10th International Conference on Data Science, Technology and Applications (DATA 2021)

Online
6 – 8 July 2021

Editors:

Christoph Quix
Slimane Hammoudi
Wil van der Aalst

ISBN: 978-1-7138-3993-4

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© (2021) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda.
at the address below.

SCITEPRESS – Science and Technology Publications, Lda.
Avenida de S. Francisco Xavier, Lote 7 Cv. C,
2900-616 Setúbal, Portugal

Phone: +351 265 520 185
Fax: +351 265520 186

info@scitepress.org

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

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

From Representation to Mediation: Modeling Information Systems in a Digital World 5
Jan Recker

A Profile-aware Methodological Framework for Collaborative Multidimensional Modeling: Agro-biodiversity Case Study 7
Sandro Bimonte

Querying Decentralized Knowledge Graphs 9
Hala Skaf-Molli

Database Systems and Information Management: Trends and a Vision 11
Volker Markl

PAPERS

FULL PAPERS

Similarity of Software Libraries: A Tag-based Classification Approach 17
Maximilian Auch, Maximilian Balluff, Peter Mandl and Christian Wolff

A Comparative Study on Inflated and Dispersed Count Data 29
Monika Arora, Yash Kalyani and Shivam Shanker

Database Recovery from Malicious Transactions: A Use of Provenance Information 39
Theppatorn Rhujittawiwat, John Ravan, Ahmed Saaudi, Shankar Banik and Csilla Farkas

An Efficient Representation of Enriched Temporal Trajectories 49
Nieves R. Brisaboa, Antonio Fariña, Diego Otero-González and Tirso V. Rodeiro

textPrep: A Text Preprocessing Toolkit for Topic Modeling on Social Media Data 60
Rob Churchill and Lisa Singh

Semantic Entanglement on Verb Negation 71
Yuto Kikuchi, Kazuo Hara and Ikumi Suzuki

A Study on the Effects of Response Time on Travel Package Attributes 79
Usha Ananthakumar and Sagun Pai

Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques 88
Ahmad Homsi, Joyce Al Nemri, Nisma Naimat, Hamzeh Abdul Kareem, Mustafa Al-Fayoumi and Mohammad Abu Snober

A Longitudinal Model for Song Popularity Prediction 96
Ahmet Çimen and Enis Kayış

Forecasting Stock Market Trends using Deep Learning on Financial and Textual Data 105
Georgios-Markos Chatziloizos, Dimitrios Gunopulos and Konstantinos Konstantinou

A Graph-based Approach at Passage Level to Investigate the Cohesiveness of Documents 115
Ghulam Sarwar and Colm O'Riordan

A Reference Process for Judging Reliability of Classification Results in Predictive Analytics 124
Simon Staudinger, Christoph G. Schuetz and Michael Schrefl

SHORT PAPERS

Determining How Different Factors Affect Police-Allegation's Sustainability in Chicago using Decision-Tree 137
Linxin Yang

Archival and Museum Information as a Component of the Common Digital Space of Scientific Knowledge 144
N. Kalenov, I. Sobolevskaya and A. Sotnikov

Building an Integrated Relational Database from Swiss Nutrition's (menuCH) and Multiple Swiss Health Datasets Acquired from 1992 to 2012 for Data Mining Purposes 150
Timo Lustenberger, Helena Jenzer and Farshideh Einsele

Data Mining for Animal Health to Improve Human Quality of Life: Insights from a University Veterinary Hospital 157
Oscar Tamburis, Elio Masciari, Christian Esposito and Gerardo Fatone

DERM: A Reference Model for Data Engineering 165
Daniel Tebernum, Marcel Altendeitering and Falk Howar

GRASP: Graph-based Mining of Scientific Papers 176
Navid Nobani, Mauro Pelucchi, Matteo Perico, Andrea Scrivanti and Alessandro Vaccarino

Motif-based Classification using Enhanced Sub-Sequence-Based Dynamic Time Warping 184
Mohammed Alshehri, Frans Coenen and Keith Dures

