Yuan-Fang Li · Wei Hu
Jin Song Dong · Grigoris Antoniou
Zhe Wang · Jun Sun
Yang Liu (Eds.)

LNCS 10055

Semantic Technology

6th Joint International Conference, JIST 2016
Singapore, Singapore, November 2–4, 2016
Revised Selected Papers

Springer

Lecture Notes in Computer Science 10055

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7409

Yuan-Fang Li · Wei Hu
Jin Song Dong · Grigoris Antoniou
Zhe Wang · Jun Sun
Yang Liu (Eds.)

Semantic Technology

6th Joint International Conference, JIST 2016
Singapore, Singapore, November 2–4, 2016
Revised Selected Papers

Springer

Editors
Yuan-Fang Li
Information Technology
Monash University
Melbourne, VIC
Australia

Wei Hu
Computer Science and Technology
Nanjing University
Nanjing
China

Jin Song Dong
Computer Science
National University of Singapore
Singapore
Singapore

Grigoris Antoniou
University of Huddersfield
Huddersfield
UK

Zhe Wang
Information and Communication Technology
Griffith University
Brisbane, QLD
Australia

Jun Sun
ISTD
Singapore University of Technology and Design
Singapore
Singapore

Yang Liu
Computer Science and Engineering
Nanyang Technological University
Singapore
Singapore

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-50111-6 ISBN 978-3-319-50112-3 (eBook)
DOI 10.1007/978-3-319-50112-3

Library of Congress Control Number: 2016959173

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the papers presented at JIST 2016: the 6th Joint International Semantic Technology Conference held during November 2–4, 2016, in Singapore. JIST 2016 was co-hosted by National University of Singapore, Nanyang Technological University (Singapore), and Monash University (Australia). JIST is a regional federation of semantic technology-related conferences. It attracts many participants from mainly the Asia Pacific region and often Europe and the USA. The mission of JIST is to bring together researchers in semantic technology research and other areas of semantic-related technologies to present their innovative research results and novel applications.

The main topics of JIST 2016 include ontology and reasoning, linked data, and knowledge graph, among others. JIST 2016 consisted of two keynotes, a main technical track, including (full and short) papers from the research and the in-use tracks, a poster and demo session, a workshop, and two tutorials. There were a total of 34 submissions for the main technical tracks from 17 countries. All papers were reviewed by at least three reviewers and the results were rigorously discussed by the program co-chairs. In all, 16 full papers (47%) and eight short papers were accepted in the technical tracks.

The paper topics are divided into six categories: Ontology and Data Management, Linked Data, Information Retrieval and Knowledge Discovery, RDF and Query, Knowledge Graph, and Applications of Semantic Technologies.

We would like to thank the JIST Steering Committee, Organizing Committee, and Program Committee for their significant contributions. We would also like to especially thank the co-hosts for their support in making JIST 2016 a successful and memorable event. Finally, we would like to express our appreciation to all speakers and participants of JIST 2016. This book is an outcome of their contributions.

