

Marenglen Biba and Fatos Xhafa (Eds.)

Learning Structure and Schemas from Documents

Studies in Computational Intelligence, Volume 375

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 353. Igor Aizenberg
Complex-Valued Neural Networks with Multi-Valued Neurons, 2011
ISBN 978-3-642-20352-7

Vol. 354. Ljupco Kocarev and Shiguo Lian (Eds.)
Chaos-Based Cryptography, 2011
ISBN 978-3-642-20541-5

Vol. 355. Yan Meng and Yaochu Jin (Eds.)
Bio-Inspired Self-Organizing Robotic Systems, 2011
ISBN 978-3-642-20759-4

Vol. 356. Slawomir Koziel and Xin-She Yang (Eds.)
Computational Optimization, Methods and Algorithms, 2011
ISBN 978-3-642-20858-4

Vol. 357. Nadia Nedjah, Leandro Santos Coelho, Viviana Cocco Mariani, and Luiza de Macedo Mourelle (Eds.)
Innovative Computing Methods and their Applications to Engineering Problems, 2011
ISBN 978-3-642-20957-4

Vol. 358. Norbert Jankowski, Włodzisław Duch, and Krzysztof Grańbczewski (Eds.)
Meta-Learning in Computational Intelligence, 2011
ISBN 978-3-642-20979-6

Vol. 359. Xin-She Yang, and Slawomir Koziel (Eds.)
Computational Optimization and Applications in Engineering and Industry, 2011
ISBN 978-3-642-20985-7

Vol. 360. Mikhail Moshkov and Beata Zielosko
Combinatorial Machine Learning, 2011
ISBN 978-3-642-20994-9

Vol. 361. Vincenzo Pallotta, Alessandro Soro, and Eloisa Vargiu (Eds.)
Advances in Distributed Agent-Based Retrieval Tools, 2011
ISBN 978-3-642-21383-0

Vol. 362. Pascal Bouvry, Horacio González-Vélez, and Joanna Kolodziej (Eds.)
Intelligent Decision Systems in Large-Scale Distributed Environments, 2011
ISBN 978-3-642-21270-3

Vol. 363. Kishan G. Mehrotra, Chilukuri Mohan, Jae C. Oh, Pramod K. Varshney, and Moonis Ali (Eds.)
Developing Concepts in Applied Intelligence, 2011
ISBN 978-3-642-21331-1

Vol. 364. Roger Lee (Ed.)
Computer and Information Science, 2011
ISBN 978-3-642-21377-9

Vol. 365. Roger Lee (Ed.)
Computers, Networks, Systems, and Industrial Engineering 2011, 2011
ISBN 978-3-642-21374-8

Vol. 366. Mario Köppen, Gerald Schaefer, and Ajith Abraham (Eds.)
Intelligent Computational Optimization in Engineering, 2011
ISBN 978-3-642-21704-3

Vol. 367. Gabriel Luque and Enrique Alba
Parallel Genetic Algorithms, 2011
ISBN 978-3-642-22083-8

Vol. 368. Roger Lee (Ed.)
Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2011, 2011
ISBN 978-3-642-22287-0

Vol. 369. Dominik Ryzko, Piotr Gawrysiak, Henryk Rybinski, and Marzena Kryszkiewicz (Eds.)
Emerging Intelligent Technologies in Industry, 2011
ISBN 978-3-642-22731-8

Vol. 370. Alexander Mehler, Kai-Uwe Kühnberger, Henning Lobin, Harald Lungen, Angelika Storrer, and Andreas Witt (Eds.)
Modeling, Learning, and Processing of Text Technological Data Structures, 2011
ISBN 978-3-642-22612-0

Vol. 371. Leonid Perlovsky, Ross Deming, and Roman Ilin (Eds.)
Emotional Cognitive Neural Algorithms with Engineering Applications, 2011
ISBN 978-3-642-22829-2

Vol. 372. António E. Ruano and Annamária R. Várkonyi-Kóczy (Eds.)
New Advances in Intelligent Signal Processing, 2011
ISBN 978-3-642-11738-1

Vol. 373. Oleg Okun, Giorgio Valentini, and Matteo Re (Eds.)
Ensembles in Machine Learning Applications, 2011
ISBN 978-3-642-22909-1

Vol. 374. Dimitri Plemenos and Georgios Miaoulis (Eds.)
Intelligent Computer Graphics 2011, 2011
ISBN 978-3-642-22909-1

Vol. 375. Marenglen Biba and Fatos Xhafa (Eds.)
Learning Structure and Schemas from Documents, 2011
ISBN 978-3-642-22912-1

Marenglen Biba and Fatos Xhafa (Eds.)

