Ljupčo Todorovski
Nada Lavrač
Klaus P. Jantke (Eds.)

LNAI 4265

Discovery Science

9th International Conference, DS 2006
Barcelona, Spain, October 2006
Proceedings

Springer

Ljupčo Todorovski Nada Lavrač
Klaus P. Jantke (Eds.)

Discovery Science

9th International Conference, DS 2006
Barcelona, Spain, October 7-10, 2006
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Ljupčo Todorovski
Nada Lavrač
Jožef Stefan Institute
Department of Knowledge Technologies
Jamova 39, 1000 Ljubljana, Slovenia
E-mail:{ljupco.todorovski, nada.lavrac}@ijs.si

Klaus P. Jantke
Technical University Ilmenau
Institut für Medien- und Kommunikationswissenschaft
PF 10 05 65, 98684 Ilmenau, Germany
E-mail: klaus-peter.jantke@tu-ilmenau.de

Library of Congress Control Number: 2006933944

CR Subject Classification (1998): I.2, H.2.8, H.3, J.1, J.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-46491-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-46491-4 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11893318 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 4265

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 4265: L. Todorovski, N. Lavrač, K.P. Jantke (Eds.), Discovery Science. XIV, 384 pages. 2006.

Vol. 4264: J.L. Balcázar, P.M. Long, F. Stephan (Eds.), Algorithmic Learning Theory. XIII, 393 pages. 2006.

Vol. 4253: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. XXXII, 1301 pages. 2006.

Vol. 4252: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. XXXIII, 1335 pages. 2006.

Vol. 4251: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LXVI, 1297 pages. 2006.

Vol. 4248: S. Staab, V. Svátek (Eds.), Engineering Knowledge in the Age of the Semantic Web. XIV, 400 pages. 2006.

Vol. 4223: L. Wang, L. Jiao, G. Shi, X. Li, J. Liu (Eds.), Fuzzy Systems and Knowledge Discovery. XXVIII, 1335 pages. 2006.

Vol. 4213: J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.), Knowledge Discovery in Databases: PKDD 2006. XXII, 660 pages. 2006.

Vol. 4212: J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.), Machine Learning: ECML 2006. XXIII, 851 pages. 2006.

Vol. 4211: P. Vogt, Y. Sugita, E. Tuci, C. Nehaniv (Eds.), Symbol Grounding and Beyond. VIII, 237 pages. 2006.

Vol. 4203: F. Esposito, Z.W. Raś, D. Malerba, G. Semeraro (Eds.), Foundations of Intelligent Systems. XVIII, 767 pages. 2006.

Vol. 4201: Y. Sakakibara, S. Kobayashi, K. Sato, T. Nishino, E. Tomita (Eds.), Grammatical Inference: Algorithms and Applications. XII, 359 pages. 2006.

Vol. 4198: O. Nasraoui, O. Zaiane, M. Spiliopoulou, B. Mobasher, B. Masand, P. Yu (Eds.), Advances in Web Minding and Web Usage Analysis. IX, 177 pages. 2006.

Vol. 4196: K. Fischer, I.J. Timm, E. André, N. Zhong (Eds.), Multiagent System Technologies. X, 185 pages. 2006.

Vol. 4188: P. Sojka, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue. XIV, 721 pages. 2006.

Vol. 4183: J. Euzenat, J. Domingue (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. XIII, 291 pages. 2006.

Vol. 4180: M. Kohlhase, OMDoc – An Open Markup Format for Mathematical Documents [version 1.2]. XIX, 428 pages. 2006.

Vol. 4177: R. Marín, E. Onaindía, A. Bugarín, J. Santos (Eds.), Current Topics in Artificial Intelligence. XV, 482 pages. 2006.

Vol. 4160: M. Fisher, W.v.d. Hoek, B. Konev, A. Lisitsa (Eds.), Logics in Artificial Intelligence. XII, 516 pages. 2006.

Vol. 4155: O. Stock, M. Schaerf (Eds.), Reasoning, Action and Interaction in AI Theories and Systems. XVIII, 343 pages. 2006.

Vol. 4149: M. Klusch, M. Rovatsos, T.R. Payne (Eds.), Cooperative Information Agents X. XII, 477 pages. 2006.

