26th Italian Symposium on Advanced Database Systems (SEBD 2018)

Castellaneta Marina, Italy
24-27 June 2018

Editors:

Sonia Bergamaschi
Andrea Maurino
Tommaso Di Noia
Letizia Tanca

ISBN: 978-1-5108-6963-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© for individual papers are retained by the authors.
All rights reserved.

Printed by Curran Associates, Inc. (2018)

Edited by:
Sonia Bergamaschi
Tommaso Di Noia
Andrea Maurino
Letizia Tanca

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

AN APPROACH TO EXTRACTING THEMATIC VIEWS FROM HIGHLY HETEROGENEOUS SOURCES OF A DATA LAKE 1
C. Diamatini, P. Giudice, L. Musarella, D. Potena, E. Storti, D. Ursino

FILLING THE GAP BETWEEN BUSINESS RULES AND TECHNICAL REQUIREMENTS IN BUSINESS ANALYTICS: THE FACT - CENTERED ETL APPROACH 13
A. Longo, M. Bochicchio, M. Zappatore, L. Vaira

FIRST STEPS TOWARDS REASONING ON BIG DATA WITH DLV 21
N. Leone, S. Perri, F. Ricca, P. Veltri, J. Zangari

NEWSVALLUM: SEMANTICS-AWARE TEXT AND IMAGE PROCESSING FOR FAKE NEWS DETECTION SYSTEM 30
G. Armano, S. Battiato, D. Bennato, L. Boratto, S. Carta, T. Noia, E. Sciascio, A. Ortis, D. Recupero

A BLOCKCHAIN INFRASTRUCTURE FOR THE SEMANTIC WEB OF THINGS 38
M. Ruta, F. Scioscia, S. Ieva, G. Capurso, A. Pinto, E. Seiascio

ETHICS-AWARE DATA GOVERNANCE 50
L. Tanca, P. Atzeni, D. Azzalini, I. Bartolini, L. Cabibbo, L. Calderoni, P. Ciaccia, V. Crescenzi, J. Martin, S. Fenoglietto, D. Firmani, S. Grego, F. Isgro, D. Maio, D. Martinenghi, M. Matera, P. Merialdo, C. Molinaro, M. Patella, R. Prevete, E. Quintarelli, A. Santangelo, A. Tagerelli, G. Tamburrini, R. Torlone

INTRODUCING ONLINE PROFILE LEARNING IN CROWDSOURCING TASK ROUTING 58
S. Castano, A. Ferrara, S. Montanelli

ON THE INTERPRETATION OF TRACES OF LOW LEVEL EVENTS IN BUSINESS PROCESS LOGS 66
B. Fazzinga, S. Flesca, F. Furfaro, E. Masciari, L. Pontieri

HOW TO ROBUSTLY COMBINE JUDGEMENTS FROM CROWD ASSESSORS WITH AWARE 74
M. Ferrante, N. Ferro, M. Maistro

MULTIDIMENSIONAL MINING OVER BIG HEALTHCARE DATA: A BIG DATA ANALYTICS FRAMEWORK 82
M. Bochicchio, A. Cuzzocrea, L. Vaira, A. Longo, M. Zappatore

COUNTING VEHICLES WITH CAMERAS 91
L. Ciampi, G. Amato, F. Falchi, C. Gennaro, F. Rabitti

AN ATTENTION-BASED RECURRENT NEURAL NETWORKS FRAMEWORK FOR HEALTH DATA ANALYSIS 99
Q. Suo, F. Ma, G. Canino, J. Gao, A. Zhang, A. Gnasso, G. Tradigo, P. Veltri

LET'S MAKE IT DIRTY WITH BART! 106
D. Santoro, P. Arocena, B. Glavic, G. Mecca, R. Miller, P. Papotti

HOW MODERN DEDUCTIVE DATABASE SYSTEMS CAN ENHANCE DATA INTEGRATION 114
F. Calimeri, S. Perri, G. Terracina, J. Zangari

TOWARDS PROGRESSIVE SEARCH-DRIVEN ENTITY RESOLUTION 122
A. Pietrangelo, G. Simonini, S. Bergamasehi, I. Koumarelas, F. Naumann

A NOVEL ACCUSATION MODEL FOR DOCUMENT FINGERPRINTING 130
B. Fazzinga, S. Flesca, F. Furfaro, E. Masciari

CONTRASTING FAKE REVIEWS IN TRIPADVISOR 142
F. Buccafurri, M. Fazzolari, G. Lax, M. Petrocchi

BEYOND SKYLINE AND RANKING QUERIES: RESTRICTED SKYLINES 150
P. Ciaccia, D. Martinenghi

APPROXIMATE QUERY ANSWERING OVER INCONSISTENT KNOWLEDGE BASES 158
S. Greco, C. Molinaro, I. Trubitsyna

