Towards Semantic Applications: from Knowledge Management to Data Publishing on the Web

Dung Xuan Thi Le
Semantic Software Asia Pacific
100 Miller Street, North Sydney, 2060,
New South Wales, Australia
dle@semanticsoftware.com.au

Michel Heon
Cotechnoe Inc
2356 Ch Bourbonniere
Lachute (Québec), Canada
heon@cotechnoe.com

Nick Volmer
Semantic Software Asia Pacific
100 Miller Street, North Sydney, 2060,
New South Wales, Australia
nvolmer@semanticsoftware.com.au

ABSTRACT
This tutorial is a three-part contribution using semantic web technologies. The first is to highlight the means to use visual representation of resources, properties, classes, etc. to create an ontology. The second is to underpin a series of pruning methods for extracting sub-domain ontologies from a given large domain ontology without losing properties, axioms or integrity. The third is to outline the templating framework that enables data transformation to create a data integration platform for semantic application purposes. Although there is still more potential work in relation to semantic applications waiting to be addressed, we hope this tutorial will bring some interest to the audience and raise some awareness of available tools for moving towards semantic applications. Additionally, we hope our applications are useful for solving some outstanding issues in relation to ontology engineering and data integration from a semantic computing aspect.

Keywords
Semantic Web; Semantic Web Technology; Ontology; Ontology Editor; Ontology Pruner; Semantic Platform.

1. INTRODUCTION
Semantic web technologies focus on supporting data expression in a common language for interoperability together with knowledge capturing, representation, reuse and sharing. In the semantic management space, an ontology can be used as a knowledge representation language, and several textual syntaxes exist such as Triples, Manchester and Turtle, etc. for ontology construction. Most situations need to represent knowledge in a graphical mode, for example, for knowledge elicitation, knowledge sharing amongst humans or knowledge based systems modelling. In addition, existing public ontologies are normally large and complex. Each of these ontologies could describe many specific sub-domains. Understanding such a large ontology structure or reusing it to address a specific sub-domain will result in a high cost, which can be avoided. Given a significantly large master data repository, we can expose the data in Resource Description Framework (RDF) triples and at the same time, extract a

knowledge representation that describes the extracted information for management and sharing purposes [11][3]. To extract data and capture the knowledge in an effective manner, it should be made more adaptable for users who face challenges due to having very little or no semantic web knowledge. In the semantic integration space, questions such as: how flexible and adaptable are the entities, attributes and relationships being captured; how can inferences be enabled without the need to use a standard rule engine or reasoners; how are RDF triples being efficiently managed for manipulation, performance and scalability purposes; etc., are important and still need to be addressed.

2. TOPIC DISCUSSION
In this tutorial, we will learn how to use a visual representation of resources, classes, properties, individuals, restrictions, etc. to build an ontology. We will demonstrate OntoCASE4GOWL [5][7] (Ontology Case tool for Graphical Web Ontology Language) to show how to represent ontological knowledge in a graphical mode [6][8].

Next, we will learn about a rule based pruning methodology for ontologies. In this part, we will highlight the objective and the need for pruning ontologies which allows us to address several existing pruning techniques [1][9][10][13]. We will then highlight the associated challenges with these techniques which motivate us to investigate a series of practical pruning methods consisting of five ontology rule-based pruning methods including *full graph*, *subclass graph*, *semi graph*, *node by node*, and *common ancestor* for pruning ontologies. We will present the Semantiro Platform's Ontology Management Suite, referred to as Ontocuro, in which the rule-based pruning methods have been implemented.

Finally, we will learn about a templating framework that allows us to create a mapping, with built-in semantic rules for inferences driven by SPARQL translation. This is used for automating the transformation process of extracting data and creating knowledge which will then be exposed as RDF triples. We will also learn about approaches to store and manage ontologies, data and inferred data to avoid unnecessary cost overheads. We will demonstrate these frameworks and the approach via a transformation process on an integration platform referred to as Datacuro, the Semantiro Platform's Data Management Suite, which follows W3C recommendations for semantic web syntaxes and languages [2][3][4][11][12].

3. DURATION AND SESSION
The tutorial will run as a full-day event that is divided into two sessions. In the first session, we will present the knowledge modelling syntax for graphical web ontology language. We will

then continue to address the challenges of deriving a specific sub-domain ontology from an existing large ontology without breaking the rules and axioms. Finally, we will present the integration framework on data transformation and the challenges in enabling inferencing on large datasets for semantic processing and integration purposes. In the second session, we will provide case studies to the audience to have hands-on experience with our ontology management and data integration suites.

4. AUDIENCE
The event targets researchers and practitioners who are interested in applying semantic computing to explore solutions for creating ontologies, pruning existing ontologies and moving towards semantic applications. The technologies and topics in this tutorial are relevant to researchers, people from IoT, cognitive computing communities, as well as social media, health, oil industries, etc., who want to put their existing information (data) into a common language for semantic applications or for publishing data on the web.

We will provide a cloud-based practical exercise environment to the audience. We encourage attendees to bring a laptop, which will allow them to participate in the hand-on exercises running in the second session of the tutorial.