Learning Structure and Schemas from Documents

Editors

Marenglen Biba, PhD

University of New York Tirana,

Rr. Komuna E Parisit,

Tirana, Albania

E-mail: marenglenbiba@unyt.edu.al

Fatos Xhafa, PhD

Technical University of Catalonia,

Campus Nord, Ed. Omega

C/Jordi Girona 1-3

08034 Barcelona, Spain

E-mail: fatos@lsi.upc.edu

ISBN 978-3-642-22912-1

e-ISBN 978-3-642-22913-8

DOI 10.1007/978-3-642-22913-8

Studies in Computational Intelligence

ISSN 1860-949X

Library of Congress Control Number: 2011936645

© 2011 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Foreword

It was not long ago that database systems were revolutionized through the birth of the relational concepts and theory, which are now materialized in most commercial database management systems. It was then thought that data should be structured in a rather simple way to support the day-to-day business operations. This has certainly been made possible and supported by the mature transaction management concepts, and mathematically based query language and processes. Whilst this has been the backbone of businesses today, it cannot be ignored that most data in the world is not structured, or at least, easily and conveniently structured in a relational way.

The rise of an Internet world has undeniably contributed to the explosion of data in the digital universe. It has been reported not so long ago that there were over 1 trillion web pages – this roughly equals to almost 150 web pages per man, woman, and child on Earth, and the number of web pages increases billions per day. Another good example to illustrate data explosion in the digital universe is *facebook*. Also, it has been reported that the number of facebook users has surpassed the 500-million mark and is still growing strongly. If facebook users were citizen of a country called facebook, this country would have been the third largest country in the world, after China and India. And the kind of data is not easily categorized as structured data, since mostly, they are free from rigidly and constraining structures.

It was estimated slightly more than a couple years ago in 2008 that our digital universe grew by almost by half *zettabytes* – this is more than the data collected in the previous 5,000 years. To give readers perspective on this size, imagine gigabytes, terabytes, and then petabytes. It doesn't stop there; it continues to exabytes, and now zettabytes. It will not be surprising that in the near future, it will even surpass zettabytes to reach yottabytes. Next year in 2012, it is predicted that the annual growth rate of data will be 5 times that of 2008 – so, yottabytes, literally, are not far away. Of this data, it is estimated that 95% of them is unstructured.

Now how about the database market? It is not surprising to know that large enterprises have almost 10 thousand databases with 30-50 major data warehouses – many in tera-peta-exa byte range and double in size every 18 months. The database market is almost \$30billion a couple of years ago with 4 major vendors growing at 7%. The database infrastructure is the largest, fastest growing and highest priority software category. Hence, the structured concept still governs. It is still expected that the structures and schemas still exist in the so-called unstructured world of data. It is undeniably that structures and schemas play an important role in any aspect of data – either it is in data processing, data management, data manipulation, data analysis and

knowledge discovery, or even simple data or information retrieval. Every organization depends critically on the notion of structures and schemas of data.

This book addresses an important issue – that is finding, discovering, learning structures and schemas from documents and unstructured data. There is nothing more critical in the data explosion age in our digital universe than knowing that data have some degree of structures and schemas, as this is the foundation in many important applications, such as digital libraries, imaging, biomedical, medicine, and bioinformatics. This book covers the most comprehensive range of topics related to the importance of learning structures and schemas from documents – from basic techniques to advanced applications. I believe readers, from different sort of fields and background; from data oriented researchers to application practitioners, will benefit from reading a fine collection of articles compiled in this book.