COMPLEXITY OF APPROXIMATE QUERY ANSWERING UNDER INCONSISTENCY IN DATALOG$^{\pm}$ 166
T. Lukasiewicz, E. Malizia, C. Molinaro

DEPENDENCY-BASED QUERY RESULT APPROXIMATION 174
L. Caruccio, V. Deufemia, G. Polese

FROM HYPERTREE WIDTH TO SUBMODULAR WIDTH AND DATA-DEPENDENT STRUCTURAL DECOMPOSITIONS 182
F. Scarcello

FORMALIZATION AND COMPLEXITY OF MONGODB QUERIES 190
E. Botoeva, D. Calvanese, B. Cogrel, G. Xiao

ONTOLOGY POPULATION FOR OPEN-SOURCE INTELLIGENCE 198
G. Ganino, D. Lembo, M. Mecella, F. Scafoglieri
AN IOT-ENABLED FRAMEWORK FOR CONTEXT-AWARE ROLE-BASED ACCESS CONTROL 206
G. Delzanno, G. Guerrini
LESSONS LEARNED AND RESEARCH AGENDA FOR BIG DATA INTEGRATION OF PRODUCT SPECIFICATIONS 217
L. Barbosa, V. Crescenzi, X. Dong, P. Merialdo, F. Piai, D. Qiu, Y. Shen, D. Srivastava
WIND SPEED FORECASTING VIA STRUCTURED OUTPUT LEARNING 225
A. Appice, A. Lanza, D. Malerba
ENSEMBLE LEARNING FOR MULTI-TYPE CLASSIFICATION IN HETEROGENEOUS NETWORKS 236
F. Serafino, G. Pio, M. Ceci
DERIVING LOCAL EXPLANATIONS FOR BLACK BOX MODELS 244
E. Pastor, E. Baralis
QUERYING RDF DATA CUBES THROUGH NATURAL LANGUAGE 252
M. Atzori, G. Mazzeo, C. Zaniolo
CONTEXT AWARE SOURCE SELECTION FOR LINKED DATA 260
B. Catania, G. Guerrini, B. Yaman
USING ONTOLOGY-BASED DATA SUMMARIZATION TO DEVELOP SEMANTICS-AWARE RECOMMENDER SYSTEMS 268
V. Anelli, T. Noia, A. Maurino, M. Palmonari, A. Rula
PROCESSING SPARQL QUERIES ON DEEP WEB SOURCES 276
A. Cali, T. Noia, T. Lynch, A. Ragone
A SIMILARITY FUNCTION FOR MULTI-LEVEL AND MULTI-DIMENSIONAL ITEMSETS 284
M. Francia, M. Golfarelli, S. Rizzi
ENABLING BIG DATA EXPLORATION IN DYNAMIC CONTEXTS 292
A. Bagozi, D. Bianchini, V. Antonellis, A. Marini, D. Ragazzi
DESIGN AND EVALUATION OF A GENOMICS VARIANT ANALYSIS PIPELINE USING GATK SPARK TOOLS 300
N. Tucci, J. Cala, J. Steyn, P. Missier
CLIQUES ARE TOO STRICT FOR REPRESENTING COMMUNITIES: FINDING LARGE K-PLEXES IN REAL NETWORKS 312
A. Conte, D. Firmani, C. Mordente, M. Patrignani, R. Torlone
TEMPORAL RECURRENT ACTIVATION NETWORKS 320
G. Maneo, G. Pirro, E. Ritacco
GRAPH DATABASES BENCHMARKING ON THE ITALIAN BUSINESS REGISTER 328
N. Ferro, L. Sinico
GRAPHDBLP RELEASED: QUERYING THE COMPUTER SCIENTISTS NETWORK AS A GRAPH 340
M. Cesarini, F. Mercorio, M. Mezzanzanica, V. Moscato, A. Picariello
INVERSE TREE-OLAP: DEFINITION, COMPLEXITY AND FIRST SOLUTION 348
D. Sacca, E. Serra, A. Cuzzocrea
AN INNOVATIVE FRAMEWORK FOR COMBINING SET SIMILARITY JOIN ALGORITHMS AND CLUSTERING 358
L. Ribeiro, A. Cuzzocrea, K. Bezerra, B. Nascimento, M. Nolich
DERIVING LOCAL INTERNAL LOGIC FOR BLACK BOX MODELS 366
E. Pastor
ON THE INCREMENTAL COMPUTATION OF ARGUMENTATION FRAMEWORKS 370
G. Alfano
EFFICIENT AND EFFECTIVE SOURCE SELECTION IN THE WEB OF DATA: A SMART CACHING APPROACH 374
F. Fino
A METHODOLOGY FOR GDPR COMPLIANT DATA PROCESSING 378
D. Desiato
Author Index