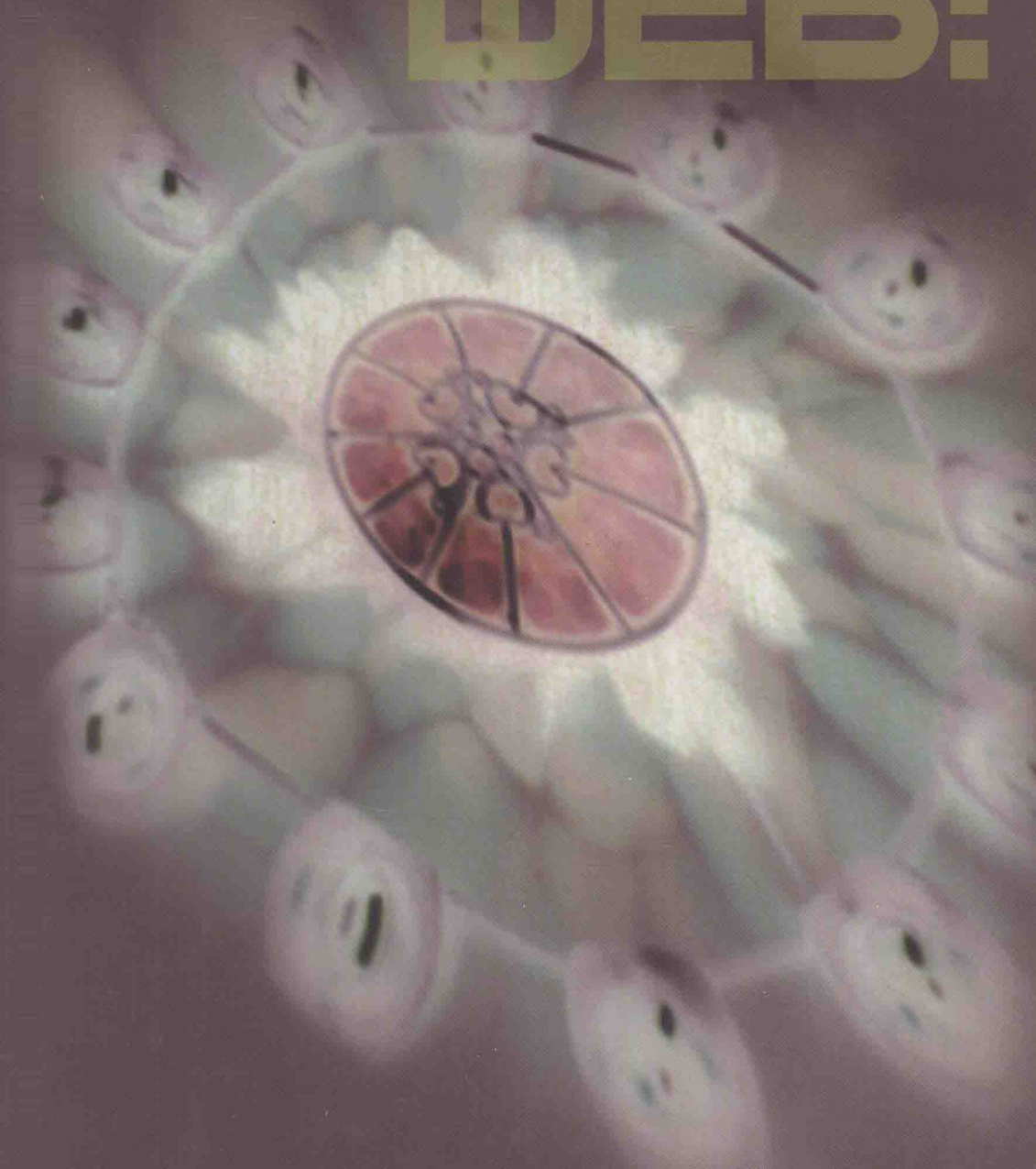


SEMANTIC WEB:



REVOLUTIONIZING KNOWLEDGE
DISCOVERY IN THE LIFE SCIENCES

SEMANTIC WEB

Revolutionizing Knowledge Discovery in the Life Sciences

edited by

Christopher I. D. Baker

Knowledge Discovery Department, Institute for Infocomm Research
Singapore

and

Kei-Hoi Cheung

Center for Medical Informatics, Yale University School of Medicine
New Haven, CT, USA



Springer

Library of Congress Control Number: 2006935819

ISBN-13: 978-0-387-48436-5

c-ISBN-13: 979-0-387-48438-9

ISBN-10: 0-387-48436-1

c-ISBN-10: 0-387-48438-8

Printed on acid-free paper.

