Preface

The advancement of modern science is growing more and more dependent on the Internet technology. Scientists rely upon it to retrieve valuable information and knowledge, communicate research ideas, have virtual meetings, exchange resources, and coordinate collaboration. As a matter of fact, the Internet has become an indispensable media for various scientific research endeavors in many kinds of disciplines. Specifically, the rapid increase in complexity and scope of modern science research brings up a protruding challenge, which is to support world-wide scale collaboration and coordinated resource sharing across disciplinary, organizational, social, and cultural boundaries. New discovery normally resides in those boundaries. Researchers working on one aspect typically are required to look into those boundaries, utilize resources from other research institutes, and explore knowledge from other fields from perspectives of a completely different scientific disciplinary. A typical case study is integrative neuroscience research. For example, the study of neurodegenerative diseases such as Parkinson, Alzheimer, Huntington needs to combine knowledge and resources from a good many research institutions. The activities normally span a wide range of the disciplines such as psychiatry, neurology, microscopic anatomy, neuronal physiology, biochemistry, genetics, molecular biology, and bioinformatics. These types of requirements require a new generation of information infrastructure to enable scientific investigation performed through distributed global collaborations between scientists and their resources. The term e-Science was thus brought up to refer to such kind of computing infrastructure that enables cross-institutions and interdisciplinary research.

The Semantic Web, coined by web inventor Tim Berners Lee, is normally deemed as the next generation of web. It aims at providing much more advanced integration capabilities that allow data, particularly the meaning of the data, to be shared and reused across application, enterprise, and community boundaries. The approach is to encode data semantics in a more formal and explicit way to make data machine-understandable and automatically consumable. A number of

---

This book started before Yimin Wang joined Lilly Singapore Centre for Drug Discovery; therefore, this book does not contain any content or opinion related to Lilly Singapore Centre for Drug Discovery and its parent company Eli Lilly & Co.

new web languages such as RDF (the Resource Description Framework) and OWL (the Web Ontology Language) have been proposed and standardized by the World Wide Web Consortium (W3C). These languages provide new capability for resource description and knowledge representation going far beyond the content presentation capabilities of HTML language and data tagging capabilities of the XML language. These languages can be used to define meaning of data, describe meta-data, represent terminologies and vocabularies, describe the input/output and functional capability of web service, and so forth.

As a synthesis of knowledge and web technologies and after over 10 years of development, the Semantic Web has been gaining tremendous attention worldwide. One noteworthy movement is the increasing recognition of this new technology in typical scientific areas such as life science,[1] earth science, and social science. For example, the World Wide Web Consortium (W3C) has established a Semantic Web interest group to focus on health care and life science (HCLSIG). This group is designed to improve collaboration, research and development, and innovation adoption of Semantic Web technologies in the health care and life science domains, aiding decision-making in clinical research and helping bridge many forms of biological and medical information across institutions. Other examples include the National Virtual Observatory, which is exploring the use of Semantic Web technologies for linking numerous astronomical resources together, and the FUSION project which is to apply, extend, and combine Semantic Web technologies and image analysis techniques to develop a knowledge management system to optimize the design of fuel cells.

As the semantic technology has been gaining momentum in various e-Science areas, it is important to offer semantic-based methodologies, tools, and middleware to facilitate scientific knowledge modeling, logical-based hypothesis checking, semantic data integration and application composition, integrated knowledge discovery, and data analyzing for different e-Science applications. However, such a research area does not yet exist in a coherent form. Influenced by the Artificial Intelligence community, the Semantic Web researchers have largely focused on formal aspects of logic languages or general-purpose semantic application development, with inadequate consideration of requirements from specific scientific areas. On the other hand, general science researchers are growing ever more dependent on the web, but they have no coherent agenda for exploring the emerging trends on Semantic Web technologies. It urgently requires the development of a multidisciplinary field to foster the growth and development of e-Science applications based on the semantic technologies and related knowledge-based approaches. We regard it as necessity to set a research agenda at this point in time, in order to steer the development and the research efforts in the most rewarding direction toward a common goal of realizing the semantic e-Science.

---

[1] Three recent relevant special issues published in life science area:
(1) *BMC Bioinformatics*: Semantic e-Science in biomedicine.
(2) *Journal of Biomedical Informatics*: Special issue on Semantic BioMed Mashup.
(3) *Journal of Web Semantics*: Special issue on Semantic Web in life science.

This book surveys the state of the art in the field of applying semantic technologies in typical e-Science applications. We anticipate it to provide a sufficient overview and valuable summary on semantic e-Science applications and to report the cutting-edge research outcome in this field. The book is conceived from a series of events.[2] We received nearly 30 submissions as a result of our public call for chapters, from which we selected 8 as the candidate chapters.

