

Preface

This book introduces theories, formalisms and applications that permit effective interaction between legal scholarly documents and their related semantic and formal descriptions, using the semantic technologies that have been employed to create the Semantic Web and Linked Data. Such an application of Semantic Web technologies to the world of documents and publishing is here termed *Semantic Publishing*.

Research and publishing in Law have a long tradition of applying such semantic technologies to real-world scenarios. In this context, the aim of this book is to describe and propose solutions to three of the main issues faced by the application of Semantic Publishing within the context of the legal scholarly documents, namely: the need for tools to link document text to a *formal representation of its meaning*; the requirement for complete *metadata schemas/ontologies* that can be used to describe documents using the language and vocabulary of the world of publishing; and the demand for effective *user interfaces* that permit such tools and schemas to be applied in practice.

The first section of this book describes *markup theory and technology*, whereby a better comprehension of a document can be derived from a description of its structural organisation and of the formal semantics contained within it, and from clarification of how these two types of markup—structural and semantic—can overlap and interact. In digital documents, the way we say something about a text is by the addition of markup, by which is meant additional statements, not normally rendered visible for human reading, that define and delimit structural or semantic features of the text. Markup has been used for years for annotating documents at all levels of granularity, from the digital document as a whole to its sub-components. However, the most commonly used document formats employed in publishing (i.e., XML and PDF) were not developed primarily to enable semantic enhancement and overlapping markup, although in principle it is possible to use them for this purpose. To facilitate the task of integrating potentially overlapping structural and semantic markup, I have developed an OWL-based markup metalanguage called *EARMARK*. *EARMARKed* documents are collections of addressable text fragments associated with OWL assertions that describe the structural and semantic properties of (parts of) that content.

Digital documents and their content form the focus of Semantic Publishing, which enables their automated discovery and connection to other document-related resources and contexts, such as other articles and raw data. Unfortunately, existing Semantic Web vocabularies are too abstract and incomplete to cover the requirements of the publishing world. There is thus an acute need for new vocabularies (ontologies) that comprehensively describe the different aspects of the publishing domain. In the central section of this book, I describe a suite of orthogonal and complementary OWL 2 DL ontology modules, the *Semantic Publishing and Referencing (SPAR) Ontologies*, that enable all aspects of the publishing process to be described in machine-readable metadata statements encoded using the Resource Description Framework (RDF).

The final section of this book describes new tools and interfaces that enable users to use and interact with such semantic technologies and to create semantic data. This provision of new tools and interfaces is particularly crucial for Semantic Publishing, since its end-users are generally publishers, researchers, readers, librarians, law scholars and the like, rather than computer scientists who are experts in semantic technologies. The use of semantic languages and vocabularies such as EARMARK and the SPAR Ontologies requires simple and easy-to-use human-computer interfaces that facilitate their application to Semantic Publishing by non-experts. Here I introduce four new tools developed to support users: two (*LODE* and *KC-Viz*) to aid in understanding an ontology; a third (*Graffoo*) that assists in presenting such an ontology, or fragments of it, in comprehensible formalised diagrams; and a fourth (*Gaffe*) that is useful when employing such an ontology to create semantic definitions of real data.

Semantic Web Technologies and Legal Scholarly
Publishing

Peroni, S.

2014, XXIII, 304 p. 211 illus., 46 illus. in color.,

Hardcover

ISBN: 978-3-319-04776-8