© 2007 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com

SEMANTIC WEB
Revolutionizing Knowledge Discovery
in the Life Sciences

Dedication

*This book is dedicated to Iris,
Rahma, Irmengard and Barrie
for sharing with me their love,
support and wisdom.*

Christopher J. O. Baker

*This book is dedicated to Candy
(my wife) for her loving support
and to Ian (my son) who is a
special gift of love in my life.*

Kei-Hoi Cheung

Contributing Authors

Almeida, Jonas

Department of Biostatistics and Applied Mathematics
The University of Texas, USA

Alper, Pinar

School of Computer Science
University of Manchester, UK

Baker[#], Christopher J. O

Knowledge Discovery Department
Institute for Infocomm Research, Singapore
cbaker@i2r.a-star.edu.sg, baker@encs.concordia.ca

Belhajjame, Khalid

School of Computer Science
University of Manchester, UK

Bodenreider, Olivier

National Library of Medicine
National Institutes of Health, USA

Bry, Francois

Institute for Informatics
University of Munich, Germany

Burger*, Albert

Human Genetics Unit, Medical Research Council, UK
Department of Computer Science, Heriot-Watt University, UK
ab@macs.hw.ac.uk

Ceusters, Werner

Center of Excellence in Bioinformatics and Life Sciences
State University of New York at Buffalo, USA

Cheung^{#*}, Kei-Hoi

Center for Medical Informatics
Yale University School of Medicine, USA
kei.cheung@yale.edu

De Roure, David

School of Electronics and Computer Science
University of Southampton, UK

Furch, Tim

Institute for Informatics
University of Munich, Germany

Gerstein, Mark B.

Molecular Biophysics and Biochemistry
Yale University, USA

Goble*, Carole

School of Computer Science
University of Manchester, UK
carole@cs.man.ac.uk

Goderis, Antoon

School of Computer Science
University of Manchester, UK

Greenbaum*, Dov

School of Law
University of California, Berkeley, USA
dsg28@boalthall.berkeley.edu

Haarslev, Volker

Department of Computer Science and Software Engineering
Concordia University, Canada

Hayes, William

Biogen Idec, USA

Hirschman*, Lynette

The MITRE Corporation, USA

lynette@mitre.org

Hongsermeier, Tonya

Clinical Informatics R&D

Partners HealthCare System, USA

Hull, Duncan

School of Computer Science

University of Manchester, UK

Hunter, Blake

Computer Science Department and

Complex Carbohydrate Research Center (CCRC)

University of Georgia, USA

Jakoniene, Vaida

Department of Computer and Information Science

Linköpings Universitet, Sweden

Kappler, Thomas

Institut für Programmstrukturen und Datenorganisation

Universität Karlsruhe (TH), Germany

Kashyap*, Vipul

Clinical Informatics R&D

Partners HealthCare System, USA

vkashyap1@partners.org

Kazic*, Toni

Department of Computer Science

University of Missouri, USA

toni@athe.mnet.missouri.edu

King, Ross D.

The Computer Science Department

The University of Wales, Aberystwyth, UK

Krivov*, Serguei

Department of Computer Science and
Gund Institute for Ecological Economics
The University of Vermont, USA
Serguei.Krivov@uvm.edu

Lambrix*, Patrick

Department of Computer and Information Science
Linköpings Universitet, Sweden
patla@ida.liu.se

Linse, Benedikt

Institute for Informatics
University of Munich, Germany

Lord, Phillip

School of Computing Science
University of Newcastle upon Tyne, UK

Lussier*, Yves

Department of Medicine
The University of Chicago, USA and
Department of Biomedical Informatics and
College of Physicians and Surgeons, Columbia University, USA
lussier@uchicago.edu

Mani, Inderjeet

The MITRE Corporation, USA

Obrst*, Leo

The MITRE Corporation, USA
lobrst@mitre.org

Oinn, Tom

The European Bioinformatics Institute, UK

Pan*, Jeff

Department of Computing Science
The University of Aberdeen, UK
jpan@csd.abdn.ac.uk

Ray, Steve

US National Institute of Standards and Technology, USA

Royer, Loïc

Biotec

Dresden University of Technology, Germany

Sahoo*, Satya S.