Melbourne, June 2011

David Taniar
Editor-in-Chief *International Journal of
Data Warehousing and Mining* (SCI-E)

Preface by Editors

The rapidly growing volume of available digital documents of various formats and the possibility to access these through internet-based technologies, have led to the necessity to develop solid methods to properly organize and structure documents in large digital libraries and repositories. Specifically, since the extremely large volumes make it impossible to manually organize such documents and since most of the documents exist in an unstructured form and do not follow any schemas, most of the efforts in this direction are dedicated to automatically infer structure and schemas that can help to better organize huge collections. This is essential in order for these documents to be effectively and efficiently retrieved.

Dealing with unstructured information is currently a hot research topic. A growing body of work is addressing the problem of recognizing structure and schemas in documents of various types. Some areas are mainly concerned about the visual representation of documents and increasing improvements are being made in the area of pattern recognition and document layout analysis to classify documents according to structure found in their layout. On the other side, extensive research is being done in the field of machine learning to exploit attributes of documents and relationships among different documents to infer structures in large collections of documents. Important work is also being performed in the data mining and knowledge discovery community which has traditionally dealt with raw data but recently is dedicating attention to learning structure from unstructured information. In addition, Semantic Web researchers are dedicating important efforts to the problem of identifying structure and schemas in order for them to achieve ontology matching or alignment. Another related area regards the database community that has long worked with integration problems but only recently this community has started considering automatic structure and schema learning as a potential approach for schema and database integration. Finally information retrieval and extraction seek to infer structure and schemas from free text in order to build efficient information seeking models from large corpora.

Main Contributions of This Book

This book covers the latest advances in structure inference in heterogeneous collections of documents and data. The goal is to present the state-of-the-art in the area in order to describe some lessons learned, identify new research issues and challenges as well as opportunities for further developments.

Overall, the book contributes with the following:

- Presenting state-of-the-art approaches to learning and inferring structure and schemas from documents
- Assessing whether methods and techniques of one approach can be extended or exploited by the other approaches in a multi-strategic effort to the problem
- Assisting in the identification of new and future challenges and opportunities that will raise the bar in the state-of-the-art of several research areas attacking the same problem.
- Case studies and best practices from real large scale digital libraries, repositories and corpora.

Organization of the Book

The book consists of 19 chapters, organized as follows:

Chapter 1: Learning Structure and Schemas from Heterogeneous Domains in Networked Systems Surveyed

Marenglen Biba and Fatos Xhafa

This chapter presents a survey of the state-of-the-art methods for inferring structure from documents and schemas in networked environments. The survey is organized around the most important application domains, namely, bio-informatics, sensor networks, social networks, P2P systems, automation and control, transportation and privacy-preserving.

Chapter 2: Handling Hierarchically Structured Resources Addressing Interoperability Issues in Digital Libraries

Maristella Agosti, Nicola Ferro, and Gianmaria Silvello

This chapter presents the NESTed SeTs for Object hieRarchies (NESTOR) Framework that allows to model, manage, access and exchange hierarchically structured resources. The authors envision the framework in the context of Digital Libraries as a mean to address the complex and multiform concept of interoperability when dealing with hierarchical structures.

Chapter 3: Administrative Document Analysis and Structure

Abdel Belaïd, Vincent Poulain D'Andecy, Hatem Hamza, and Yolande Belaïd

This chapter presents a novel technique about the analysis and recognition of scanned administrative documents. This technique is related to the case-based reasoning already used in data mining and various problems of machine learning.

Chapter 4: Automatic Document Layout Analysis through Relational Machine Learning

Stefano Ferilli, Teresa M.A. Basile, Nicola Di Mauro, and Floriana Esposito

This chapter presents a prototypical version of the document processing system DOMINUS, using the incremental first-order logic learner INTHELEX. The authors propose the application of supervised Machine Learning techniques to infer correction rules to be applied to forthcoming documents.

