

John Davies, Marko Grobelnik and Dunja Mladenic: Semantic knowledge management

Springer, 2009, 252 pp., Price: \$99 (ISBN: 978-3-540-88844-4)

David Vallet

© Springer Science+Business Media, LLC 2010

Semantic Knowledge Management edited by John Davies, Marko Grobelnik and Dunja Mladenic, provides an overview of the results of the Semantic Knowledge Technologies (SEKT) project,¹ co-funded by the EU 6th Framework programme, in which the contributors of this book have been mostly involved. This project spanned over three years from 2004 and focused on the integration of the research areas of Ontology Management, Machine Learning and Human Language Technologies. The results of this project are a number of tools and applications drawing from these areas of research. This book reviews the underlying technologies, the design, and the evaluation of the tools and applications. Most of the presented applications are also related to information retrieval, such as search engines, faceted browsers, knowledge extraction and sharing systems, etc. The book thus provides a number of examples on how to apply Semantic Web technologies to Information Retrieval applications.

The book is structured in four parts, in a bottom up approach. The first part describes the three semantic technologies involved in the project: ontology management, semantic knowledge discovery and natural language processing. The second part of the book presents several knowledge management tools that integrate these technologies. The third part presents a number of applications that make use of the introduced technologies and tools. The fourth part introduces three real industry-related uses cases, which, in turn, make use of these applications. The latter part has an emphasis on the evaluation methodologies that were designed and applied to validate the use of the introduced semantic technologies.

In the introductory chapter, the authors motivate the need for semantic technologies from the business point of view, and highlight the different goals that the Semantic Web community has established to achieve significant advances in this area: data integration, data sharing, and semantic search.

The first part of the books comprises three chapters. Chapter 2 introduces ontology management techniques. In particular, it describes a general ontology management

¹ <http://www.sekt-project.com/>

architecture, inspired in two existing management systems. It also briefly describes and gives examples of four more specific research areas related to ontology management: ontology mapping, ontology learning, ontology inconsistency management, and query answering. Chapter 3 provides a short introduction to knowledge discovery applied to ontology constructions from both unstructured and structured data. Three tools are presented in this chapter: OntoClassify (Grobelt and Mladenic 2005), an automatic instance classification system, OntoGen (Fortuna et al. 2007), an assisted ontology editor, and SEKTbar, an automatic ontology constructor based on the monitored actions of the user over the Web browser. Chapter 4 introduces ontology-based information extraction technologies, describing a learning algorithm to automatically classify ontology instances and extract relations between them from an unstructured corpus. Additionally, CLIE, a Controlled Language Information Extraction tool is presented, which is an ontology editing tool that supports the manipulation of ontologies by non-expert users in a simple semi-natural language.

Although this part of the book is introductory, it is not intended to be a survey of semantic technologies, but rather a compendium of the different solutions applied in the context of SEKT project. Still, the presented approaches and applications are suitably related to the state of the art. Thus, rather than a general overview of the state of the art of semantic technologies, the book is meant to provide examples on how these technologies can be applied in real world industry-related situations. Researchers and practitioners from the Information Retrieval community should not have major difficulty reading this book, although an inexperienced reader might need to get some prerequisite background from provided references in order to understand some of the involved concepts.

The second and third part of the book cover applications and tools that integrate the above technologies. Chapter 5 describes OntoSTUDIO,² an extendable ontology engineering environment which integrates reasoning, text mining and ontology learning capabilities, and also includes extensions such as ontology visualization, mapping, and querying. A brief survey of the state of the art on ontology engineering environments is included. The following chapters describe three different approaches for semantically-enhanced retrieval systems. Chapter 6 presents an extension over the well known KIM semantic annotation platform (Popov et al. 2004), which provides a user interface that allows searching over a selected set of instances. The system computes a co-occurrence value between instances and allows the user to track the trends and popularity of annotated instances. The browsing capabilities of this system are compared with a traditional faceted search paradigm, inasmuch as the ontology instances can represent the different facets of the annotated corpus. Chapter 7 presents Squirrel, a search tool that allows the execution of structured queries in natural language. Squirrel is able to present a meta-result page where the semantic information of the annotated documents is used to give an overview description of the results obtained from the search query. The authors contend that this can help the user understand the delivered results, and the search tasks themselves as a whole, as results can be further refined using the documents' related metadata. In Chap. 8, the Squidz system is introduced, which facilitates knowledge sharing between users. Using this tool, users can create semantic annotations of Web resources and share them within their community. As a user navigates over the Web, resources are recommended based on the inferred topic of the pages and the community's semantic annotations.