Computer Science Department and

Complex Carbohydrate Research Center (CCRC)

University of Georgia, USA

satya30@uga.edu

Schroeder*, Michael

Biotec

Dresden University of Technology, Germany

ms@biotec.tu-dresden.de

Sheth, Amit P.

Computer Science Department and

Complex Carbohydrate Research Center (CCRC)

University of Georgia, USA

Smith, Andrew K.

Computer Science

Yale University, USA

Smith, Barry

Department of Philosophy

State University of New York at Buffalo, USA

Soldatova*, Larisa N.

The Computer Science Department

The University of Wales, Aberystwyth, UK

lss@aber.ac.uk

Stevens, Robert

School of Computer Science

University of Manchester, UK

Strömbäck, Lena

Department of Computer and Information Science

Linköpings Universitet, Sweden

Tan, He

Department of Computer and Information Science

Linköpings Universitet, Sweden

Turi, Daniele

School of Computer Science
University of Manchester, UK

Valencia, Alfonso

Spanish National Cancer Research Center, Spain

Villa, Ferdinando

The Botany Department and
Gund Institute for Ecological Economics
The University of Vermont, USA

Wächter, Thomas

Biotec
Dresden University of Technology, Germany

Wang*, Xiaoshu

Department of Biostatistics, Bioinformatics and Epidemiology
Medical University of South Carolina, USA
wangxiao@musc.edu

Williams, Richard

Rocky Mountain Biological Laboratory, USA

Witte*, René

Institut für Programmstrukturen und Datenorganisation
Universität Karlsruhe (TH), Germany
witte@ipd.uka.de

Wolstencroft*, Katy

School of Computer Science
University of Manchester, UK
kwolstencroft@cs.man.ac.uk

Wroe, Chris

British Telecom, UK

Wu, Xindong

Department of Computer Science
The University of Vermont, USA

Yip, Kevin Y. L.

Computer Science
Yale University, USA

York, S. William

Computer Science Department and
Complex Carbohydrate Research Center (CCRC)
University of Georgia, USA

Zhao, Jun

School of Computer Science
University of Manchester, UK

***Corresponding Authors
#Editors**

Preface

The rapid growth of the Web has led to the proliferation of information sources and content accessible via the Internet. While improvements in hardware capabilities continue to help the speed and the flow of information across networked computers, there remains a major problem for the human user to keep up with the rapid expansion of the Web information space. Although there is plenty of room for computers to help humans to discover, navigate, and integrate information in this vast information space, the way the information is currently represented and structured through the Web is not easily readable to computers. To address this issue, the Semantic Web has emerged. It envisions a new information infrastructure that enables computers to better address the information needs of human users.

To realize the Semantic Web vision, a number of standard technologies have been developed. These include the Uniform Resource Identifiers (URI) for identifying objects in the Web space as well as Resource Description Framework (RDF) and Web Ontology Language (OWL) for encoding knowledge in the form of standard machine-readable ontologies. The goal is to migrate from the syntactic Web of documents to the semantic Web of ontologies. The leading organization for facilitating, developing, and promoting these Web-based standards is the World Wide Web Consortium (W3C) (<http://www.w3.org>). Since 1994, W3C has published more than ninety such standards, called “W3C Recommendations”, which are specifications or sets of guidelines that, after extensive consensus-building (e.g., through working drafts), have received the endorsement of W3C. As these standard SW technologies are becoming mature and robust, it is important to provide test-beds for these technologies. Many believe that the

life science domain can serve as a rich test-bed for Semantic Web technologies. This belief is substantiated by the following developments.

Publicity. The “Semantic-Web-for-life-science” theme has been brought up and emphasized through keynotes, workshops and special sessions at major international Semantic Web conferences (e.g., ISWC, WWW, and Semantic Technology conferences) and bioinformatics conferences (e.g., Bio-IT World and PSB 2005). The Semantic Web wave also reaches Asia, the first Asian Semantic Web Conference (ASWC) will be held in Beijing, China in September of 2006.