Chapter 5: Dataspace: where structure and schema meet

Maurizio Atzori and Nicoletta Dessì

This chapter investigates the crucial problem that poses the bases to the concept of dataspace: the need for human interaction/intervention in the process of organizing (getting the structure of) unstructured data. The authors survey the existing techniques behind dataspace to overcome that need

Chapter 6: Transductive Learning of Logical Structures from Document Images

Michelangelo Ceci, Corrado Loglisci, and Donato Malerba

This chapter investigates the induction of a classifier for the automated recognition of semantically relevant layout components. In particular, it investigates the combination of transductive inference with principled relational classification.

Chapter 7: Progressive Filtering on the Web: the Press Reviews Case Study

Andrea Addis, Giuliano Armano and Eloisa Vargiu

This chapter presents NEWS.MAS, a multiagent system aimed at: (i) extracting information from online newspapers by using suitable wrapper agents, each associated with a specific information source, (ii) categorizing news articles according to a given taxonomy, and (iii) providing user feedback to improve the performance of the system depending on user needs and preferences.

Chapter 8: A Hybrid Binarization Technique for Document Images

Vavilis Sokratis, Ergina Kavallieratou, Roberto Paredes, and Kostas Sotiropoulos

This chapter presents a binarization technique specifically designed for historical document images. The technique is based on a hybrid approach that first applies a global thresholding technique and, then, identifies the image areas that are more likely to still contain noise.

Chapter 9: Digital Libraries and Document Image Retrieval Techniques:
a Survey*Simone Marinai, Beatrice Miotti, and Giovanni Soda*

This chapter presents a survey about the more recent techniques applied in the field of recognition and retrieval of text and graphical documents is presented. In particular it describes techniques related to recognition-free approaches.

Chapter 10: Mining Biomedical Text Towards Building a Quantitative
Food-disease-gene Network*Hui Yang, Rajesh Swaminathan, Abhishek Sharma, Vilas Ketkar, and
Jason D'Silva*

This chapter presents an information extraction system that analyzes publications in the emerging discipline of Nutritional Genomics. The system is used to systematically extract relationships among bio-entities from biomedical articles.

Chapter 11: Mining Tinnitus Data based on Clustering and new Temporal
Features*Xin Zhang, Pamela Thompson, Zbigniew W. Raś, and Pawel Jastreboff*

This chapter presents an initial research on data mining of medical data about Tinnitus Retraining Therapy. The authors present Interesting rules about the relationship recovery against symptoms, audiological therapy parameters, and other factors revealed during the experiments.

Chapter 12: DTW-GO Based Microarray Time Series Data Analysis for
Gene-Gene Regulation Prediction*Andy C. Yang and Hui-Huang Hsu*

This chapter introduces a novel approach that provides an effective distance measurement for genes based on gene ontology annotations. The authors show that the proposed approach facilitates analysis for microarray time series data.

Chapter 13: Integrating Content and Structure into a Comprehensive Framework
for XML Document Similarity Represented in 3D Space*Eric Draken, Tamer N. Jarada, Keivan Kianmehr, and Reda Alhajj*

In this chapter the authors have presented a novel way to represent XML document similarity in 3D space. Their approach benefits from the characteristics of the XML documents to produce a measure to be used in clustering and classification techniques, information retrieval and searching methods for the case of XML documents.

Chapter 14: Modelling User Behaviour on Page Content and Layout in Recommender Systems*Kin Fun Li and Kosuke Takano*

This chapter presents a literature survey on Web browsing behaviour, a recommender system based on user browsing behaviour, and the representation and manipulation of attributes associated with the design and layout of Web pages. A proof-of-concept system is designed with the objective to study whether there is a correlation between all these factors.

Chapter 15: MANENT: An Infrastructure for Integrating, Structuring and Searching Digital Libraries*Angela Locoro, Daniele Grignani, and Viviana Mascardi*

This chapter describes MANENT, an infrastructure for integrating, structuring and searching Digital Libraries. The technologies and standards enabling the realisation of infrastructures of this kind have been reviewed, as well as the related work.

Chapter 16: Low-Level Document Image Analysis and Description: From Appearance to Structure.*Emanuele Salerno, Pasquale Savino, and Anna Tonazzini*

This chapter deals with the problem of processing and analyzing digital images of ancient or degraded documents to increase the possibilities of inferring their structures.

