Preface

This volume of *Lecture Notes in Bioinformatics (LNBI)* contains selected papers from the workshop of the BioLINK Special Interest Group (SIG) on Linking Literature, Information and Knowledge for Biology. The workshop was held June 28–29, 2009 at the ISMB/ECCB 2009 conference in Stockholm (http://www.iscb.org/ismbeccb2009).

The BioLINK SIG meeting has been regularly held in association with the ISMB conferences since 2001, focusing on the development and application of resources and tools for biomedical text mining. The SIG is interdisciplinary in nature, and brings together researchers applying natural language processing, text mining, information extraction and retrieval in the biomedical domain, with scientists from bioinformatics and biology. This year the SIG included two special sessions, one on the analysis of images and figures, which play a critical role in scientific papers, and the other on the future of scientific publication. Overview papers concerned with both these topics are included in this volume.

The meeting featured invited talks, presentations of peer-reviewed contributed papers, reports from recent evaluations and workshops, as well as a poster session. The SIG solicited papers that discuss current biological image- and text-related needs, the challenges of meeting them, and the tools that may help address them. Specific topics include: biological image analysis and its relationship to biomedical text mining, application of text-mining tools to biomedical research, automated indexing of the biomedical literature, generation of structured digital abstracts, new evaluation measures to measure utility and usability of text mining tools integrated into the biologist’s workflow, as well as new tools and applications, and future directions in biomedical text mining. The publications in this volume were selected from the workshop papers and represent extended versions that have undergone an additional review process by the members of the Program Committee.

We thank the Program Committee for thoroughly reviewing the submitted manuscripts and for their valuable comments that significantly improved the final publications.

December 2009

Christian Blaschke
Hagit Shatkay
Organization

Workshop Program Chairs

Christian Blaschke  
Bioalma, Spain

Lynette Hirschman  
MITRE Corporation, USA

Scott Markel  
Accelrys, USA

Robert Murphy  
Carnegie Mellon University, USA

Hagit Shatkay  
School of Computing, Queen’s University, Canada

Alfonso Valencia  
Spanish National Cancer Research Centre, Spain

Program Committee

Alan Aronson  
NLM, USA

Olivier Bodenreider  
NLM, USA

Monica Chagoyen  
Spanish National Centre for Biotechnology, Spain

Aaron Cohen  
Oregon Health Sciences University, USA

Kevin B. Cohen  
University of Colorado, Health Sciences Center, USA

Mark Craven  
University of Wisconsin, Madison, USA

Bob Futrelle  
Northeastern University, USA

William Hayes  
Biogen IDEC, USA

Martin Krallinger  
Spanish National Cancer Research Centre, Spain

Alberto Pascual Montano  
Complutense University of Madrid, Spain

Andreas Rechsteiner  
University of California, Santa Cruz, USA

Luis Rocha  
Indiana University, Bloomington, USA

Karin Verspoor  
Los Alamos National Laboratories, USA

John Wilbur  
NCBI, NLM, USA

Hong Yu  
University of Wisconsin, Milwaukee, USA

Ralf Zimmer  
Ludwig-Maximilians University, Munich, Germany

Sponsoring Institutions

We gratefully acknowledge:
The International Society for Computational Biology (ISCB)
The Spanish National Cancer Research Centre
BioCreative II.5 challenge
Table of Contents

BioLINK 2009

Overview of the Ninth Annual Meeting of the BioLINK SIG at ISMB:
Linking Literature, Information and Knowledge for Biology .......... 1
    Christian Blaschke, Lynette Hirschman, Hagit Shatkay, and
    Alfonso Valencia

Principles of Bioimage Informatics: Focus on Machine Learning of Cell
Patterns ............................................................... 8
    Luis Pedro Coelho, Estelle Glory-Afshar, Joshua Kangas,
    Shannon Quinn, Aabid Shariff, and Robert F. Murphy

Summary of the BioLINK Special Interest Group Session on the Future
of Scientific Publishing ........................................... 19
    Scott Markel

Structured Literature Image Finder: Extracting Information from Text
and Images in Biomedical Literature ................................. 23
    Luis Pedro Coelho, Amr Ahmed, Andrew Arnold, Joshua Kangas,
    Abdul-Saboor Sheikh, Eric P. Xing, William W. Cohen, and
    Robert F. Murphy

Toward Computer-Assisted Text Curation: Classification Is Easy
(Choosing Training Data Can Be Hard...) ......................... 33
    Robert Denroche, Ramana Madupu, Shibu Yooseph,
    Granger Sutton, and Hagit Shatkay

Mining Protein-Protein Interactions from GeneRIFs with
OpenDMAP ............................................................... 43
    Andrew D. Fox, William A. Baumgartner Jr., Helen L. Johnson,
    Lawrence E. Hunter, and Donna K. Slonim

Combining Semantic Relations and DNA Microarray Data for Novel
Hypotheses Generation ............................................. 53
    Dimitar Hristovski, Andrej Kastrin, Borut Peterlin, and
    Thomas C. Rindflesch

Learning from Positive and Unlabeled Documents for Retrieval of
Bacterial Protein-Protein Interaction Literature ..................... 62
    Hongfang Liu, Manabu Torii, Guixian Xu, Zhangzhi Hu, and
    Johannes Goll
Table of Contents

Extracting and Normalizing Gene/Protein Mentions with the Flexible and Trainable Moara Java Library ........................................ 71

Mariana L. Neves, José Maria Carazo, and Alberto Pascual-Montano

Author Index ............................................................. 81