Community Support. The W3C Semantic Web for Health Care and Life Science Interest Group (SW HCLSIG; <http://www.w3.org/2001/sw/hcls>) was inaugurated in September of 2005, and is chartered to develop and support the use of Semantic Web technologies to improve collaboration, research and development, and innovation adoption in the Health Care and Life Science domains. In addition, the e-Science initiative in UK and other major Semantic Web communities including REVERSE (<http://reverso.net/>) and AKT (<http://www.aktors.org/akt/>) have launched projects involving life science applications of the Semantic Web. These communities include both academic and industrial participants across different nations.

Publications. There are a growing number of papers describing Semantic Web use cases for the life sciences, which were published in prestigious journals (e.g., Science and Nature) and conference proceedings (e.g., ISMB and ISWC). A special issue on “Semantic Web for the Life Sciences” was published in the Journal of Web Semantics this year. (http://www.elsevier.com/wps/find/journaldescription.cws_home/671322/description).

Tools. A significant number of Semantic-Web-aware tools have been developed over the past several years. While some of them are proprietary tools developed by commercial vendors, others were developed by academic institutions as open source software. These tools (more tools will be needed) are critical in bringing Semantic Web to bear on behalf on the life scientist.

This book was conceived at the juncture of these exciting developments, in step with the growing awareness and interest of the Semantic Web in the Life Sciences. It encompasses a collection of representative topics written by leading experts who have contributed their technical expertise, experience, and knowledge. This selection of topics and experts is by no means exhaustive and represents the tip of the iceberg. Continued exploration and investigation are required before the potential of the Semantic Web can be fully realized in the life sciences. This book documents encouraging and important first steps.

Acknowledgments

The editors would like to thank the publisher for providing the opportunity to work on this book project and in addition wish to express gratitude to Joe, Marcia and Deborah for their timely help and advice.

The editors also express their gratitude to the numerous reviewers who contributed their time and invaluable expert knowledge to this project, objectively ensuring the quality of contributions to the book.

Contents

Dedication	v
Contributing Authors	xi
Preface	xix
Acknowledgments	xxi
Introduction	1
PART I: Database and Literature Integration	9
Chapter 1	11
Semantic Web Approach to Database Integration in the Life Sciences	
KEI-HOI CHEUNG, ANDREW K. SMITH, KEVIN Y. L. YIP,	
CHRISTOPHER J. O. BAKER AND MARK B. GERSTEIN	
Chapter 2	31
Querying Semantic Web Contents: A Case Study	
LOÏC ROYER, BENEDIKT LINSE, THOMAS WÄCHTER, TIM FURCH,	
FRANÇOIS BRY AND MICHAEL SCHROEDER	
Chapter 3	53
Knowledge Acquisition from the Biomedical Literature	
LYNETTE HIRSCHMAN, WILLIAM HAYES AND ALFONSO VALENCIA	

PART II: Ontologies in the Life Sciences	83
Chapter 4	85
Biological Ontologies	
PATRICK LAMBRIX, HE TAN, VAIDA JAKONIENE AND LENA STRÖMBÄCK	
Chapter 5	101
Clinical Ontologies for Discovery Applications	
YVES LUSSIER AND OLIVIER BODENREIDER	
Chapter 6	121
Ontology Engineering for Biological Applications	
LARISA N. SOLDATOVA AND ROSS D. KING	
Chapter 7	139
The Evaluation of Ontologies:	
Toward Improved Semantic Interoperability	
LEO OBRST, WERNER CEUSTERS, INDERJEET MANI, STEVE RAY	
AND BARRY SMITH	
Chapter 8	159
OWL for the Novice: A Logical Perspective	
JEFF PAN	
PART III: Ontology Visualization	183
Chapter 9	185
Techniques for Ontology Visualization	
XIAOSHU WANG AND JONAS ALMEIDA	
Chapter 10	205
On Visualization of OWL Ontologies	
SERGUEI KRIVOV, FERDINANDO VILLA, RICHARD WILLIAMS	
AND XINDONG WU	
PART IV: Ontologies in Action	223
Chapter 11	225
Applying OWL Reasoning to Genomic Data	
KATY WOLSTENCROFT, ROBERT STEVENS AND VOLKER HAARSLEV	