In Chapter 1, Prof. David De Roure and Carole Goble introduce the field being known as the 'Semantic Grid' that emphasizes on the Semantic Web and on the joined-up infrastructure to support the increasing scale of data, computation, collaboration, and automation as science and computing advance. Since its instigation in 2001 the Semantic Grid community has established a significant body of work on the role of the ideas and technologies of the Semantic Web in providing and utilizing the infrastructure for science.

In Chapter 2, Dr. M. Scott Marshall and colleagues introduce the Virtual Laboratory for e-Science (VL-e) project, which is a project with academic and industrial partners where e-Science has been applied to several domains of scientific research. The authors explain what 'semantic disclosure' means and how it is essential to knowledge sharing in e-Science. The authors describe several Semantic Web applications and how they were built using components of the proposed application toolkit.

In Chapter 3, Dr. Hock Beng Lim and colleagues describe a smart e-Science cyber-infrastructure for cross-disciplinary scientific collaborations. The proposed infrastructure forms the key resource sharing backbone that enables each participating scientific community to expose their sensor, computational, data, and intellectual resources in a service-oriented manner, accompanied by the domain-specific knowledge and semantic descriptions.

In Chapter 4, Dr. Alexander Garcia and colleagues propose a new methodology for developing ontologies within a decentralized setting in the e-Science context. The Melting Point (MP) approach is the product of direct first-hand experience and observation of biological communities of practice in which some of the authors have been involved.

In Chapter 5, Dr. Amitava Biswas and colleagues investigate semantic technologies for searching in e-Science grids. They present a survey of meaning representation and comparison technologies, followed by a design of meaning representation and comparison technique which is coherent to the cognitive science and linguistics models. This meaning comparison technique discerns complex meaning while enabling search query relaxation and search key interchangeability.

In Chapter 6, Dr. Enrico Pontelli and colleagues present a semantic e-Science system called *BioService Integration System (BSIS)*. BSIS is aimed at automating bioinformatics tasks through intelligent workflow construction. It provides a

[2] Four recent events:
(1) WWW2008 Semantic Web for health care and life sciences workshop.
(2) WWW2007 Workshop on health care and life science data integration for the Semantic Web.
(3) AAAI2007 Workshop on Semantic e-Science.
(4) ASWC2006 Workshop on Semantic e-Science.

graphic-based workflow language that can enable the description of workflows composed of abstract Semantic Web services and references to biologically relevant data. The workflows constructed in BSIS can be instantiated through automated planning techniques and automatically executed by adapting the *Web Service Integration Framework (WSIF)*.

In Chapter 7, Mikhail Simonov and Flavia Mazzitelli introduce the applications of rules and semantics in near-miss detection in nursing. Nursing science deals with human-to-human interaction delivering care service, showing several peculiarities compared with other life science domains. Semantic technologies can be applied to clinical nursing, where the risk management forms a particular application class and requires error detection and semantic technologies complementing best nursing strategies. The chapter investigates on possible risk control measures in above-said nursing, suggesting a combination of Information and Communication Technologies (ICT) and knowledge technologies.

In Chapter 8, Dr. Peisheng Zhao and colleagues describe a semantic e-Science application in the earth science domain. The Sensor Web provides an interoperable way to discover, task, and retrieve sensor data over the network. This chapter leverages recent advances in autonomous control and the Semantic Web technologies to present a systematic approach in which the Sensor Web and Earth science models are annotated with semantic meta-data and chained together in an autonomous way as a service chain to simulate the process of Sensor Web mining based on semantic inference. This approach significantly advances sensor data exploration with semantics in an interoperable way to improve the accuracy and timeliness of monitoring and predicting rapidly changing Earth phenomena.

In Chapter 9, Vít Nováček and colleagues investigate the application of semantic technologies in scientific publication. It describes the CORAAL system for knowledge-based life science publication search. It presents the approach from three different perspectives: (1) extraction of asserted publication meta-data together with the knowledge implicitly present in the respective text; (2) integration, refinement, and extension of the emergent content; (3) release of the processed content via a meaning-sensitive search and browse interface catering for services complementary to the current full-text search.

In summary, the merits of the Semantic Web with respect to scientific research are exhibited in different perspectives encompassing collective knowledge acquisition, semantic data integration, scientific workflow management and composition, integrative knowledge discovery and data mining, logic-based hypothesis checking, and so forth. The semantic e-Science is still an evolving area with rapid development. Although this book cannot give a complete account on all issues and topics, we hope it can shed some light on the most important aspects. We also hope it can push and promote the semantic technologies within various science fields, particularly in interdisciplinary scientific research.

Hangzhou, China — Huajun Chen
Immunos, Singapore — Yimin Wang
New Haven, Connecticut — Kei-Hoi Cheung

Springer
http://www.springer.com/978-1-4419-5902-7

Semantic e-Science
Chen, J.; Wang, Y.; Cheung, K.-H. (Eds.)
2010, VII, 352 p. 127 illus., Softcover
ISBN: 978-1-4419-5902-7