November 2016

Yuan-Fang Li
Wei Hu
Jin Song Dong
Grigoris Antoniou
Zhe Wang
Jun Sun
Yang Liu

Organization

Program Committee

Paolo Bouquet	University of Trento, Italy
Nopphadol Chalortham	Silpakorn University, Thailand
C. Chantrapornchai	Kasetsart University, Thailand
Gong Cheng	Nanjing University, China
Paola Di Maio	ISTCS.org/IIT Mandi, India
Stefan Dietze	L3S Research Center, Germany
Dejing Dou	University of Oregon, USA
Jae-Hong Eom	Seoul National University, South Korea
Naoki Fukuta	Shizuoka University, Japan
Volker Haarslev	Concordia University, Canada
Armin Haller	Australian National University, Australia
Masahiro Hamasaki	National Institute of Advanced Industrial Science and Technology (AIST), Japan
Sungkook Han	Wonkwang University, South Korea
Koiti Hasida	AIST, Japan
Wei Hu	Nanjing University, China
Eero Hyvönen	Aalto University, Finland
Ryutaro Ichise	National Institute of Informatics, Japan
Vahid Jalali	Indiana University, USA
Jason Jung	Chung-Ang University, South Korea
Yong-Bin Kang	Monash University, Australia
Takahiro Kawamura	Japan Science and Technology Agency, Japan
Pyung Kim	Jeonju National University of Education, South Korea
Seiji Koide	Ontolonomy, LLC, Japan
Kouji Kozaki	Osaka University, Japan
Seungwoo Lee	KISTII, South Korea
Tony Lee	Saltlux, Inc., South Korea
Yuan-Fang Li	Monash University, Australia
Riichiro Mizoguchi	Japan Advanced Institute of Science and Technology, Japan
Takeshi Morita	Keio University, Japan
Ralf Möller	Universität zu Lübeck, Germany
Shinichi Nagano	Toshiba Corporation, Japan
Ikki Ohmukai	National Institute of Informatics, Japan
Artemis Parvizi	Oxford University Press, UK
Yuzhong Qu	Nanjing University, China
Ulrich Reimer	University of Applied Sciences St. Gallen, Switzerland

Giorgos Stoilos	National Technical University of Athens (NTUA), Greece
Umberto Straccia	ISTI-CNR, Italy
Boontawee Suntisrivaraporn	DTAC, Thailand
Hideaki Takeda	National Institute of Informatics, Japan
Holger Wache	University of Applied Science Northweastern Switzerland, Switzerland
Haofen Wang	East China University of Science and Technology, China
Peng Wang	Southeast University, China
Xin Wang	Tianjin University, China
Zhe Wang	Griffith University, Australia
Krzysztof Wecel	Poznan University of Economics, Poland
Gang Wu	College of Information Science and Engineering, Northeastern University, China
Guohui Xiao	KRDB Research Centre, Free University of Bozen-Bolzano, Italy
Bin Xu	DCST, Tsinghua University, China
Yasunori Yamamoto	Database Center for Life Science, Japan
Xiang Zhang	Southeast University, China
Yuting Zhao	IBM Italy, Italy
Amal Zouaq	Royal Military College of Canada, Canada

Organizing Committee

General Chairs

Jin Song Dong	National University of Singapore, Singapore
Grigoris Antoniou	University of Huddersfield, UK

Program Co-chairs

Yuan-Fang Li	Monash University, Australia
Wei Hu	Nanjing University, China

Publicity Chair

Zhe Wang	Griffith University, Australia

Local Chair

Jun Sun	Singapore University of Technology and Design, Singapore

Workshop Co-chairs

Xin Wang	Tianjing University, China
Hanmin Jung	Korea Institute of Science and Technology Information, Korea

Tutorial Co-chairs

Armin Haller	Australian National University, Australia
Gong Cheng	Nanjing University, China

Poster and Demo Co-chairs

Zhichun Wang	Beijing Normal University, China
Kouji Kozaki	Osaka University, Japan

Finance Chair

Yang Liu	Nanyang Technological University, Singapore

Publicity Chair

Haofen Wang	East China University of Science and Technology, China

Keynotes

Managing Dynamic Ontologies: Belief Revision and Forgetting

Kewen Wang

Griffith University, Brisbane, Australia
k.wang@griffith.edu.au

Ontologies have recently been used in a wide range of practical domains such as e-Science, e-Commerce, medical informatics, bio-informatics, and the Semantic Web. An *ontology* is a formal model of some domain knowledge of the world. It specifies the *formalization* of the domain knowledge as well as the *meaning* (semantics) of the formalization. The Web Ontology Language (OWL), with its latest version, OWL 2, is based on description logics (DLs). Thus, an ontology is often expressed as a knowledge base (KB) in DLs, which consists of both terminological knowledge (or schema information) in the TBox and assertional knowledge (or data information) in the ABox. As with all formal knowledge structures, ontologies are not static, but may evolve over time. Indeed, ontology engineering is described as a life-cycle, which is based on evolving prototypes and specific techniques peculiar to each ontology engineering activity. An important and challenging problem is thus how to effectively and efficiently modify ontologies.

