to Medical Visualization Understanding	599
<i>Marek R. Ogiela</i>	

Automatic Understanding of Speech Pathology	609
<i>Wiesław Wszótek, Tadeusz Wszótek</i>	

Part X. Invited Session: New Trends in Preprocessing and Pattern Classification

Application of Zernike Moments, Independent Component Analysis, and Rough and Fuzzy Classifier for Hand-Written Character Recognition	623
<i>Roman Swiniarski</i>	

Input preprocessing for neural feedforward architecture	633
<i>Bartłomiej Beliczynski</i>	