Biomedical Dataset Recommendation 192
Xu Wang, Frank van Harmelen and Zhisheng Huang

A Comparison of Methods for the Evaluation of Text Summarization Techniques 200
Marcello Barbella, Michele Risi and Genoveffa Tortora

Deep Learning for RF-based Drone Detection and Identification using Welch's Method 208
Mahmoud Almasri

DQ-MeeRKat: Automating Data Quality Monitoring with a Reference-Data-Profile-Annotated Knowledge Graph 215
Lisa Ehrlinger, Alexander Gindlhumer, Lisa-Marie Huber and Wolfram Wöß

A Survey of Social Emotion Prediction Methods 223
Abdullah Alsaedi, Phillip Brooker, Floriana Grasso and Stuart Thomason

WFDU-net: A Workflow Notation for Sovereign Data Exchange 231
Heinrich Pettenpohl, Daniel Tebernum and Boris Otto

A Network based Approach for Reducing Variant Diversity in Production Planning and Control 241
Shailesh Tripathi, Sonja Strasser and Herbert Jodlbauer

Predicting Shopping Intent of e-Commerce Users using LSTM Recurrent Neural Networks 252
Konstantinos Diamantaras, Michail Salampasis, Alkiviadis Katsalis and Konstantinos Christantonis

Invers Natural Number System to Maintain User-defined Sequence of Data Records 260
Seyfettin Öztürk

Using BPMN for ETL Conceptual Modelling: A Case Study 267
Bruno Oliveira, Óscar Oliveira and Orlando Belo

Well-Being in Plastic Surgery: Deep Learning Reveals Patients' Evaluations 275
Joschka Kersting and Michaela Geierhos

Applied Feature-oriented Project Life Cycle Classification 285
Oliver Böhme and Tobias Meisen

Semantic Enrichment of Vital Sign Streams through Ontology-based Context Modeling using Linked Data Approach 292
Sachiko Lim, Rahim Rahmani and Paul Johannesson

Data Driven Hybrid Approach for Health Monitoring and Fault Detection in Military Ground Vehicles 300
Indu Shukla, Antoinette Silas, Haley Dozier, Brandon E. Hansen and W. Glenn Bond

Impact of Duplicating Small Training Data on GANs 308
Yuki Eizuka, Kazuo Hara and Ikumi Suzuki

Making Data Big for a Deep-learning Analysis: Aggregation of Public COVID-19 Datasets of Lung Computed Tomography Scans 316
Francesca Lizzi, Francesca Brero, Raffaella Fiamma Cabini, Maria Evelina Fantacci, Stefano Piffer, Ian Postuma, Lisa Rinaldi and Alessandra Retico

Toward a Multimodal Multitask Model for Neurodegenerative Diseases Diagnosis and Progression Prediction 322
Sofia Lahrichi, Maryem Rhanoui, Mounia Mikram and Bouchra El Asri

Estimating Territory Risk Relativity for Auto Insurance Rate Regulation using Generalized Linear Mixed Models 329
Shengkun Xie, Chong Gan and Clare Chua-Chow

Knowledge Graph Analysis of Russian Trolls 335
Chih-yuan Li, Soon Ae Chun and James Geller

Aspect Based Sentiment Analysis on Online Review Data to Predict Corporate Reputation 343
R. E. Loke and W. Reitter

Evo-Path: Querying Data Evolution through Complex Changes 353
Theodora Galani, Yannis Stavrakas, George Papastefanatos and Yannis Vassiliou

Enhanced AI On-the-Edge 3D Vision Accelerated Point Cloud Spatial Computing Solution 362
Gaurav Kumar Wankar and Shubham Vohra

Tailoring Taint Analysis for Database Applications in the K Framework 370
Md. Imran Alam and Raju Halder

A Company's Corporate Reputation through the Eyes of Employees Measured with Sentiment Analysis of Online Reviews 378
R. E. Loke and R. Lam-Lion

AUTHOR INDEX 387