In this talk, we discuss some recent developments and challenges for two paradigms of ontology changes. We focus on model-based approaches.

Knowledge Update: Outdated and incorrect axioms in an ontology have to be eliminated from the ontology and newly formed axioms have to be incorporated into the ontology. In the field of belief change, extensive work has been done on formalising various kinds of changes over logical knowledge bases. In particular, elimination of old knowledge is called contraction and incorporation of new knowledge is called revision. The dominant approach in belief change is the so called AGM framework. Regardless of its wide acceptance, the AGM framework is incompatible with DLs due to its assumption on an underlying logic that includes propositional logic. The incompatibility is the major difficulty in defining DL contraction and revision. Additionally, DL revision is more involved than AGM revision. AGM revision aims to resolve any inconsistency caused while incorporating a new formula. Since a meaningful DL ontology has to be both consistent and coherent (i.e, absence of unsatisfiable concepts), DL revision has to resolve not only inconsistency but also incoherence. Finally, DL contraction and revision should lead to tractable instantiations and at the same time respecting the mathematical properties of AGM contraction and revision.

Forgetting: To support the reuse and combination of ontologies in Semantic Web applications, it is often necessary to obtain smaller ontologies from existing larger ontologies. In particular, applications may require the omission of many terms, e.g., concept names and role names, from an ontology. However, the task of omitting terms from an ontology is challenging because the omission of some terms may affect the relationships between the remaining terms in complex ways. The technique of forgetting provides an effective way for extracting modules from a large ontology.

The Rise of Approximate Ontology Reasoning: Is It Mainstream Yet?

Jeff Z. Pan

University of Aberdeen, Aberdeen, UK

The last five years have seen a growing volume and complexity of ontologies and large-scale linked data available,[1] which present a pressing need for efficient and scalable ontology reasoning services. Major technology vendors are starting to embrace semantic technologies by supporting new standards and integrating with state of the art semantic tools. For example, in their new release 12.1, Oracle Spatial and Graph supports both RDF and OWL2-EL natively,[2] and integrates with an OWL2-DL reasoner (TrOWL) via OWL-DBC.[3]

The second version of the ontology standard OWL (Web Ontology Language) offers a family of ontology languages, including OWL2-DL, the most expressive decidable language in the family, and three tractable sub-languages of OWL2-DL, i.e. OWL2-EL, OWL2-QL and OWL2-RL. Such a two-layered language architecture allows approximate reasoning for OWL2-DL, by approximating OWL2-DL ontologies to those in its tractable sub-languages, so as to exploit efficient and scalable reasoners of the sublanguages. This is motivated by the fact that real-world knowledge and data are hardly perfect or completely digitalised. State of the art approximate reasoners, such as the TrOWL reasoner, can out-perform sound and complete reasoners in time constrained sound-and-complete reasoner competitions, such as the ORE competitions.

In this talk, we will look into how and why approximate reasoners work. Indeed, approximation approaches bring a new dimension – quality, in terms of completeness and soundness of reasoning, into the trade-off between expressiveness and performance, attempting to strike a balance among the three. Once we start to consider such a third dimension, many interesting questions follows: What are the typical approximate reasoning approaches? Should we approximate the input ontology or the input query? Are approximations always finite and unique? Given an ontology and some target queries, are there any best approximations? Why do some approximate reasoning algorithms lose many reasoning results, while others can enjoy high recall? Are approximate reasoning algorithms relevant to optimisations for sound and complete reasoners? Can we extend approximate reasoning algorithm with some post-processing to ensure soundness and completeness? I will discuss many of these questions, in the context of the TrOWL reasoner and related work, and share some thoughts on what approximate reasoning might bring in the near future.

