Lecture Notes in Bioinformatics 4544

Edited by S. Istrail, P. Pevzner, and M. Waterman

Subseries of Lecture Notes in Computer Science

Sarah Cohen-Boulakia Val Tannen (Eds.)

Data Integration in the Life Sciences

4th International Workshop, DILS 2007
Philadelphia, PA, USA, June 27-29, 2007
Proceedings

Volume Editors

Sarah Cohen-Boulakia
University of Pennsylvania, Department of Computer and Information Science
303 Levine Hall, 3330 Walnut St., Philadelphia, PA 19104, USA
E-mail: sarahcb@seas.upenn.edu

Val Tannen
University of Pennsylvania, Department of Computer and Information Science
570 Levine Hall, 3330 Walnut St., Philadelphia, PA 19104, USA
E-mail: val@cis.upenn.edu

Library of Congress Control Number: 2007928915

CR Subject Classification (1998): H.2, H.3, H.4, J.3

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-540-73254-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-73254-9 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12081769 06/3180 5 4 3 2 1 0

Preface

Understanding the mechanisms involved in life (e.g., discovering the biological function of a set of proteins, inferring the evolution of a set of species) is becoming increasingly dependent on progress made in mathematics, computer science, and molecular engineering. For the past 30 years, new high-throughput technologies have been developed generating large amounts of data, distributed across many data sources on the Web, with a high degree of semantic heterogeneity and different levels of quality. However, one such dataset is not, by itself, sufficient for scientific discovery. Instead, it must be combined with other data and processed by bioinformatics tools for patterns, similarities, and unusual occurrences to be observed. Both data integration and data mining are thus of paramount importance in life science.

DILS 2007 was the fourth in a workshop series that aims at fostering discussion, exchange, and innovation in research and development in the areas of data integration and data management for the life sciences. Each previous DILS workshop attracted around 100 researchers from all over the world. This year, the number of submitted papers again increased. The Program Committee selected 19 papers out of 52 full submissions. The DILS 2007 papers cover a wide spectrum of theoretical and practical issues including scientific workflows, annotation in data integration, mapping and matching techniques, and modeling of life science data. Among the papers, we distinguished 13 papers presenting research on new models, methods, or algorithms and 6 papers presenting implementation of systems or experience with systems in practice. In addition to the presented papers, DILS 2007 featured two keynote talks by Kenneth H. Buetow, National Cancer Institute, and Junhyong Kim, University of Pennsylvania.

The workshop was held at the University of Pennsylvania, in Philadelphia, USA. It was kindly sponsored by the School of Engineering and Applied Science of the University of Pennsylvania, the Penn Genomics Institute, and Microsoft Research, who also made available their conference management system. As editors of this volume, we thank all the authors who submitted papers, the Program Committee members, and the external reviewers for their excellent work. Special thanks go to Susan Davidson, General Chair, Chris Stoeckert, PC Co-chair, as well as Olivier Biton, Tara Betterbid, and Howard Bilowsky. Finally, we are grateful for the cooperation and help of Springer in putting this volume together.

June 2007

Sarah Cohen-Boulakia
Val Tannen

Organization

Executive Committee

General Chair		
	Susan Davisdon,	University of Pennsylvania, USA
Program Chairs		
	Chris Stoeckert,	University of Pennsylvania, USA
	Val Tannen,	University of Pennsylvania, USA

Program Committee

Judith Blake	Jackson Laboratory, USA
Sarah Cohen-Boulakia	University of Pennsylvania, USA
Marie-Dominique Devignes	LORIA, Nancy, France
Barbara Eckman	IBM
Christine Froidevaux	LRI, University of Paris-Sud XI, France
Cesare Furlanello	ITC-irst, Trento, Italy
Jim French	University of Virginia, USA
Floris Geerts	University of Edinburgh, UK
Amarnath Gupta	University of California San Diego, USA
Ela Hunt	ETH Zurich, Switzerland
Jacob Koehler	Rothamsted Research, UK
Anthony Kosky	Axiope Inc.
Hilmar Lapp	NESCENT
Ulf Leser	Humboldt-Universität zu Berlin, Germany
Bertram Ludäscher	University of California Davis, USA
Victor Markowitz	Lawrence Berkeley Labs
Peter Mork	MITRE
Tom Oinn	European Bioinformatics Institute, UK
Meral Ozsoyoglu	Case Western Reserve University, USA
John Quackenbush	Harvard, USA
Louiqa Raschid	University of Maryland, USA
Fritz Roth	Harvard Medical School, USA
Susanna-Assunta Sansone	European Bioinformatics Institute, UK
Kai-Uwe Sattler	Technical University of Ilmena, Germany
Chris Stoeckert	University of Pennsylvania, USA
Val Tannen	University of Pennsylvania, USA
Olga Troyanskaya	Princeton University, USA

External Reviewers

Jérôme Azé
Jana Bauckmann
Julie Bernauer
Shawn Bowers
William Bug
Amy Chen
Adnan Derti
Francis Gibbons
Philip Groth
Timothy McPhillips
Joe Mellor
Krishna Palaniappan
Norbert Podhorszki
Murat Tasan
Weidong Tian
Silke Trißl
Daniel Zinn

Sponsorship Chair

Howard Bilofsky, University of Pennsylvania

Sponsoring Institutions

School of Engineering and Applied Science at the University of Pennsylvania
http://www.seas.upenn.edu/

Penn Genomics Institute
http://www.genomics.upenn.edu/

Microsoft Research
http://research.microsoft.com/

Web Site and Publicity Chairs

Olivier Biton	University of Pennsylvania
Sarah Cohen-Boulakia	University of Pennsylvania
DILS 2007 Web site	http://dils07.cis.upenn.edu/

Table of Contents

Keynote Presentations

New Architectures and Experience on Using Systems

Managing and Designing Scientific Workflows

Mapping and Matching Techniques

Modeling of Life Science Data

Annotation in Data Integration