Vol. 4140: J.S. Sichman, H. Coelho, S.O. Rezende (Eds.), Advances in Artificial Intelligence - IBERAMIA-SBIA 2006. XXIII, 635 pages. 2006.

Vol. 4139: T. Salakoski, F. Ginter, S. Pyysalo, T. Pahikkala, Advances in Natural Language Processing. XVI, 771 pages. 2006.

Vol. 4133: J. Gratch, M. Young, R. Aylett, D. Ballin, P. Olivier (Eds.), Intelligent Virtual Agents. XIV, 472 pages. 2006.

Vol. 4130: U. Furbach, N. Shankar (Eds.), Automated Reasoning. XV, 680 pages. 2006.

Vol. 4120: J. Calmet, T. Ida, D. Wang (Eds.), Artificial Intelligence and Symbolic Computation. XIII, 269 pages. 2006.

Vol. 4114: D.-S. Huang, K. Li, G.W. Irwin (Eds.), Computational Intelligence, Part II. XXVII, 1337 pages. 2006.

Vol. 4108: J.M. Borwein, W.M. Farmer (Eds.), Mathematical Knowledge Management. VIII, 295 pages. 2006.

Vol. 4106: T.R. Roth-Berghofer, M.H. Göker, H. A. Güvenir (Eds.), Advances in Case-Based Reasoning. XIV, 566 pages. 2006.

Vol. 4099: Q. Yang, G. Webb (Eds.), PRICAI 2006: Trends in Artificial Intelligence. XXVIII, 1263 pages. 2006.

Vol. 4095: S. Nolfi, G. Baldassarre, R. Calabretta, J.C. T. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, D. Parisi (Eds.), From Animals to Animats 9. XV, 869 pages. 2006.

Vol. 4093: X. Li, O.R. Zaïane, Z. Li (Eds.), Advanced Data Mining and Applications. XXI, 1110 pages. 2006.

Vol. 4092: J. Lang, F. Lin, J. Wang (Eds.), Knowledge Science, Engineering and Management. XV, 664 pages. 2006.

Vol. 4088: Z.-Z. Shi, R. Sadananda (Eds.), Agent Computing and Multi-Agent Systems. XVII, 827 pages. 2006.

Vol. 4087: F. Schwenker, S. Marinai (Eds.), Artificial Neural Networks in Pattern Recognition. IX, 299 pages. 2006.

Vol. 4068: H. Schärfe, P. Hitzler, P. Øhrstrøm (Eds.), Conceptual Structures: Inspiration and Application. XI, 455 pages. 2006.

Vol. 4065: P. Perner (Ed.), Advances in Data Mining. XI, 592 pages. 2006.

Vol. 4062: G. Wang, J.F. Peters, A. Skowron, Y. Yao (Eds.), Rough Sets and Knowledge Technology. XX, 810 pages. 2006.

Vol. 4049: S. Parsons, N. Maudet, P. Moraitis, I. Rahwan (Eds.), Argumentation in Multi-Agent Systems. XIV, 313 pages. 2006.

Vol. 4048: L. Goble, J.-J.C.. Meyer (Eds.), Deontic Logic and Artificial Normative Systems. X, 273 pages. 2006.

Vol. 4045: D. Barker-Plummer, R. Cox, N. Swoboda (Eds.), Diagrammatic Representation and Inference. XII, 301 pages. 2006.

Vol. 4031: M. Ali, R. Dapoigny (Eds.), Advances in Applied Artificial Intelligence. XXIII, 1353 pages. 2006.

Vol. 4029: L. Rutkowski, R. Tadeusiewicz, L.A. Zadeh, J.M. Zurada (Eds.), Artificial Intelligence and Soft Computing – ICAISC 2006. XXI, 1235 pages. 2006.

Vol. 4027: H.L. Larsen, G. Pasi, D. Ortiz-Arroyo, T. Andreasen, H. Christiansen (Eds.), Flexible Query Answering Systems. XVIII, 714 pages. 2006.

Vol. 4021: E. André, L. Dybkjær, W. Minker, H. Neumann, M. Weber (Eds.), Perception and Interactive Technologies. XI, 217 pages. 2006.