More examples on using the semantic tools are presented in the following chapters, including a natural language generator from ontology instances (Chap. 9), an example of a

² <http://www.ontoprise.de/en/home/products/ontostudio/>

social ontology extraction technique from an intranet mailing list database (Chap. 10), and a multi-document summarization technique with detection of duplicate instances (Chap. 11). Chapter 12 presents an interesting semantic space visualization technique, which applies a dimensionality reduction approach and takes into account a time dimension. Chapter 13 presents a semantic extension of a media wiki, where the identification of concepts and the representation of relations and valued attributes are supported.

The described tools and applications have an important focus on semantic search technologies, going from approaches for the execution of structured queries on natural language, to faceted search applications that exploit the concepts, taxonomies and properties represented within an ontology. Following the overall trend of the book, the description of these approaches is not technical, but rather consists of high-level explanations. This allows the reader to get a general idea about the approaches, their integration and evaluation. The provided references need to be followed in order to get a deeper understanding of the presented approaches. Chapter 12 and 13 felt somewhat off-topic to me in relation with the other presented applications as, contrary to the latter, they are not later integrated and evaluated in the last part of the book.

In the last part, three business use cases are described, which integrate technologies and tools presented earlier in the book. Chapter 14 introduces a digital library application, which integrates the semantic search and browsing technologies described in earlier chapters, on a large collection of over 5 million articles. A specialised frequently asked question system is presented in Chap. 15, where the semantic search capabilities and the ontology engineering and management tools are integrated and evaluated. Finally, Chap. 16 presents a use case focused on knowledge sharing and discovery through the set of semantic tools presented earlier in the book. Chap. 17 further complements the evaluation methodology described in the first two use cases. The evaluation methodology and the obtained results are detailed at length. The authors adopt a user-centred evaluation, where users are interviewed in early stages of the development phase in order to obtain a set of user requirements and perform a user evaluation over the deployed system, obtaining both qualitative and quantitative measures.

In this last part of the book, there is an emphasis on the evaluation methodology. Regarding this, the book is a very good reference for those readers who are planning to evaluate a semantic technology in similar conditions, as this part of the book gives specific insight and details on how to design the evaluation over real application and real use cases. Additionally, the book provides extensive details on the metrics that should be measured in a quantitative evaluation, and how to design user questionnaires in the requirement elicitation and qualitative evaluation phases.

To conclude, this book presents the design, implementation, and evaluation of several semantically-enhanced information retrieval applications. These applications cover a number of use cases described in the book, which exemplify how Semantic Web technologies can be successfully applied in the context of specific business areas. The book covers all the aspects of their development, ranging from the core Semantic Web technologies that have been integrated into these applications, to the design of a proper user-centred evaluation, including the description of the semantic tools shared by these applications and the elicitation of the user requirements. Perhaps the first part of the book, where the core semantic technologies are surveyed and introduced, may be somewhat concise for readers from the Information Retrieval community to fully understand the presented semantic-based approaches, but all in all the required references are provided for such readers who may need to get some previous background. The main value of this book is its comprehensiveness in the coverage of all the phases of the design and development of

semantic-based retrieval applications, including the complex and challenging issues involved in evaluating semantic-based approaches. I would thus recommend this book to anyone who wants to get a good grasp on how semantic technologies can be applied to enhance traditional information retrieval systems.

References

- Fortuna, B., Grobelnik, M., & Mladenic, D. (2007). Ontogen: Semi-automatic ontology editor. In *Proceedings of human interface and the management of information. Interacting in information environments* (pp. 309–318). Berlin: Springer.
- Grobelnik, M., & Mladenic, D. (2005). Simple classification into large topic ontology of web documents. In *Proceedings of 27th international conference on information technology interfaces* (pp. 188–193). New York: IEEE Press.
- Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D., & Kirilov, A. (2004). Kim—a semantic platform for information extraction and retrieval. *Journal of Natural Language Engineering*, 10(3–4), 375–392.

Semantic Knowledge Management
Integrating Ontology Management, Knowledge
Discovery, and Human Language Technologies
Davies, J.F.; Grobelnik, M.; Mladenić, D. (Eds.)
2009, X, 252 p., Hardcover
ISBN: 978-3-540-88844-4