Preface

This book is a compilation of the International Journal on Semantic Web and Information Systems (IJSWIS) Volume 7. It is composed of high quality manuscripts pertaining to various aspects of Semantic Web that are relevant to computer science and information systems communities. In addition to being a premier outlet of high quality archival research in Semantic Web, IJSWIS also had high impact factor of 2.3 in the year of the publications covered here and belongs among the top journals in World Wide Web with over 7 citations per publication on average (Microsoft Academic Search, http://bit.ly/www-j1).

The chapters in this book are divided in three parts. The first part on Ontology Engineering and Knowledge management has three chapters that deal with ontology engineering, relational learning and knowledge extraction. The second part on Reasoning and Ontology Development has four chapters that focus on reasoning. Three of them use induction form of reasoning for various uses in formal ontology development. The final and the third section of the book on Web of Data and Applications consist of five chapters that cover Web of Data and variety of applications involving semantic processing of data. After presenting the overview of twelve chapters in this book, I briefly provide my take on some of the exciting new progress and exciting things to come in near future in Semantic Web.

SECTION 1: ONTOLOGY ENGINEERING AND KNOWLEDGE MANAGEMENT

The first chapter “Online Semantic Knowledge Management for Product Design Based on Product Engineering Ontologies,” by Lijuan Zhu, Uma Jayaram, and Okjoon Kim, formulates an approach to use the semantic web for knowledge management in the product design domain to provide enhanced capabilities of authoring/updating, querying/reasoning, searching, and visualization of information. Engineering has unique challenges, due to the pervasive use of CAD models and underlying interoperability and integration issues. The authors propose a distributed model composed of a host hybrid-data repository, external public linked data sources, a semantic data management engine, and a web-based user interface layer. The hybrid-data repository consists of ontologies to preserve knowledge for the product design domain and a conventional product data base to utilize legacy design data. Near full integration with a web based environment is achieved. The importance of accessing product related CAD data that has been instantiated in ontology models, query them, and then display the data on a web interface in real time with other legacy data, such as hand sketches and notes that have been scanned and relevant information from public linked data sites, is a useful and transformational capability. The system clearly facilitates design and information management beyond traditional CAD capabilities and creates a foundation for important capability improvements in the domain.
The agentivity of social entities has posed problems for ontologies of social phenomena, especially in the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) designed for use in the semantic web. The second chapter, “A Theory of Social Agentivity and its Integration into the Descriptive Ontology for Linguistic and Cognitive Engineering,” Edward Heath Robinson elucidates a theory by which physical and social objects can take action, but that also recognizes the different ways in which they act. It introduces the “carry” relationship, by which it allows social actions to occur when a physical action is taken in the correct circumstances. For example, the physical action of a wave of a hand may carry the social action of saying hello when entering a room. This article shows how a system can simultaneously and in a noncontradictory manner handle statements and queries in which both nonphysical social agents and physical agents take action by the carry relationship and the use of representatives. A revision of DOLCE’s taxonomic structure of perdurants is also proposed. This revision divides perdurants into physical and nonphysical varieties at the same ontological level at which endurants are so divided.

The third chapter “AL-QuIn: An Onto-Relational Learning System for Semantic Web Mining” is by Francesca A. Lisi. Onto-Relational Learning is an extension of Relational Learning aimed at accounting for ontologies in a clear, well-founded and elegant manner. The system AL-QuIn supports a variant of the frequent pattern discovery task by following the Onto-Relational Learning approach. It takes taxonomic ontologies into account during the discovery process and produces descriptions of a given relational database at multiple granularity levels. The functionalities of the system are illustrated by means of examples taken from a Semantic Web Mining case study concerning the analysis of relational data extracted from the on-line CIA World Fact Book.

SECTION 2: REASONING AND ONTOLOGY DEVELOPMENT

Chapter 4 titled “Semi-Automatic Ontology Construction by Exploiting Functional Dependencies and Association Rules” by Luca Cagliero, Tania Cerquitelli, and Paolo Garza presents a novel semi-automatic approach to construct conceptual ontologies over structured data by exploiting both the schema and content of the input dataset. It effectively combines two well-founded database and data mining techniques, i.e., functional dependency discovery and association rule mining, to support domain experts in the construction of meaningful ontologies, tailored to the analyzed data, by using Description Logic (DL). To this aim, functional dependencies are first discovered to highlight valuable conceptual relationships among attributes of the data schema (i.e., among concepts). The set of discovered correlations effectively support analysts in the assertion of the Tbox ontological statements (i.e., the statements involving shared data conceptualizations and their relationships). Then, the analyst-validated dependencies are exploited to drive the association rule mining process. Association rules represent relevant and hidden correlations among data content and they are used to provide valuable knowledge at the instance level. The pushing of functional dependency constraints into the rule mining process allows analysts to look into and exploit only the most significant data item recurrences in the assertion of the Abox ontological statements (i.e., the statements involving concept instances and their relationships).