---

1 http://lod-cloud.net/state/.

2 http://download.oracle.com/otndocs/tech/semantic_web/pdf/semtech_datamining_v8.pdf.

3 http://download.oracle.com/otndocs/tech/semantic_web/pdf/trowl_integration_with_orasag.pdf.

Contents

Information Retrieval and Knowledge Discovery

RDF and Query

Knowledge Graph

Applications of Semantic Technologies

Ontology and Data Management

How Can Reasoner Performance of ABox Intensive Ontologies Be Predicted?

Isa Guclu[1], Carlos Bobed[2], Jeff Z. Pan[1(✉)], Martin J. Kollingbaum[1], and Yuan-Fang Li[3]

[1] University of Aberdeen, Aberdeen, UK
jeff.z.pan@abdn.ac.uk
[2] University of Zaragoza, Zaragoza, Spain
[3] Monash University, Melbourne, Australia

Abstract. Reasoner performance prediction of ontologies in OWL 2 language has been studied so far from different dimensions. One key aspect of these studies has been the prediction of how much time a particular task for a given ontology will consume. Several approaches have adopted different machine learning techniques to predict time consumption of ontologies already. However, these studies focused on capturing general aspects of the ontologies (i.e., mainly the complexity of their TBoxes), while paying little attention to ABox intensive ontologies. To address this issue, in this paper, we propose to improve the representativeness of ontology metrics by developing new metrics which focus on the ABox features of ontologies. Our experiments show that the proposed metrics contribute to overall prediction accuracy for all ontologies in general without causing side-effects.

Keywords: Semantic web · Ontology reasoning · Prediction · Random forests · Knowledge graph · Practical reasoning

1 Introduction

Semantic technologies have been utilized in various application domains for assisting knowledge management thus far, e.g., data management [13] and software engineering [17]. The worst case complexity 2NEXPTIME-complete [6] of OWL 2 DL, the most expressive profile of OWL 2, constitutes a bottleneck for performance critical environments. Empirical studies show that even the EL profile, with PTIME-complete complexity and less expressiveness, can become too time-consuming [4,11]. To have a scalable environment for implementing semantic technologies, an accurate prediction of ontology time consumption which will guide us about the feasibility of ontology reasoning is needed.

There have been several studies regarding the performance prediction of ontologies. Kang et al. [10] investigated the *hardness category* (categories according to reasoning time) for reasoner-ontology pairs and used machine learning techniques to make a prediction. Using FaCT++ [25], HermiT [5], Pellet [23], and TrOWL [16,18,20,24], they reached high accuracy in terms of hardness category, but not reasoning time.

Y.-F. Li et al. (Eds.): JIST 2016, LNCS 10055, pp. 3–14, 2016.
DOI: 10.1007/978-3-319-50112-3_1

In another study, Kang et al. [12] investigated regression techniques to predict reasoning time. They made experiments using reasoners FaCT++, HermiT, JFact, MORe [21], Pellet and TrOWL with *their syntactic metrics* as features. These metrics are generally effective when there is a balance between TBox axioms and ABox axioms. Our experiments show that accuracy of these metrics decreases as ABox axiom sizes increase. As ABox constitutes the data in an ontology [1,8,27], where TBox constitutes the schema, an approach that can capture the changes in the ABox in a more detailed way is needed to make accurate overall predictions. As observed by Bobed et al. [2], there is an interest in using semantic technologies in mobile devices. In such scenarios, TBox axioms are expected to be more static and the ABox axioms (data) tend to be more frequently changing which necessitates high accuracy in ABox performance prediction. In this paper, we aim to investigate what metrics could help further improve reasoner predictions of ABox intensive ontologies.

Our main contributions can be summarized as follows.