Chapter 17: Model Learning from Published Aggregated Data*Janusz Wojtusiak and Ancha Baranova*

This chapter describes the problem of machine learning of models from aggregated data as compared to traditional learning from individual examples. It presents a method of rule induction from such data as well as an application of this method.

Chapter 18: Data De-duplication: A Review*Gianni Costa, Alfredo Cuzzocrea, Giuseppe Manco, and Riccardo Ortale*

This chapter presents an overview of research on data de-duplication, with the goal of providing a general understanding and useful references to fundamental concepts concerning the recognition of similarities in very large data collections.

Chapter 19: A Survey on Integrating Data in Bioinformatics

Andrea Manconi and Patricia Rodriguez-Tomé

This chapter discusses the current approaches, methods, systems, and open issues of integrating biological data.

Targeted Audience and Last Words

The audiences of this book are those working in or interested in joining interdisciplinary and trans-disciplinary works in the areas of data mining, machine learning, pattern recognition, document analysis and understanding, semantic web, databases, artificial intelligence and digital libraries, whose main focus is that of learning structure and schemas from unstructured information. The book also brings contributions from application areas such as from bioinformatics, web mining, text mining, information retrieval, real-world digital libraries, data warehouses and ontology building. Specifically, audiences who are broadly involved in the domains of computer science, web technologies, applied informatics, business or management information systems are: (1) researchers or senior graduates working in academia; (2) academics, instructors and senior students in colleges and universities, and (3) business analysts from industries interested in data integration, information retrieval and enterprise search.

Finally, academic researchers, professionals and practitioners in the field can also be inspired and put in practice the ideas and experiences proposed in the book in order to evaluate them for their specific research and work.

We hope that the readers will find this book useful in their academic and professional activities!

Acknowledgements

We would like to first thank all the authors of the chapters for their contributions and their efforts to produce high quality manuscripts to this volume. We are grateful to the referees who have carefully reviewed the chapters and gave useful suggestions and constructive feedback to the authors. We gratefully acknowledge the support and encouragement received from Prof. Janusz Kacprzyk, the editor in chief of Springer series “Studies in Computational Intelligence”, Dr Thomas Ditzinger, Ms Heather King and the whole Springer’s editorial team for their great and continuous support throughout the development of this book.

Finally, we are deeply indebted to our families for their love, patience and support!

Fatos Xhafa’s work is partially done at Birkbeck, University of London, during his Sabbatical Leave from Technical University of Catalonia (Barcelona, Spain). His research is supported by a grant from the General Secretariat of Universities of the Ministry of Education, Spain.

June 2011

Marenglen Biba
Fatos Xhafa

Contents

Learning Structure and Schemas from Heterogeneous Domains in Networked Systems Surveyed	1
<i>Marenglen Biba, Fatos Xhafa</i>	
Handling Hierarchically Structured Resources Addressing Interoperability Issues in Digital Libraries	17
<i>Maristella Agosti, Nicola Ferro, Gianmaria Silvello</i>	
Administrative Document Analysis and Structure	51
<i>Abdel Belaïd, Vincent Poulain D’Andecy, Hatem Hamza, Yolande Belaïd</i>	
Automatic Document Layout Analysis through Relational Machine Learning	73
<i>Stefano Ferilli, Teresa M.A. Basile, Nicola Di Mauro, Floriana Esposito</i>	
Dataspaces: Where Structure and Schema Meet	97
<i>Maurizio Atzori, Nicoletta Dessì</i>	
Transductive Learning of Logical Structures from Document Images ...	121
<i>Michelangelo Ceci, Corrado Loglisci, Donato Malerba</i>	
Progressive Filtering on the Web: The Press Reviews Case Study	143
<i>Andrea Addis, Giuliano Armano, Eloisa Vargiu</i>	
A Hybrid Binarization Technique for Document Images	165
<i>Vavilis Sokratis, Ergina Kavallieratou, Roberto Paredes, Kostas Sotiropoulos</i>	
Digital Libraries and Document Image Retrieval Techniques: A Survey	181
<i>Simone Marinai, Beatrice Miotti, Giovanni Soda</i>	