The next chapter, “Concept Induction in Description Logics Using Information-Theoretic Heuristics,” by Nicola Fanizzi presents an approach to ontology construction pursued through the induction of concept descriptions expressed in Description Logics. The author surveys the theoretical foundations of the standard representations for formal ontologies in the Semantic Web. After stating the learning problem in this peculiar context, a FOIL-like algorithm is presented that can be applied to learn DL
concept descriptions. The algorithm performs a search through a space of candidate concept definitions by means of refinement operators. This process is guided by heuristics that are based on the available examples. The author discusses related theoretical aspects of learning with the inherent incompleteness underlying the semantics of this representation. The experimental evaluation of the system DL-Foil, which implements the learning algorithm, was carried out in two series of sessions on real ontologies from standard repositories for different domains expressed in diverse description logics.

Chapter 6 titled “Using Similarity-Based Approaches for Continuous Ontology Development” by Maryam Ramezani presents novel algorithms for learning semantic relations from an existing ontology or concept hierarchy. The author suggest recommendation of semantic relations for supporting continuous ontology development, i.e., the development of ontologies during their use in social semantic bookmarking, semantic wiki, or other Web 2.0 style semantic applications. This chapter assists users in placing a newly added concept in a concept hierarchy. The proposed algorithms are evaluated using datasets from Wikipedia category hierarchy and provide recommendations.

The motivation behind the seventh chapter, titled “A Modal Defeasible Reasoner of Deontic Logic for the Semantic Web,” by Efstratios Kontopoulos, Nick Bassiliades, Guido Governatori, and Grigoris Antoniou, is to develop a practical defeasible reasoner for the Semantic Web that takes advantage of the expressive power offered by modal logics, accompanied by the flexibility to define diverse agent behaviours. Defeasible logic is a non-monotonic formalism that deals with incomplete and conflicting information, whereas modal logic deals with the concepts of necessity and possibility. These types of logics play a significant role in the emerging Semantic Web, which enriches the available Web information with meaning, leading to better cooperation between end-users and applications. Defeasible and modal logics, in general, and, particularly, deontic logic provide means for modeling agent communities, where each agent is characterized by its cognitive profile and normative system, as well as policies, which define privacy requirements, access permissions, and individual rights. Toward this direction, this article discusses the extension of DR-DEVICE, a Semantic Web-aware defeasible reasoner, with a mechanism for expressing modal logic operators, while testing the implementation via deontic logic operators, concerned with obligations, permissions, and related concepts. A further incentive is to study the various motivational notions of deontic logic and discuss the cognitive state of agents, as well as the interactions among them.

SECTION 3: WEB OF DATA AND APPLICATIONS

By specifying that published datasets must link to other existing datasets, the 4th linked data principle ensures a Web of data and not just a set of unconnected data islands. In Chapter 8 titled “Data Linking for the Semantic Web,” Alfio Ferrara, Andriy Nikolov, and François Scharffe propose the term data linking to name the problem of finding equivalent resources on the Web of linked data. In order to perform data linking, many techniques were developed, finding their roots in statistics, database, natural language processing and graph theory. The authors begin this paper by providing background information and terminological clarifications related to data linking. Then a comprehensive survey over the various techniques available for data linking is provided. These techniques are classified along the three criteria of granularity, type of evidence, and source of the evidence. Finally, the authors survey eleven recent tools performing data linking and we classify them according to the surveyed techniques.
The next chapter is “ACRONYM: Context Metrics for Linking People to User-Generated Media Content” by Fergal Monaghan, Siegfried Handschuh, and David O’Sullivan. With the advent of online social networks and User-Generated Content (UGC), the social Web is experiencing an explosion of audio-visual data. But the usefulness of the collected data is in doubt, given that the means of retrieval are limited by the semantic gap between them and people’s perceived understanding of the memories they represent. Whereas machines interpret UGC media as series of binary audio-visual data, humans perceive the context under which the content is captured and the people, places, and events represented. The Annotation CReatiON for Your Media (ACRONYM) framework addresses the semantic gap by supporting the creation of a layer of explicit machine-interpretable meaning describing UGC context. This paper presents an overview of a use case of ACRONYM for semantic annotation of personal photographs. The authors define a set of recommendation algorithms employed by ACRONYM to support the annotation of generic UGC multimedia. This paper introduces the context metrics and combination methods that form the recommendation algorithms used by ACRONYM to determine the people represented in multimedia resources. For the photograph annotation use case, these result in an increase in recommendation accuracy. Context-based algorithms provide a cheap and robust means of UGC media annotation that is compatible with and complimentary to content-recognition techniques.

The tenth chapter “An Enhanced Semantic Layer for Hybrid Recommender Systems: Application to News Recommendation” written by Iván Cantador, Pablo Castells, and Alejandro Bellogín. Recommender systems have achieved success in a variety of domains, as a means to help users in information overload scenarios by proactively finding items or services on their behalf, taking into account or predicting their tastes, priorities, or goals. Challenging issues in their research agenda include the sparsity of user preference data and the lack of flexibility to incorporate contextual factors in the recommendation methods. To a significant extent, these issues can be related to a limited description and exploitation of the semantics underlying both user and item representations. The authors propose a three-fold knowledge representation, in which an explicit, semantic-rich domain knowledge space is incorporated between user and item spaces. The enhanced semantics support the development of contextualisation capabilities and enable performance improvements in recommendation methods. As a proof of concept and evaluation testbed, the approach is evaluated through its implementation in a news recommender system, in which it is tested with real users. In such scenario, semantic knowledge bases and item annotations are automatically produced from public sources.