1. We propose an initial set of metrics which estimate the complexity of the TBox concepts and propagates it into the estimated complexity of the ABox.
2. We show that our proposed new metrics for representing the structure of ontologies from the ABox perspective indicate a good research path to improve the accuracy of predicting time consumption of ontology reasoning.

The rest of the paper is as follows. In Sect. 2, we present some related works to place our proposal. In Sect. 3, we define the metrics that we propose in our ongoing work. In Sects. 4 and 5, we explain our experimental settings and the achieved results, respectively. Finally, in Sect. 6, we make some conclusions and draw some future work.

2 Related Work and Background

Ontology metrics, which are features of the ontology expressed numerically or categorically to represent the structure of an ontology, have been effectively utilised in analysing the complexity [28], energy consumption on mobile devices [7], cohesion [26], quality [3] and population task [15] of ontology reasoning.

Kang et al. [10] proposed a set of metrics in 2012 to classify raw reasoning times of ontologies into five large categories: [0 s.–100 ms.], (100 ms.–1 s.], (1 s.–10 s.], (10 s.–100 s.] and (100 s.–∞). Despite the high accuracy of prediction, over an 80%, this approach does not provide actual reasoning time but time categories, which may become obsolete or meaningless according to needs of implementation.

In 2014, Kang et al. [12] extended their work and proposed a new set of metrics to predict actual reasoning time by developing regression models. They extended the previous 27 metrics [10,28] and developed a set of 91 metrics that include 24 ontology-level (ONT) metrics, 15 class-level (CLS) metrics, 22 anonymous class expression (ACE) metrics, and 30 property definition and axiom (PRO) metrics.

While a high number of metrics are usually proposed by researchers, Sazonau et al. [22] proposed instead a local method which involved selecting a *suitable*, small subset of the ontology, and making extrapolation to predict total time consumption of ontology reasoning using the data coming from the processing of such small subset. To do so, they used *Principal Component Analysis* (PCA) [9]. In their experiments, Sazonau et al. [22] observed that 57 of the studied features can be replaced by just one or two features. Using a sample of size of a 10 % of the ontology for reasoning, they argue that they reached good predictions with simple extrapolations. They list advantages of their method as: (1) more accurate performance predictions, (2) not relying on an ontology corpus, (3) not being biased by this corpus, and (4) being able to obtain information about reasoner's behaviour of linear/nonlinear predictability on the corpus. A remarkable contribution of this approach is that it saves researchers from the difficulty/risk of selecting an unbiased corpus [14], which is very difficult while checking the validity of the prediction model and accuracy of the prediction. However, making reasoning with the 10 % of an ontology may not always be applicable especially when the ontology requires high reasoning times.

3 Our Approach

Our claim is that increasing the expressivity of ontology metrics directly helps increasing the accuracy of all the above studies, and enables new studies that target a more feasible implementation environment for semantic technologies.

Part of 91 metrics proposed by Kang et al. [12] are obtained by transforming an ontology into a graph which grasps the relationship between of ABox and TBox axioms. However, their approach calculates the effect of ABox axioms up to a certain extent. It is apparent that connected ABox axioms are more prone to cause more inferences than disconnected ABox axioms. These connections can trigger reasoning time enormously when they come along with a complex TBox. In our work, we have observed that the models trained with this set of 91 metrics begin to lose accuracy in predicting time consumption of ontologies as the ratio between the amount of ABox axioms and TBox axioms increases.

Thus, we propose to include the propagation of the complexity of the TBox into the ABox. To do so, we extend this set of metrics with our 15 *Class Complexity Assertions (CCA)* metrics, which can contribute to performance prediction of ontologies especially when we deal with ontologies which are ABox intensive (i.e., they exhibit a high ABox/TBox ratio). Experiment results and source codes are accessible[1].

3.1 Class Complexity Assertions Metrics

As above mentioned, to capture the interactions between the complexity of the different elements of the TBox and the individuals asserted in the ABox, we have

[1] http://sid.cps.unizar.es/projects/OWL2Predictions/JIST16/.