Mining Biomedical Text towards Building a Quantitative Food-Disease-Gene Network	205
<i>Hui Yang, Rajesh Swaminathan, Abhishek Sharma, Vilas Ketkar, Jason D'Silva</i>	
Mining Tinnitus Data Based on Clustering and New Temporal Features	227
<i>Xin Zhang, Pamela Thompson, Zbigniew W. Raś, Pawel Jastreboff</i>	
DTW-GO Based Microarray Time Series Data Analysis for Gene-Gene Regulation Prediction	247
<i>Andy C. Yang, Hui-Huang Hsu</i>	
Integrating Content and Structure into a Comprehensive Framework for XML Document Similarity Represented in 3D Space	275
<i>Eric Draken, Tamer N. Jarada, Keivan Kianmehr, Reda Alhajj</i>	
Modelling User Behaviour on Page Content and Layout in Recommender Systems	289
<i>Kin Fun Li, Kosuke Takano</i>	
MANENT: An Infrastructure for Integrating, Structuring and Searching Digital Libraries	315
<i>Angela Locoro, Daniele Grignani, Viviana Mascardi</i>	
Low-Level Document Image Analysis and Description: From Appearance to Structure	343
<i>Emanuele Salerno, Pasquale Savino, Anna Tonazzini</i>	
Model Learning from Published Aggregated Data	369
<i>Janusz Wojtusiak, Ancha Baranova</i>	
Data De-duplication: A Review	385
<i>Gianni Costa, Alfredo Cuzzocrea, Giuseppe Manco, Riccardo Ortale</i>	
A Survey on Integrating Data in Bioinformatics	413
<i>Andrea Manconi, Patricia Rodriguez-Tomé</i>	
Author Index	433
Subject Index	435

List of Contributors

Marenglen Biba
Department of Computer Science,
University of New York, Tirana
Rr. "Komuna e Parisit", Tirana, Albania
marenglenbiba@unyt.edu.al

Fatos Xhafa
Department of Languages and
Information Systems
Technical University of Catalonia
Campus Nord. Ed. Omega
C/ Jordi Girona, 1-3
08034 Barcelona, Spain
fatos@lsi.upc.edu

Maristella Agosti
Department of Information Engineering
University of Padua
DEI - Via Gradenigo 6/B - 35131
Padova, Italy
agosti@dei.unipd.it

Nicola Ferro
Department of Information Engineering,
University of Padua
DEI - Via Gradenigo 6/B - 35131
Padova, Italy
ferro@dei.unipd.it

Gianmaria Silvello
Department of Information Engineering
University of Padua
DEI - Via Gradenigo 6/B - 35131
Padova, Italy
silvello@dei.unipd.it

Maurizio Atzori
University of Cagliari
Via Ospedale 72, 09124 Cagliari, Italy
atzori@unica.it

Nicoletta Dessì
University of Cagliari
Via Ospedale 72, 09124 Cagliari, Italy
dessi@unica.it

Abdel Belaïd
LORIA, University of Nancy 2
UMR 7503
Campus scientifique B.P. 239
54506 Vandoeuvre Lès Nancy Cedex,
France
abdel.belaïd@loria.fr

Vincent Poulain D'Andecy
ITESOFT, France,
Vincent.PoulainDAndecy@itesoft.com

Hatem Hamza
ITESOFT, France,
Hatem.Hamza@itesoft.com

Yolande Belaïd
LORIA, University of Nancy 2
UMR 7503
Campus scientifique B.P. 239
54506 Vandoeuvre Lès Nancy Cedex,
France
Yolande.Belaïd@loria.fr

Michelangelo Ceci,
Dipartimento di Informatica,
Università degli Studi di Bari "Aldo
Moro" via Orabona 4 - 70125
Bari, Italy.
ceci,@di.uniba.it

Corrado Loglisci
Dipartimento di Informatica,
Università degli Studi di Bari "Aldo
Moro" via Orabona 4 - 70125
Bari, Italy.
loglisci@di.uniba.it