Vol. 4020: A. Bredenfeld, A. Jacoff, I. Noda, Y. Takahashi (Eds.), RoboCup 2005: Robot Soccer World Cup IX. XVII, 727 pages. 2006.

Vol. 4013: L. Lamontagne, M. Marchand (Eds.), Advances in Artificial Intelligence. XIII, 564 pages. 2006.

Vol. 4012: T. Washio, A. Sakurai, K. Nakajima, H. Takeda, S. Tojo, M. Yokoo (Eds.), New Frontiers in Artificial Intelligence. XIII, 484 pages. 2006.

Vol. 4008: J.C. Augusto, C.D. Nugent (Eds.), Designing Smart Homes. XI, 183 pages. 2006.

Vol. 4005: G. Lugosi, H.U. Simon (Eds.), Learning Theory. XI, 656 pages. 2006.

Vol. 3978: B. Hnich, M. Carlsson, F. Fages, F. Rossi (Eds.), Recent Advances in Constraints. VIII, 179 pages. 2006.

Vol. 3963: O. Dikenelli, M.-P. Gleizes, A. Ricci (Eds.), Engineering Societies in the Agents World VI. XII, 303 pages. 2006.

Vol. 3960: R. Vieira, P. Quaresma, M.d.G.V. Nunes, N.J. Mamede, C. Oliveira, M.C. Dias (Eds.), Computational Processing of the Portuguese Language. XII, 274 pages. 2006.

Vol. 3955: G. Antoniou, G. Potamias, C. Spyropoulos, D. Plexousakis (Eds.), Advances in Artificial Intelligence. XVII, 611 pages. 2006.

Vol. 3949: F. A. Savacı (Ed.), Artificial Intelligence and Neural Networks. IX, 227 pages. 2006.

Vol. 3946: T.R. Roth-Berghofer, S. Schulz, D.B. Leake (Eds.), Modeling and Retrieval of Context. XI, 149 pages. 2006.

Vol. 3944: J. Quiñonero-Candela, I. Dagan, B. Magnini, F. d'Alché-Buc (Eds.), Machine Learning Challenges. XIII, 462 pages. 2006.

Vol. 3937: H. La Poutré, N.M. Sadeh, S. Janson (Eds.), Agent-Mediated Electronic Commerce. X, 227 pages. 2006.

Vol. 3930: D.S. Yeung, Z.-Q. Liu, X.-Z. Wang, H. Yan (Eds.), Advances in Machine Learning and Cybernetics. XXI, 1110 pages. 2006.

Vol. 3918: W.K. Ng, M. Kitsuregawa, J. Li, K. Chang (Eds.), Advances in Knowledge Discovery and Data Mining. XXIV, 879 pages. 2006.

Vol. 3913: O. Boissier, J. Padget, V. Dignum, G. Lindemann, E. Matson, S. Ossowski, J.S. Sichman, J. Vázquez-Salceda (Eds.), Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems. XII, 259 pages. 2006.

Vol. 3910: S.A. Brueckner, G.D.M. Serugendo, D. Hales, F. Zambonelli (Eds.), Engineering Self-Organising Systems. XII, 245 pages. 2006.

Vol. 3904: M. Baldoni, U. Endriss, A. Omicini, P. Torroni (Eds.), Declarative Agent Languages and Technologies III. XII, 245 pages. 2006.

Vol. 3900: F. Toni, P. Torroni (Eds.), Computational Logic in Multi-Agent Systems. XVII, 427 pages. 2006.

Vol. 3899: S. Frintrop, VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search. XIV, 216 pages. 2006.

Vol. 3898: K. Tuyls, P.J. 't Hoen, K. Verbeeck, S. Sen (Eds.), Learning and Adaption in Multi-Agent Systems. X, 217 pages. 2006.

Vol. 3891: J.S. Sichman, L. Antunes (Eds.), Multi-Agent-Based Simulation VI. X, 191 pages. 2006.

Vol. 3890: S.G. Thompson, R. Ghanea-Hercock (Eds.), Defence Applications of Multi-Agent Systems. XII, 141 pages. 2006.

