Lecture Notes in Bioinformatics

Subseries of Lecture Notes in Computer Science

LNBI Series Editors

Sorin Istrail Brown University, Providence, RI, USA
Pavel Pevzner University of California, San Diego, CA, USA
Michael Waterman University of Southern California, Los Angeles, CA, USA

LNBI Editorial Board

Søren Brunak Technical University of Denmark, Kongens Lyngby, Denmark Mikhail S. Gelfand IITP, Research and Training Center on Bioinformatics, Moscow, Russia Thomas Lengauer Max Planck Institute for Informatics, Saarbrücken, Germany Satoru Miyano University of Tokyo, Tokyo, Japan Eugene Myers Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany Marie-France Sagot Université Lyon 1, Villeurbanne, France David Sankoff University of Ottawa, Ottawa, Canada Ron Shamir Tel Aviv University, Ramat Aviv, Tel Aviv, Israel Terry Speed Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia Martin Vingron Max Planck Institute for Molecular Genetics, Berlin, Germany W. Eric Wong University of Texas at Dallas, Richardson, TX, USA

More information about this series at http://www.springer.com/series/5381

Data Integration in the Life Sciences

13th International Conference, DILS 2018 Hannover, Germany, November 20–21, 2018 Proceedings



Editors Sören Auer TIB and Leibniz University Hannover, Germany

Maria-Esther Vidal D TIB and Leibniz University Hannover, Germany

 ISSN 0302-9743
 ISSN 1611-3349 (electronic)

 Lecture Notes in Bioinformatics
 ISBN 978-3-030-06015-2
 ISBN 978-3-030-06016-9 (eBook)

 https://doi.org/10.1007/978-3-030-06016-9
 ISBN 978-3-030-06016-9
 ISBN 978-3-030-06016-9

Library of Congress Control Number: 2018964607

LNCS Sublibrary: SL8 - Bioinformatics

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume comprises the proceedings of the 13th International Conference on Data Integration in the Life Sciences (DILS 2018), held in Hannover (Germany) during November 20–21, 2018. DILS 2018 was hosted by TIB Leibniz Information Centre for Science and Technology, L3S Research Center at Leibniz University of Hannover, the Information Centre for Life Sciences (ZB MED) and the Hannover Medical School (MHH).

The articles included in this volume went through a peer-review process where each submission was reviewed by at least three reviewers and one senior program chair. The submissions were evaluated in terms of relevance, novelty, significance, soundness, and quality of the presentation. Three types of submissions were received: (1) full papers describing solid and complete research contributions; (2) short papers presenting results of on-going research work; and (3) poster and demonstration papers. We accepted five full papers; eight short papers; four poster papers; and four demo papers. Our sincere thanks go to the Program Committee members and external reviewers for their valuable input, and for accepting our invitation to contribute to the review process.

The DILS 2018 submissions cover a wide variety of topics related to data management in the life sciences. The articles tackle open problems and technical solutions for data integration, query processing, and analytics on big life science data coming from diverse data sources, e.g., genomic data collections, biomedical literature, or clinical records, and in the challenges of transforming big data into actionable insights.

We composed an exciting program that included four research sessions: (1) Big Biomedical Data Integration and Management; (2) Data Exploration in the Life Sciences; (3) Biomedical Data Analytics; and (4) Big Biomedical Applications. Additionally, the program included two invited talks; the first invited talk was on "Matching Biomedical Ontologies for Semantic Data Integration" by Dr. Catia Pesquita, and the second invited talk was on "The de.NBI network–A Bioinformatics Infrastructure in Germany for Handling Big Data in Life Sciences" by Prof. Alfred Pühler. Posters and demos were presented in a plenary session where the authors and attendees had the opportunity to interact in an informal environment.

November 2018

Sören Auer Maria-Esther Vidal

Organization

General Chair

Sören Auer	TIB Leibniz Information Centre for Science and Technology, and L3S Research Center at Leibniz University of Hannover, Germany
Program Chair	
Maria-Esther Vidal	TIB Leibniz Information Centre for Science and Technology, and L3S Research Center at Leibniz University of Hannover, Germany
Organizing Co-chairs	
Thomas Illig Wolfgang Nejdl	Hannover Medical School (MHH), Germany L3S Research Center at Leibniz University of Hannover, Germany
Dietrich Nelle	ZB-MED Information Centre for Life Sciences, Germany
Local Organizers	
Alexandra Garatzogianni	TIB Leibniz Information Centre for Science and Technology, and L3S Research Center at the Leibniz University of Hannover, Germany
Katrin Hanebutt	TIB Leibniz Information Centre for Science and Technology, Germany
Program Committee	
Maribel Acosta José Luis Ambite Naveen Ashish Diego Collarana Benjamin Lang Marcos Da Silveira Michel Dumontier Kemele Endris Juliane Eluck	 AIFB, Karlsruhe Institute of Technology, Germany University of Southern California, USA InferLink Corporation, USA University of Bonn, Germany Centre for Genomic Regulation (CRG), Spain Luxembourg Institute of Science and Technology, Luxembourg Maastricht University, The Netherlands L3S Research Centre at the Leibniz University of Hannover, Germany Fraunhofer SCAL Germany