Donato Malerba
Dipartimento di Informatica,
Università degli Studi di Bari
“Aldo Moro” via Orabona 4 - 70125
Bari, Italy.
malerba@di.uniba.it

Gianni Costa
ICAR-CNR, Via P. Bucci, 41C,
87036 Rende (CS)
Italy.
costa@icar.cnr.it

Alfredo Cuzzocrea
ICAR-CNR, Via P. Bucci, 41C,
87036 Rende (CS) - Italy.
cuzzocrea@icar.cnr.it

Giuseppe Manco
ICAR-CNR, Via P. Bucci, 41C,
87036 Rende (CS) Italy.
manco@icar.cnr.it

Riccardo Ortale
ICAR-CNR, Via P. Bucci, 41C,
87036 Rende (CS) - Italy.
ortale@icar.cnr.it

Vavilis Sokratis
Dept. of Information and
Communication Systems Engineering
University of the Aegean
83200 Karlovasi
Samos, Greece
sokratisvav@gmail.com

Ergina Kavallieratou
Dept. of Information and
Communication Systems Engineering
University of the Aegean
83200 Karlovasi
Samos, Greece
kavallieratou@aegean.gr

Roberto Paredes
PRHLT - Universidad Politecnica de
Valencia
Camino de Vera s/n
46071 Valencia
Spain
rparedes@dsic.upv.es

Kostas Sotiropoulos
University of Patras,
University campus, 26504, Rio, Patras,
Greece
kosotiro@upatras.gr

Stefano Ferilli
Dipartimento di Informatica,
Università degli Studi di Bari “Aldo
Moro” via Orabona 4 - 70125
Bari, Italy.
ferilli@di.uniba.it

Teresa M.A. Basile and
Dipartimento di Informatica,
Università degli Studi di Bari “Aldo
Moro” via Orabona 4 - 70125
Bari, Italy.
basile@di.uniba.it

Nicola Di Mauro
Dipartimento di Informatica,
Università degli Studi di Bari “Aldo
Moro” via Orabona 4 - 70125
Bari, Italy.
ndm@di.uniba.it

Floriana Esposito
Dipartimento di Informatica,
Università degli Studi di Bari “Aldo
Moro” via Orabona 4 - 70125
Bari, Italy.
esposito@di.uniba.it

Hui Yang,
Department of Computer Science, 1600
Holloway Avenue . San Francisco . CA
94132, USA
huiyang@cs.sfsu.edu

Rajesh Swaminathan
Department of Computer Science, 1600
Holloway Avenue . San Francisco . CA
94132, USA
rajeshs@sfsu.edu

Abhishek Sharma
Department of Computer Science, 1600
Holloway Avenue . San Francisco . CA
94132, USA
asharmal@sfsu.edu

Vilas Ketkar
Department of Computer Science, 1600
Holloway Avenue . San Francisco . CA
94132, USA
vilask@sfsu.edu

Jason D'Silva
Department of Computer Science, 1600
Holloway Avenue . San Francisco . CA
94132, USA

Andy C. Yang
Department of Computer Science &
Information Engineering, Tamkang
University, Taipei, 25137,
Taiwan R.O.C.
andyyung0215@gmail.com

Hui-Huang Hsu
Department of Computer Science &
Information Engineering, Tamkang
University, Taipei, 25137,
Taiwan R.O.C.
huihuanghsu@gmail.com

Eric Draken
Computer Science Department
University of Calgary
2500 University Dr. NW
Calgary, Alberta Canada

Tamer N. Jarada
Computer Science Department
University of Calgary
2500 University Dr. NW
Calgary, Alberta Canada
tjarada@gmail.com

Keivan Kianmehr
Computer Engineering Department,
University of Western Ontario, London,
Ontario, Canada
kkianmeh@uwo.ca

Reda Alhajj
Computer Science Department
University of Calgary
2500 University Dr. NW
Calgary, Alberta Canada
alhajj@ucalgary.ca