Vol. 3885: V. Torra, Y. Narukawa, A. Valls, J. Domingo-Ferrer (Eds.), Modeling Decisions for Artificial Intelligence. XII, 374 pages. 2006.

Vol. 3881: S. Gibet, N. Courty, J.-F. Kamp (Eds.), Gesture in Human-Computer Interaction and Simulation. XIII, 344 pages. 2006.

Vol. 3874: R. Missaoui, J. Schmidt (Eds.), Formal Concept Analysis. X, 309 pages. 2006.

Vol. 3873: L. Maicher, J. Park (Eds.), Charting the Topic Maps Research and Applications Landscape. VIII, 281 pages. 2006.

Vol. 3864: Y. Cai, J. Abascal (Eds.), Ambient Intelligence in Everyday Life. XII, 323 pages. 2006.

Vol. 3863: M. Kohlhase (Ed.), Mathematical Knowledge Management. XI, 405 pages. 2006.

Vol. 3862: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 267 pages. 2006.

¥444.00元

Preface

The 9th International Conference on Discovery Science (DS 2006) was held in Barcelona, Spain, on 7–10 October 2006. The conference was collocated with the 17th International Conference on Algorithmic Learning Theory (ALT 2006). The two conferences shared the invited talks.

This LNAI volume, containing the proceedings of the 9th International Conference on Discovery Science, is structured in three parts. The first part contains the papers/abstracts of the invited talks, the second part contains the accepted long papers, and the third part the accepted regular (short) papers. Out of 87 submitted papers, 23 were accepted for publication as long papers, and 18 as regular papers. All the submitted papers were reviewed by two or three referees. In addition to the presentations of accepted papers, the DS 2006 conference program consisted of three invited talks, two tutorials, the collocated ALT 2006 conference and the Pascal Dialogues workshop.

We wish to express our gratitude to

- the authors of submitted papers,
- the program committee and other referees for their thorough and timely paper evaluation,
- DS 2006 invited speakers Carole Goble and Padhraic Smyth, as well as Andrew Ng as joint DS 2006 and ALT 2006 invited speaker,
- invited tutorial speakers Luis Torgo and Michael May,
- the local organization committee chaired by Ricard Gavaldà,
- DS 2006 conference chair Klaus P. Jantke,
- the DS steering committee, chaired by Hiroshi Motoda,
- Andrei Voronkov for the development of EasyChair which provided excellent support in the paper submission, evaluation and proceedings production process,
- Alfred Hofmann of Springer for the co-operation in publishing the proceedings,
- the ALT 2006 PC chairs Phil Long and Frank Stephan, as well as Thomas Zeugman and José L. Balcázar, for the cooperation and coordination of the two conferences, and finally
- we gratefully acknowledge the financial support of the Universitat Politècnica de Catalunya, Idescat — the Statistical Institute of Catalonia (for providing support to tutorial speakers), the PASCAL Network of Excellence (for supporting the Pascal Dialogues), the Spanish Ministry of Science and Education, the Slovenian Ministry of Higher Education, Science, and Technology, and Yahoo! Research for sponsoring the Carl Smith Student Award.

We hope that the week in Barcelona in early October 2006 was a fruitful, challenging and enjoyable scientific and social event.

August 2006 Nada Lavrač and Ljupčo Todorovski

Conference Organization

Conference Chair

Klaus Jantke	Technical University of Ilmenau, Germany

Steering Committee Chair

Hiroshi Motoda	AFOSR/AOARD & Osaka University, Japan

Program Committee Chairs

Nada Lavrač	Jožef Stefan Institute, Slovenia
Ljupčo Todorovski	Jožef Stefan Institute, Slovenia