Konrad Förstner	University of Würzburg, Germany
Mikhail Galkin	University of Bonn, Germany
Martin Gaedke	Chemnitz University of Technology, Germany
Matthias Gietzelt	Hannover Medical School (MHH), Germany
Irlan Grangel	University of Bonn, Germany
Anika Groß	University of Leipzig, Germany
Udo Hahn	Jena University, Germany
Robert Hoehndorf	King Abdullah University of Science and Technology, Saudi Arabia
Patrick Lambrix	Linköping University, Sweden
Pierre Larmande	Institute of Research for Development, France
Ulf Leser	Humboldt-Universität zu Berlin, Germany
Angeli Möller	Bayer Business Services, Germany
Bernd Müller	ZB MED Leibniz Information Centre for Life Sciences,
	Germany
Isaiah Mulang	University of Bonn, Germany
Guillermo Palma	L3S Research Centre at the Leibniz University
	of Hannover, Germany
Catia Pesquita	University of Lisbon, Portugal
Cédric Pruski	Luxembourg Institute of Science and Technology, Luxembourg
Erhard Rahm	University of Leipzig, Germany
Alejandro Rodriguez	Universidad Politécnica de Madrid, Spain
Kuldeep Singh	Fraunhofer IAIS, Germany
Dietrich	Information Centre for Life Sciences (ZB MED),
Rebholz-Schuhmann	Germany
Andreas Thor	University of Applied Sciences
	for Telecommunications Leipzig, Germany
Johanna Völker	Bayer Business Services, Germany
Amrapali Zaveri	Maastricht University, The Netherlands

Sponsoring Institutions

TIB Leibniz Information Centre for Science and Technology, Germany

Contents

Big Biomedical Data Integration and Management

Do Scaling Algorithms Preserve Word2Vec Semantics? A Case Study	
for Medical Entities Janus Wawrzinek, José María González Pinto, Philipp Markiewka, and Wolf-Tilo Balke	3
Combining Semantic and Lexical Measures to Evaluate Medical Terms Similarity	17
Construction and Visualization of Dynamic Biological Networks: Benchmarking the Neo4J Graph Database Lena Wiese, Chimi Wangmo, Lukas Steuernagel, Armin O. Schmitt, and Mehmet Gültas	33
A Knowledge-Driven Pipeline for Transforming Big Data into Actionable Knowledge	44
Leaving No Stone Unturned: Using Machine Learning Based Approaches for Information Extraction from Full Texts of a Research Data Warehouse Johanna Fiebeck, Hans Laser, Hinrich B. Winther, and Svetlana Gerbel	50
Data Exploration in the Life Sciences	
Towards Research Infrastructures that Curate Scientific Information: A Use Case in Life Sciences Markus Stocker, Manuel Prinz, Fatemeh Rostami, and Tibor Kempf	61
Interactive Visualization for Large-Scale Multi-factorial Research Designs Andreas Friedrich, Luis de la Garza, Oliver Kohlbacher, and Sven Nahnsen	75
FedSDM: Semantic Data Manager for Federations of RDF Datasets Kemele M. Endris, Maria-Esther Vidal, and Sören Auer	85

Poster Paper Data Integration for Supporting Biomedical Knowledge	
Graph Creation at Large-Scale	91
Samaneh Jozashoori, Tatiana Novikova, and Maria-Esther Vidal	
DISBi: A Flexible Framework for Integrating Systems Biology Data Rüdiger Busche, Henning Dannheim, and Dietmar Schomburg	97

Biomedical Data Analytics

Using Machine Learning to Distinguish Infected from Non-infected Subjects at an Early Stage Based on Viral Inoculation Ghanshyam Verma, Alokkumar Jha, Dietrich Rebholz-Schuhmann, and Michael G. Madden	105
Automated Coding of Medical Diagnostics from Free-Text: The Role of Parameters Optimization and Imbalanced Classes Luiz Virginio and Julio Cesar dos Reis	122
A Learning-Based Approach to Combine Medical Annotation Results (Short Paper)	135
Knowledge Graph Completion to Predict Polypharmacy Side Effects Brandon Malone, Alberto García-Durán, and Mathias Niepert	144
Big Biomedical Applications	
Lung Cancer Concept Annotation from Spanish Clinical Narratives Marjan Najafabadipour, Juan Manuel Tuñas, Alejandro Rodríguez-González, and Ernestina Menasalvas	153
Linked Data Based Multi-omics Integration and Visualization for Cancer Decision Networks	164
The Hannover Medical School Enterprise Clinical Research Data Warehouse: 5 Years of Experience Svetlana Gerbel, Hans Laser, Norman Schönfeld, and Tobias Rassmann	182
User-Driven Development of a Novel Molecular Tumor Board Support Tool	195

Using Semantic Programming for Developing a Web Content Management System for Semantic Phenotype Data	
and Peter Grobe	
Converting Alzheimer's Disease Map into a Heavyweight Ontology: A Formal Network to Integrate Data	207
Author Index	217