5. RELEVANCE
 In the context of the emergence of intelligent solutions carried by the web, government organisations and industry are undertaking a massive shift towards semantic technologies. Two needs related to the future success of the web are currently expressed by the industry players: the need to graphically represent the content of an ontology and the need to simply and quickly publish linked data on the web. This tutorial was prepared by presenters, who have been semantic and data specialists for a number of years.

6. EQUIPMENT
We require a projector and microphone for the presentation. For running the hands-on tutorials with attendees, we require a good internet connection for connecting to Amazon cloud.

7. SUPPORT MATERIALS
We will provide further details such as links to download software and information related to this tutorial at http://semanticsoftware.com.au/whats-on/press-and-media/www-tutorial-2017/

8. BIO
Dung Xuan Thi Le holds a PhD in Semantic Transformation for XML Queries from Macquarie University. Dr. Le joined Semantic Software (SSAP) as Chief Data Scientist in 2013 to lead the research team to conduct and prototype research activities in semantic computing space and to make recommendations towards the development of Semantiro suite. Prior to joining SSAP, she spent 9 years working globally in software development, technical support and was made responsible for overseas markets and high-profile customers. She had successfully commissioned many projects in South-East Asia and Middle-East. She graduated in 2006 with a Master of Science in Information Systems at La Trobe University. She was a Senior Global Support Analyst at Oracle Corporation in Australia for 4 years. Dzung joined Business Intelligence (BI) development team at Downer Group in Sydney between 2011 and 2013, where she specialised in Oracle BI Enterprise Suite.

Michel Héon holds a PhD in cognitive computing from the University of Québec at Montréal and founding president of Cotechnoe a semantic web consulting company. Over the past two decades, he has developed strong skills in computing and development of artificial intelligence applications in the research and industry context. He is particularly expert in software engineering, as well as in knowledge engineering and ontology modeling. Currently, Dr. Héon is interested in the design of a Graphical syntax for the Web Ontology Language (GOWL) in addition as developing OntoCASE4GOWL an Ontology Case Tool for GOWL. For the French community, Michel is the author of the book "*Web sémantique et modélisation ontologique avec G-OWL*" a book intended for programmers wishing to develop a semantic web application in Java.

Nick Volmer joined Semantic Software (SSAP) to lead an innovative development team and to challenge his conventional IT skills and thinking by contributing to the delivery of cognitive computing solutions which form part of the third wave of computing. Prior to joining SSAP, Nick spent a number of years as a consultant in the USA before he moved to Australia initially contracting for AGL. He later joined permanently working on web based solutions (primarily Java) and continued to expand his leadership competencies. Most recently Nick held the position of Associate Director within Technology at Macquarie Group, where he managed the largest portfolio of technology work in Risk Management IT, leading 3 development teams supporting various business units.

9. REFERENCES
[1] J. Conesa and A. Olive. Method for Pruning Ontologies in the Development of Conceptual Schemas of Information Systems, 2006. In Journal of Data Semantics. Vol. 3870, 2006. pp. 64-90.

[2] S. Das and S. Sundarar, R. Cyganiak. R2RML: RDF to RDF Mapping Language, 2012. https://www.w3.org/TR/r2rml/

[3] P. Hayes. RDF Semantics, 2004. https://www.w3.org/TR/2004/REC-rdf-mt-20040210/

[4] P. Hitzler, M. Krotzsch, B. Parsia, P. Patel-Schneider and S. Rudolph. OWL 2 Web Ontology Language Primer (Second Edition), 2012.https://www.w3.org/TR/2012/REC-owl2-primer-20121211/

[5] M. Héon. OntoCASE4GOWL a modeling tool for Graphical Semantic Web Ontology Language (GOWL), 2006. Available: http://www.cotechnoe.com/ontocase4gowl

[6] M. Héon, R. Nkambou, C. Langheit. Toward G-OWL: A Graphical, Polymorphic And Typed Syntax For Building Formal OWL2 Ontologies. In 25th International Conference Companion on World Wide Web. 2016.

[7] M. Héon, R. Nkambou, M. Gaha. OntoCASE4G-OWL: Towards an modeling tool for G-OWL a visual syntax for RDF/RDFS/OWL2. In 5th International Semantic Web Conference DEMO-Session, 2016.

[8] M. Héon. Web sémantique et modélisation ontologique (avec G-OWL): Guide du développeur Java sous Eclipse, Collection Epsilon ed.: Editions ENI, 2014.

[9] J-U. Kietz, A. Maedche and R. Volz. A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet. In Proceedings of EKAW-2000 Workshop, 2000.

[10] J. Kim, J. Conesa and J. Hilliard. Pruning Bio-Ontologies. In 40th Annual Hawaii International Conference, 2007. pp.196c-196c.

[11] F. Manola and E. Miller. RDF Primer, 2004. https://www.w3.org/TR/2004/REC-rdf-primer-20040210/

[12] D. McGuinness and F. Harmelen. OWL Web Ontology Language, 2004. https://www.w3.org/TR/2004/REC-owl-features-20040210/

[13] J. Zhang and Y. Lv. An Approach of Refining the Merged Ontology. In 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), 2012. pp.802-807.