Program Committee

José Luis Balcázar	Universitat Politècnica de Catalunya, Spain
Michael Berthold	University of Konstanz, Germany
Elisa Bertino	Purdue University, USA
Vincent Corruble	Pierre and Marie Curie University, France
Andreas Dress	Shanghai Institutes for Biological Sciences, China
Sašo Džeroski	Jožef Stefan Institute, Slovenia
Tapio Elomaa	Tampere University of Technology, Finland
João Gama	University of Porto, Portugal
Dragan Gamberger	Rudjer Bošković Institute, Croatia
Gunter Grieser	Technical University of Darmstadt, Germany
Fabrice Guillet	University of Nantes, France
Mohand-Saïd Hacid	Claude Bernard University Lyon 1, France
Udo Hahn	Jena University, Germany
Tu Bao Ho	Japan Advanced Institute of Science and Technology, Japan
Achim Hoffmann	University of New South Wales, Australia
Szymon Jaroszewicz	Szczecin University of Technology, Poland
Kristian Kersting	University of Freiburg, Germany
Ross King	University of Wales, UK
Kevin Korb	Monash University, Australia
Ramamohanarao Kotagiri	University of Melbourne, Australia
Stefan Kramer	Technical University of Munich, Germany
Nicolas Lachiche	Louis Pasteur University, France
Aleksandar Lazarević	University of Minnesota, USA
Jinyan Li	Institute for Infocomm Research, Singapore

Ashesh Mahidadia	University of New South Wales, Australia
Michael May	Fraunhofer Institute for Autonomous Intelligent Systems, Germany
Dunja Mladenić	Jožef Stefan Institute, Slovenia
Igor Mozetič	Jožef Stefan Institute, Slovenia
Ion Muslea	Language Weaver, USA
Lourdes Peña Castillo	University of Toronto, Canada
Bernhard Pfahringer	University of Waikato, New Zealand
Jan Rauch	University of Economics, Prague, Czech Republic
Domenico Saccà	University of Calabria, Italy
Rudy Setiono	National University of Singapore, Singapore
Einoshin Suzuki	Kyushu University, Japan
Masayuki Takeda	Kyushu Univeristy, Japan
Kai Ming Ting	Monash University, Australia
Alfonso Valencia	National Centre for Biotechnology, Spain
Takashi Washio	Osaka University, Japan
Gerhard Widmer	Johannes Kepler University, Austria
Akihiro Yamamoto	Kyoto University, Japan
Mohammed Zaki	Rensselaer Polytechnic Institute, USA
Filip Železný	Czech Technical University, Czech Republic
Djamel A. Zighed	Lumière University Lyon 2, France

Local Organization

Ricard Gavaldà (Chair)	Universitat Politècnica de Catalunya, Spain
José Luis Balcázar	Universitat Politècnica de Catalunya, Spain
Albert Bifet	Universitat Politècnica de Catalunya, Spain
Gemma Casas	Universitat Politècnica de Catalunya, Spain
Jorge Castro	Universitat Politècnica de Catalunya, Spain
Jesús Cerquides	Universitat de Barcelona, Spain
Pedro Delicado	Universitat Politècnica de Catalunya, Spain
Gábor Lugosi	Universitat Pompeu Fabra, Spain
Victor Dalmau	Universitat Pompeu Fabra, Spain

External Reviewers

Zeyar Aung
Jose Campos Avila
Hideo Bannai
Maurice Bernadet
Julien Blanchard
Bruno Bouzy
Janez Brank
Agnès Braud
Jérôme David
Stéphane Daviet
Fabian Dill
Tomaž Erjavec
Jure Ferlež
Blaž Fortuna
Gemma C. Garriga
Gianluigi Greco
El Ghazel Haytham
Hakim Hacid
Kohei Hatano
Le Minh Hoang
Zujun Hou
Qingmao Hu
Salvatore Iritano
Daisuke Ikeda
Ignazio Infantino
Aneta Ivanovska
Jussi Kujala
Ha Thanh Le
Aydano Machado
Giuseppe Manco
Blaž Novak
Riccardo Ortale
François Pachet
Kwok Pan Pang
Panče Panov
Nguyen Thanh Phuong
Gerard Ramstein
Lothar Richter
Pedro Rodrigues
Massimo Ruffolo
Ulrich Rückert
Ivica Slavkov
James Tan Swee-Chuan
Miha Vuk
Bernd Wiswedel
Bernard Ženko

Table of Contents

I Invited Papers

II Long Papers

III Regular Papers

e-Science and the Semantic Web: A Symbiotic Relationship

Carole Goble[1], Oscar Corcho[1], Pinar Alper[1], and David De Roure[2]