Kin Fun Li
Department of Electrical and Computer
Engineering, University of Victoria
P.O. Box 3055 STN CSC
Victoria, B.C. V8W 3P6
Canada
kinli@uvic.ca

Kosuke Takano
Department of Information and
Computer Sciences, Kanagawa
Institute of Technology, Japan
takano@ic.kanagawa-it.ac.jp

Angela Locoro
Computer Science Department,
University of Genova, Via Dodecaneso
35, 16146 Genova, Italy
locoro@disi.unige.it

Daniele Grignani
Department of Modern and
Contemporary History, University of
Genova, Via Balbi 6, 16126 Genova,
Italy
daniele.grignani@gmail.com

Viviana Mascardi
Computer Science Department,
University of Genova, Via Dodecaneso
35, 16146 Genova, Italy
mascardi@disi.unige.it

Andrea Manconi
Institute for Biomedical Technologies,
National Research Council, Via F.lli
Cervi 93, 20090, Seg-
rate (MI), Italy.
andrea.manconi@itb.cnr.it

Patricia Rodriguez-Tomé
CRS4 - Center for Advanced Studies,
Research and Development in Sardinia,
Loc. Piscina Manna,
Ed.1, 09010 Pula (CA), Italy
prtome@crs4.it

Simone Marinai
Dipartimento di Sistemi e Informatica.
Via di Santa Marta, 3 - 50139, Firenze,
Italy
simone.marinai@unifi.it

Beatrice Miotti
Dipartimento di Sistemi e Informatica.
Via di Santa Marta, 3 - 50139, Firenze,
Italy
miotti@dsi.unifi.it

Giovanni Soda
Dipartimento di Sistemi e Informatica.
Via di Santa Marta, 3 - 50139, Firenze,
Italy
soda@dsi.unifi.it

Xin Zhang,
University of North Carolina at
Pembroke, Dept. of Math. Comp.
Science, Pembroke, NC 28372, USA
xin.zhang@uncp.edu

Pamela Thompson,
University of North Carolina,
Dept. of Computer Science, Charlotte,
NC 28223, USA
pthomps@catawba.edu

Zbigniew W. Ras,
University of North Carolina,
Dept. of Computer Science, Charlotte,
NC 28223, USA & Warsaw
University of Technology, Institute of
Comp. Science, 00-665 Warsaw, Poland
ras@uncc.edu

Pawel Jastreboff
Emory University School of Medicine,
Dept. of Otolaryngology, Atlanta, GA
30322, USA
pjastre@emory.

Emanuele Salerno
CNR - Istituto di Scienza e Tecnologie
dell'Informazione, Via Moruzzi 1,
56124 Pisa, Italy
emanuele.salerno@isti.cnr.it

Pasquale Savino
CNR - Istituto di Scienza e Tecnologie
dell'Informazione, Via Moruzzi 1,
56124 Pisa, Italy
pasquale.savino@isti.cnr.it

Anna Tonazzini
CNR - Istituto di Scienza e Tecnologie
dell'Informazione, Via Moruzzi 1,
56124 Pisa, Italy
anna.tonazzini@isti.cnr.it

Andrea Addis
Dept. of Electrical and Electronic Engi-
neering, University of Cagliari, Piazza
D'Armi, I-09123 Cagliari, Italy
addis@diee.unica.it

Giuliano Armano
Dept. of Electrical and Electronic
Engineering, University of Cagliari,
Piazza D'Armi, I-09123 Cagliari, Italy
armano@diee.unica.it

Eloisa Vargiu
Dept. of Electrical and Electronic
Engineering, University of Cagliari,
Piazza D'Armi, I-09123 Cagliari, Italy
vargiu@diee.unica.it

Janusz Wojtusiak
Department of Health Administration
and Policy, George Mason University
Northeast Module, Room 108
4400 University Drive, MSN 1J3
Fairfax, VA 22030, USA
jwojt@mli.gmu.edu

Ancha Baranova
The Center for Biomedical Genomics
Room 182 Discovery Hall, MSN 4D7
10900 University Blvd
Manassas VA, 20110
abaranov@gmu.edu