

Preface

It has become sort of a cliché nowadays to mention how rapidly textual information is growing and how the World Wide Web has assisted in this growth. This, however, does not shadow the fact that such explosive growth will only intensify for years to come, and more new challenges and opportunities will arise. Advances in fundamental areas such as information retrieval, machine learning, data mining, natural language processing, and knowledge representation and reasoning have provided us with some relief by uncovering and representing facts and patterns in text to ease the management, retrieval, and interpretation process. *Information retrieval*, for instance, provides various algorithms to analyse associations between components of a text using vectors, matrices, and probabilistic theorems. *Machine learning* and *data mining*, on the other hand, offer the ability to learn rules and patterns out of massive datasets in a supervised or unsupervised manner based on extensive statistical analysis. *Natural language processing* provides the tools for analysing natural language text on various language levels (e.g. morphology, syntax, semantics) to uncover manifestations of concepts and relations through linguistic cues. *Knowledge representation and reasoning* enable the extracted knowledge to be formally specified and represented such that new knowledge can be deduced.

The realization that a more systematic way of consolidating the discovered facts and patterns into an organised, higher level construct to enhance everyday applications (e.g. Web search) and enable intelligent systems (e.g. Semantic Web) eventually gave rise to ontology learning and knowledge discovery. *Ontologies* are effectively formal and explicit specifications, in the form of concepts and relations, of shared conceptualisations, while *knowledge bases* can be obtained by populating the ontologies with instances. Occasionally, ontologies contain axioms for validation and constraint definition. As an analogy, consider an ontology as a cupcake mould and knowledge bases as the actual cupcakes of assorted colours, tastes, and so on. *Ontology learning* from text is then essentially the process of deriving the high-level concepts and relations from textual information. Considering this perspective, *knowledge discovery* can refer to two things, the first denotation being the uncovering of relevant instances from data to populate the ontologies (also known as *ontology population*), and the second, more general sense being the searching of data for useful patterns. In this book, knowledge discovery can mean either one of the two.

Being a young and exciting field, ontology learning has witnessed a relatively fast progress due to its adoption of established techniques from the related areas discussed above. Aside from the inherent challenges of processing natural language, one of the remaining obstacles preventing the large-scale deployment of ontology learning systems is the bottleneck in handcrafting structured knowledge sources (e.g. dictionaries, taxonomies, knowledge bases) and training data (e.g. annotated text corpora). It is gradually becoming apparent that in order to minimize human efforts in the learning process, and to

improve the scalability and robustness of the system, static and expert crafted resources may no longer be adequate. An increasing amount of research effort is being directed towards harnessing *collective intelligence on the Web* as an attempt to address this major bottleneck. At the same time, as with many fields before ontology learning, the process of maturing has triggered an increased awareness of the difficulties in automatically discovering all components of an ontology, i.e. terms, concepts, relations, and especially axioms. This gives rise to the question of whether the ultimate goal of achieving *full-fledged formal ontologies* automatically can be achieved. While some individuals dwell on the question, many others have moved on with a more pragmatic goal, which is to focus on learning *lightweight ontologies* first, and extend them later if possible. With high hopes and achievable aims, we are now witnessing a growing interest in ontologies across different domains that require interoperability of semantics and a touch of intelligence in their applications.

This book brings together some of the latest work on three popular research directions in ontology learning and knowledge discovery today, namely, (1) the use of Web data to address the knowledge and training data preparation bottleneck, (2) the focus on lightweight ontologies, and (3) the application of ontologies in different domains and across different languages. Section I of the book contains chapters covering the use of a wide range of existing, adapted and emerging techniques for extracting terms, concepts and relations to construct ontologies and knowledge bases. For instance, in addition to traditional clustering techniques reported in Chapter III, a new topic extraction technique is being devised as in Chapter IV to offer alternative ways for discovering concepts. Chapter II, on the other hand, promotes the new application of existing deep semantic analysis methods for ontology learning in general. The use of semi-structured Web data such as Wikipedia for named entity recognition, and the question of how can this be applicable to ontology learning are also investigated in Chapter V. The focus of Chapter I is on the construction of practical, lightweight ontologies for three domains. As for Chapter VI and VII, the authors mainly investigate the use of a combination of data sources, both local and from the Web, to discover hierarchical and non-taxonomic relations. In Section II, the authors look at how ontologies and knowledge bases are currently being applied across different domains. Some of the domains covered by the chapters in this section include biomedical (Chapter VIII and IX), humanities (Chapter X) and enterprise knowledge management (Chapter XI). This book ends with Section III that covers chapters on the use of social data (Chapter XII) and parallel texts (Chapter XIII and XIV), which may or may not be from the Web for learning social ontologies, incorporating trust into ontologies, and improving the process learning ontologies.

This volume is both a valuable standalone as well as a great complement to the existing books on ontology learning that have been published since the turn of the millennium. Some of the previous books focus mainly on techniques and evaluations, while others look at more abstract concerns such as ontology languages, standards, and engineering environments. While the background discussions on the techniques and evaluations are indispensable, the focal point of this book remains on emerging research directions involving the use of Web data for ontology learning, the learning of lightweight as well as cross-language ontologies, and the involvement of ontologies in real-world applications. We are certain that the content of this book will be of interest to a wide ranging audience. From a teaching viewpoint, the book is intended for undergraduate students at the final year level, or postgraduate students who wish to learn about the basic techniques for ontology learning. From a researcher's and practitioner's point of view, this volume will be an excellent addition outlining the most recent progress to complement basic references in ontology learning. A basic familiarity with natural language processing, probability

and statistics, and some fundamental Web technologies such as wikis and search engines is beneficial to the understanding of this text.

Wilson Wong

The University of Western Australia, Australia

Wei Liu

The University of Western Australia, Australia

Mohammed Bennamoun

The University of Western Australia, Australia