The next chapter is “Numeric Query Answering on the Web,” by Steven O’Hara and Tom Bylander. Query answering usually assumes that the asker is looking for a single correct answer to the question. When retrieving a textual answer this is often the case, but when searching for numeric answers, there are additional considerations. In particular, numbers often have units associated with them, and the asker may not care whether the raw answer is in feet or meters. Also, numbers usually denote a precision. In a few cases, the precision may be explicit, but normally, there is an implied precision associated with every number. Finally, an association between different reliability levels to different sources can be made. In this paper, the authors experimentally show that, in the context of conflicting answers from multiple sources, numeric query accuracy can be improved by taking advantage of units, precision, and the reliability of sources.

The effective acquisition of (semantic) metadata is crucial for many present day applications. Games with a purpose address this issue by transforming computational problems into computer games. In the final and the twelfth chapter 12 “Semantics Discovery via Human Computation Games,” Jakub Šimko, Michal Tvarožek, and Mária Bieliková present a novel approach to metadata acquisition via Little Search
Game (LSG) – a competitive web search game, whose purpose is the creation of a term relationship network. From a player perspective, the goal is to reduce the number of search results returned for a given search term by adding negative search terms to a query. The authors describe specific aspects of the game’s design, including player motivation and anti-cheating issues. The authors have performed a series of experiments with Little Search Game, acquired real-world player input, gathered qualitative feedback from the players, constructed and evaluated term relationship network from the game logs and examined the types of created relationships.

RECENT PROGRESS AND UPCOMING CHALLENGES

Against the backdrop of the research presented in the above twelve chapters, let me now review my views on some of the most exciting progresses, current challenges and the times to come in the Semantic Web arena in near future.

Growing Role of Background Knowledge: Semantic Web researchers and early entrepreneurs knew (as exemplified by the first patent on developing and using Semantic Web technologies awarded in 2001) that with moderate effort, it is possible to create background knowledge and populated ontologies by aggregating and disambiguating high quality information and facts from multiple sources. It has also been long known that by using such knowledge bases, we can substantially improve information extraction and develop a variety of semantic tools and applications including semantic search, browsing, personalization, advertisement, etc. Over the past 3-5 years, several efforts to create such knowledge bases took place, of which Freebase is a showcase. What has drawn everyone’s attention to this aspect of semantic approach is Google’s acquisition of the company that created Freebase and significantly extending techniques largely known, but scaling it to the next level, to create Google Knowledge Base (GKB). Further on, applying GKB to enhance search (and I am sure other applications in future), has forever changed the importance of creating and using background or domain models for semantic applications. I believe this form of semantic application building will see the fastest growth in the near future. I have discussed related thoughts in my article titled “Semantics Scales Up”.

Growing Pains for Linked Open Data (LOD): Publication of over 300 large data sets with 30+ billion triples certainly draws the attention of many. Data holders will continue to find LOD an attractive vehicle to publish and share their data, so it will continue to grow at a rapid pace. Some of the data sets, more than others, will find additional usage as data reference, interlinking, and transformation. But in the near term, broader or aggregate usage of LOD will be a slog because we are running into some of the harder technical challenges: questionable quality of data and provenance, unconstrained and uneven use of semantics (e.g. same-as used inconsistently) and limited use of richer relationship types (part-of relationship, causality), and poor interlinking (lack of high quality alignment). We will need to have better handle of these issues along with a better ability to identify the most relevant and high quality data sets (a semantic search for LOD) and better alignment tools (not limited to just same-as), before we can start realizing the true promise of LOD. So, I would give it another five years to fully develop.

Democratization of Semantics: So far, we have paid the majority of our attention to knowledge representation, languages, and reasoning. Furthermore, a majority of the work focuses on documents in enterprises and on the Web, or uses structured data transformed into triples. But, what is even more exciting, is how semantics and Semantic Web technology (primarily through annotating data with respect to background knowledge or ontologies) is being used for improving interoperability and analysis of
different types of textual and non-textual data, esp. social data and data generated by sensors, devices, or Internet of Things. These types of data have long overtaken traditional document-centric data and structured databases in terms of volume, velocity, and variety. The type of semantics one needs to deal with for such (relatively) nontraditional data is of amazing variety. For example, in the Twitris system (http://twitris.knoesis.org), besides semantic annotation for spatial, temporal, and thematic elements associated with the tweets, semantics (aka meaning) also includes understanding people (about the poster and receiver), network (about interactions and flow of message), sentiment, emotion, and intent. For more in-depth treatment, see our just published book on semantics empowered Web 3.0. This is probably the most important development in my view and is likely to garner a much larger share of attention related to the application of semantics and semantic web technologies.

Amit Sheth
Kno.e.sis Center, Wright State University, USA

REFERENCES