

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>

Marcos Da Silveira · Cédric Pruski
Reinhard Schneider (Eds.)


Data Integration in the Life Sciences

12th International Conference, DILS 2017
Luxembourg, Luxembourg, November 14–15, 2017
Proceedings

Editors

Marcos Da Silveira
Luxembourg Institute of Science
and Technology
Esch-sur-Alzette
Luxembourg

Reinhard Schneider
Université du Luxembourg
Esch-sur-Alzette
Luxembourg

Cédric Pruski 
Luxembourg Institute of Science
and Technology
Esch-sur-Alzette
Luxembourg

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Bioinformatics
ISBN 978-3-319-69750-5 ISBN 978-3-319-69751-2 (eBook)
<https://doi.org/10.1007/978-3-319-69751-2>

Library of Congress Control Number: 2017956781

LNCS Sublibrary: SL8 – Bioinformatics

© Springer International Publishing AG 2017

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.

Printed on acid-free paper

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

Preface

The 12th edition of the international conference on Data Integration in the Life Sciences (DILS), organized by the Luxembourg Institute of Science and Technology, was held in Hotel Parc-Bellevue during November 14–15, 2017, in Luxembourg. This year again, DILS 2017 brought together researchers (including students), developers, users, and practitioners involved in the integration, management, and analysis of heterogeneous data in the life sciences. DILS 2017 was a forum for computer scientists and members of different life science communities including bioinformatics, biodiversity, medicine, clinical health management, and pharmacy. This year, topics focused on life science data modeling and the management of life science datasets and models.

The call for papers attracted excellent contributions and a rigorous reviewing process was conducted to select the 10 best research and demo papers. Moreover, we had the pleasure to welcome two renowned speakers: Prof. Robert Stevens from the University of Manchester and Dr. Alex Bateman, Head of Protein Sequence Resources cluster of the European Bioinformatics Institute (EMBL-EBI). In his presentation, Prof. Stevens discussed the advances in the use of ontologies in the biomedical domain while Dr. Bateman discussed the UniProt Knowledge-Base, which has been providing data for protein sequence and function for over 30 years.

As the event co-chairs and editors of this volume, we would like to thank all authors who submitted papers, as well as the Program Committee members and additional reviewers for their excellent contribution in evaluating the submissions. Special thanks to Marylène Martin and the Luxembourg Institute of Science and Technology for their precious support in the organization of this event. Finally, we would like to thank Alfred Hofmann and his team at Springer for their cooperation and help in putting this volume together.

September 2017

Marcos Da Silveira
Cédric Pruski
Reinhard Schneider

Organization

Conference Chairs

Marcos Da Silveira	Luxembourg Institute of Science and Technology, Luxembourg
Cédric Pruski	Luxembourg Institute of Science and Technology, Luxembourg
Reinhard Schneider	University of Luxembourg, Luxembourg

Program Committee

Jose-Luis Ambite	University of Southern California, USA
Sarah Cohen-Boulakia	University of Paris-Sud, France
Stefan Darmoni	Rouen University Hospital, France
Juliette Dibie	AgroParisTech, France
Julio Cesar Dos Reis	University of Campinas, Brazil
Helena Galhardas	University of Lisbon, Portugal
Anika Gross	University of Leipzig, Germany
Wei Gu	University of Luxembourg, Luxembourg
Zhisheng Huang	Vrije Universiteit Amsterdam, The Netherlands
Graham Kemp	Chalmers University of Technology, Sweden
Andreas Kremer	ITTM, Luxembourg
Patrick Lambrix	Linköping University, Sweden
Ulf Leser	Humboldt-Universität zu Berlin, Germany
Dilvan Moreira	University of São Paulo, Brazil
Patrick Pirrotte	TGen, USA
Louïqa Raschid	University of Maryland, USA
Erhard Rahm	University of Leipzig, Germany
Patrick Ruch	University of Applied Sciences Western Switzerland, Switzerland
Venkata Satagopam	University of Luxembourg, Luxembourg
Annette ten Teije	Vrije Universiteit Amsterdam, The Netherlands
Christophe Trefois	University of Luxembourg, Luxembourg

Additional Reviewer

Ying-Chi Lin	University of Leipzig, Germany
--------------	--------------------------------

Contents

Evaluating and Improving Annotation Tools for Medical Forms	1
<i>Ying-Chi Lin, Victor Christen, Anika Groß, Silvio Domingos Cardoso, Cédric Pruski, Marcos Da Silveira, and Erhard Rahm</i>	
A Simple Tool to Enrich Clinical Trial Data with Multiontology-Based Conceptual Tags	17
<i>Holger Stenzhorn</i>	
Variant-DB: A Tool for Efficiently Exploring Millions of Human Genetic Variants and Their Annotations	22
<i>Joachim Kutzera and Patrick May</i>	
Semantic Interoperability of Clinical Data	29
<i>Jacqueline Midlej do Espírito Santo and Claudia Bauzer Medeiros</i>	
An Integrated Ontology-Based Approach for Patent Classification in Medical Engineering	38
<i>Sandra Geisler, Christoph Quix, Rihan Hai, and Sanchit Alekh</i>	
A Conceptual Approach for Modelling Social Care Services: The INSPIRE Project.	53
<i>Elaheh Pourabbas, Antonio D'Uffizi, and Fabrizio L. Ricci</i>	
DyKOSMap: From a Research Prototype to a Web Application	67
<i>Marwa Chaabane, Silvio Domingos Cardoso, Cédric Pruski, and Marcos Da Silveira</i>	
Layout-Aware Semi-automatic Information Extraction for Pharmaceutical Documents	71
<i>Simon Harmata, Katharina Hofer-Schmitz, Phuong-Ha Nguyen, Christoph Quix, and Bujar Bakiu</i>	
Enabling Data Integration Using MIPMap	86
<i>Giorgos Stoilos, Despoina Trivela, Vasilis Vassalos, Tassos Venetis, and Yannis Xarchakos</i>	
Interactive Map Visualization System Based on Integrated Semi-structured and Structured Healthcare Data	94
<i>Milena Frtunić Gligorijević, Darko Puflović, Evgenija Stevanoska, Tatjana Jevtović Stoimenov, Goran Velinov, and Leonid Stoimenov</i>	
Author Index	109