

# Contents

## Similarity Measure and Clustering

Determining Query Readiness for Structured Data . . . . .	3
<i>Farid Alborzi, Rada Chirkova, Jon Doyle, and Yahya Fathi</i>	
Efficient Cluster Detection by Ordered Neighborhoods . . . . .	15
<i>Emin Aksehirli, Bart Goethals, and Emmanuel Müller</i>	
Unsupervised Semantic and Syntactic Based Classification of Scientific Citations . . . . .	28
<i>Mohammad Abdullatif, Yun Sing Koh, and Gillian Dobbie</i>	

## Data Mining

HI-Tree: Mining High Influence Patterns Using External and Internal Utility Values . . . . .	43
<i>Yun Sing Koh and Russel Pears</i>	
Balancing Tree Size and Accuracy in Fast Mining of Uncertain Frequent Patterns . . . . .	57
<i>Carson Kai-Sang Leung and Richard Kyle MacKinnon</i>	
Secure Outsourced Frequent Pattern Mining by Fully Homomorphic Encryption . . . . .	70
<i>Junqiang Liu, Jiuyong Li, Shijian Xu, and Benjamin C.M. Fung</i>	
Supervised Evaluation of Top- $k$ Itemset Mining Algorithms . . . . .	82
<i>Claudio Lucchese, Salvatore Orlando, and Raffaele Perego</i>	
Finding Banded Patterns in Data: The Banded Pattern Mining Algorithm. . . .	95
<i>Fatimah B. Abdullahi, Frans Coenen, and Russell Martin</i>	
Discrimination-Aware Association Rule Mining for Unbiased Data Analytics . . . . .	108
<i>Ling Luo, Wei Liu, Irena Koprinska, and Fang Chen</i>	

## Social Computing

Big Data Analytics of Social Networks for the Discovery of “Following” Patterns . . . . .	123
<i>Carson Kai-Sang Leung and Fan Jiang</i>	

Sentiment Extraction from Tweets: Multilingual Challenges . . . . .	136
<i>Nantia Makrynioti and Vasilis Vassalos</i>	

TiDE: Template-Independent Discourse Data Extraction. . . . .	149
<i>Jayendra Barua, Dhaval Patel, and Vikram Goyal</i>	

**Heterogeneous Networks and Data**

A New Relevance Measure for Heterogeneous Networks . . . . .	165
<i>Mukul Gupta, Pradeep Kumar, and Bharat Bhasker</i>	

UFOMQ: An Algorithm for Querying for Similar Individuals in Heterogeneous Ontologies . . . . .	178
<i>Yinuo Zhang, Anand Panangadan, and Viktor K. Prasanna</i>	

Semantics-Based Multidimensional Query Over Sparse Data Marts . . . . .	190
<i>Claudia Diamantini, Domenico Potena, and Emanuele Storti</i>	

**Data Warehouses**

Automatically Tailoring Semantics-Enabled Dimensions for Movement Data Warehouses . . . . .	205
<i>Juarez A.P. Sacenti, Fabio Salvini, Renato Fileto, Alessandra Raffaetà, and Alessandro Roncato</i>	

Real-Time Snapshot Maintenance with Incremental ETL Pipelines in Data Warehouses. . . . .	217
<i>Weiping Qu, Vinanthi Basavaraj, Sahana Shankar, and Stefan Dessloch</i>	

Eco-Processing of OLAP Complex Queries . . . . .	229
<i>Amine Roukh and Ladjel Bellatreche</i>	

Materializing Baseline Views for Deviation Detection Exploratory OLAP . . .	243
<i>Pedro Furtado, Sergi Nadal, Veronika Peralta, Mahfoud Djedaini, Nicolas Labroche, and Patrick Marcel</i>	

**Stream Processing**

Binary Shapelet Transform for Multiclass Time Series Classification . . . . .	257
<i>Aaron Bostrom and Anthony Bagnall</i>	

StreamXM: An Adaptive Partitional Clustering Solution for Evolving Data Streams . . . . .	270
<i>Robert Anderson and Yun Sing Koh</i>	

Data Stream Mining with Limited Validation Opportunity: Towards Instrument Failure Prediction . . . . .	283
<i>Katie Atkinson, Frans Coenen, Phil Goddard, Terry Payne, and Luke Riley</i>	
Distributed Classification of Data Streams: An Adaptive Technique. . . . .	296
<i>Alfredo Cuzzocrea, Mohamed Medhat Gaber, and Ary Mazharuddin Shiddiqi</i>	
New Word Detection and Tagging on Chinese Twitter Stream . . . . .	310
<i>Yuzhi Liang, Pengcheng Yin, and S.M. Yiu</i>	

## Applications of Big Data Analysis

Text Categorization for Deriving the Application Quality in Enterprises Using Ticketing Systems . . . . .	325
<i>Thomas Zimmer, Florian Lemmerich, Susanna Schwarzmann, Matthias Hirth, Peter Karg, and Andreas Hotho</i>	
MultiSpot: Spotting Sentiments with Semantic Aware Multilevel Cascaded Analysis . . . . .	337
<i>Despoina Chatzakou, Nikolaos Passalis, and Athena Vakali</i>	
Online Urban Mobility Detection Based on Velocity Features. . . . .	351
<i>Fernando Terroso-Saenz, Mercedes Valdes-Vela, and Antonio F. Skarmeta-Gomez</i>	

## Big Data

Partition and Conquer: Map/Reduce Way of Substructure Discovery . . . . .	365
<i>Soumyava Das and Sharma Chakravarthy</i>	
Implementation of Multidimensional Databases with Document-Oriented NoSQL . . . . .	379
<i>M. Chevalier, M. El Malki, A. Kopliku, O. Teste, and R. Tournier</i>	
A Graph-Based Concept Discovery Method for n-Ary Relations . . . . .	391
<i>Nazmiye Ceren Abay, Alev Mutlu, and Pinar Karagoz</i>	
Exact Detection of Information Leakage in Database Access Control. . . . .	403
<i>Farid Alborzi, Rada Chirkova, and Ting Yu</i>	

<b>Author Index</b> . . . . .	417
-------------------------------	-----

Big Data Analytics and Knowledge Discovery  
17th International Conference, DaWaK 2015, Valencia,  
Spain, September 1-4, 2015, Proceedings  
Madria, S.; Hara, T. (Eds.)  
2015, XIII, 418 p. 116 illus., Softcover  
ISBN: 978-3-319-22728-3