[1] School of Computer Science, University of Manchester,
Manchester M13 9PL, UK
{carole, ocorcho, penpecip}@cs.man.ac.uk
[2] School of Electronics and Computer Science, University of Southampton,
Southampton SO17 1BJ UK
dder@ecs.soton.ac.uk

Abstract. e-Science is scientific investigation performed through distributed global collaborations between scientists and their resources, and the computing infrastructure that enables this. Scientific progress increasingly depends on pooling know-how and results; making connections between ideas, people, and data; and finding and reusing knowledge and resources generated by others in perhaps unintended ways. It is about harvesting and harnessing the "collective intelligence" of the scientific community. The Semantic Web is an extension of the current Web in which information is given well-defined meaning to facilitate sharing and reuse, better enabling computers and people to work in cooperation. Applying the Semantic Web paradigm to e-Science has the potential to bring significant benefits to scientific discovery. We identify the benefits of lightweight and heavyweight approaches, based on our experiences in the Life Sciences.

1 Introduction

The term e-Science is normally used to describe computationally intensive science that is carried out collaboratively in highly distributed network environments [1]. Typically, a feature of such collaborative scientific enterprises is that they require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists. Brain neuroscientists remotely control and collect data from the world's largest and most powerful transmission electron microscope in Japan (telescience.ucsd.edu). Astronomers steer telescopes from their offices, collect the data using remote archive repositories, and process it by exploiting the availability of machines of other institutions. The International Virtual Observatory Alliance (www.ivoa.net) makes available vast digital sky archives to all astronomers not just a lucky few. ^{my}Grid-Taverna (www.mygrid.org.uk) allows biologists to assemble personalised exploratory *in silico* experiments that interoperate between remotely and locally available applications, code-bases and databases to identify new genes [2]. The Human Genome effort is an

N. Lavrač, L. Todorovski, and K.P. Jantke (Eds.): DS 2006, LNAI 4265, pp. 1 – 12, 2006.

example of e-Science – over 500 datasets and tools are available on the web for bioinformaticians to piece together our understanding of life [3, 4].

The e-Science infrastructure supports and enhances the scientific process by enabling scientists to generate, analyse, share and discuss their insights, experiments and results in a more effective manner, particularly in the context of the deluge of data resulting from new experimental practices [5, 6].

Scientific progress increasingly depends on pooling resources, know-how and results; making connections between ideas, people, and data; and finding and interpreting knowledge generated by others, in ways that may not have been anticipated when it was created. It is about harvesting and harnessing the "collective intelligence" of the scientific community. It has as much to do with intelligent information management as with sharing scarce resources like large scale computing power or expensive instrumentation. It is about making connections between decoupled resources and people in the broadest context of diverse scientific activity, outside the bounds of localised experiments and closed projects, and enabling scientific endeavour "in the wild".

The Semantic Web is defined as an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation [7]. Applying Semantic Web to Science [8] has attracted great interest particularly in the Life Sciences [9-13] which has been proposed as a "nursery" for incubating the required technological developments [14]. We take a perspective in which the Semantic Web is seen foremost as an infrastructure for gathering and exploiting collective intelligence; i.e. the capacity of human communities to evolve towards higher order complexity and integration through collaboration and innovation.

Section 2 introduces the Semantic Web and distinguishes between lightweight and heavyweight approaches. In Sections 3 to 5 we present three aspects of the Semantic Web – annotation, integration and inference – and sketch how the methods, techniques and tools used for each of them could provide benefits to scientists. In Section 6 we reflect on the symbiosis between e-Science and the Semantic Web.

2 The Semantic Web

Annotation is the process of associating metadata with an object. Metadata, defined as structured data about an object that supports functions associated with it, can be generated for any entity for which contextual data can be recorded [15]. We can annotate any object, be it a document, dataset, publication, codes, notebooks, and so on, within the scientific process – even a person or scientific instrument. Metadata can be expressed in a wide range of languages (from natural to formal ones) and with a wide range of vocabularies (from simple ones, based on a set of agreed keywords, to complex ones, with agreed taxonomies and formal axioms). It can be available in different formats: electronically or even physically (written down in the margins of a textbook). It can be created and maintained using different types of tools (from text editors to metadata generation tools), either manually or automatically.