Knowledge-Based Systems 80 (2015) 34–47

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Ontology usage analysis in the ontology lifecycle: A state-of-the-art review

Jamshaid Ashraf [a,*], Elizabeth Chang [b], Omar Khadeer Hussain [b], Farookh Khadeer Hussain [c]

[a] Jazba Technologies Pty Ltd, PO Box 848, Cannington, Perth, WA 6987, Australia
[b] School of Business, University of New South Wales, ADFA, Canberra, Australia
[c] Decision Support and e-Service Intelligence Lab, Quantum Computation and Intelligent Systems Lab, School of Software, University of Technology, Sydney, NSW, Australia

ARTICLE INFO

Article history:
Received 30 October 2014
Received in revised form 14 February 2015
Accepted 19 February 2015
Available online 5 March 2015

Keywords:
Ontology
Ontology lifecycle
Usage analysis
Ontology usage analysis framework
Semantic web

ABSTRACT

The Semantic Web envisions a Web where information is accessible and processable by computers as well as humans. Ontologies are the cornerstones for realizing this vision of the Semantic Web by capturing domain knowledge through the defined terms and the relationship between them to provide a formal representation of the domain with machine-understandable semantics. Given the importance of ontologies, a significant amount of work in the literature has been done on knowledge representation on the Web which includes ontology development (ontology engineering), ontology evaluation, ontology population and ontology evolution. As a result, numerous domain ontologies have been developed however, not much attention has been given to the area of ontology usage analysis that shows how the developed ontologies are being used. Such a study is very important as we explain in this paper by using some motivational scenarios of a Semantic Web user in different roles. The discussion is followed by a summary of the state-of-the-art of the existing literature on ontology usage to highlight gaps in this area. We define Ontology Usage Analysis and discuss our proposed Ontology Usage Analysis Framework (OUSAF) to measure the usage of an ontology on the web from different perspectives. On a real-world collected dataset, the results obtained from OUSAF demonstrate the practical utilization of OUSAF in measuring ontology usage analysis.

Contents

1. Introduction . . . 35
2. Motivational scenarios to show how ontology usage analysis will help in the ontology lifecycle . . . 36
   2.1. Different phases in the ontology lifecycle . . . 36
      2.1.1. Ontology engineering . . . 36
      2.1.2. Ontology evaluation . . . 37
      2.1.3. Ontology population . . . 37
      2.1.4. Ontology evolution . . . 38
   2.2. What is missing in the ontology lifecycle? . . . 38
3. Ontology usage analysis: state-of-the-art overview . . . 39
   3.1. Analysis of RDF data on the Web . . . 39
   3.2. Analysis of ontologies and vocabularies in RDF data usage . . . 40
   3.3. What is missing with respect to ontology usage analysis: critical evaluation . . . 40
4. Defining ontology usage analysis and positioning it in ontology engineering lifecycle . . . 40
   4.1. Instantiation . . . 41
   4.2. Usefulness . . . 41
   4.3. Usage . . . 41
   4.4. Commercial advantage . . . 41

* Corresponding author.
E-mail addresses: jam@jazbatech.com.au (J. Ashraf), e.chang@adfa.edu.au (E. Chang), o.hussain@adfa.edu.au (O.K. Hussain), farookh.hussain@uts.edu.au (F.K. Hussain).

http://dx.doi.org/10.1016/j.knosys.2015.02.026
0950-7051/

5. Ontology Usage Analysis Framework (OUSAF) . . . 41
5.1. Identification phase . . . 42
5.2. Investigation phase . . . 43
5.3. Representation phase . . . 43
5.4. Utilization phase . . . 43
6. Ontology usage analysis from OUSAF . . . 43
6.1. Scenario 1: Bob as an ontology developer . . . 43
6.1.1. Requirement 1 . . . 43
6.1.2. Requirement 2 . . . 43
6.1.3. Requirement 3 . . . 43
6.2. Scenario 2: Bob as a data consumer . . . 44
6.2.1. Requirement . . . 44
6.2.2. Requirement . . . 45
6.3. Scenario 3: Bob as a data publisher . . . 45
6.3.1. Requirement . . . 45
6.3.2. Requirement . . . 45
6.3.3. Requirement . . . 46
7. Conclusion . . . 46
References . . . 47

1. Introduction

In the recent past, the Internet has transformed the way we communicate, interact and do business across the globe. Described and dubbed as the *information highway*, the Internet provides an unprecedented seamless infrastructure to assimilate and disseminate information with an ease and speed never before witnessed by mankind. Today, as a result of this, 34.3% of the world's population is using the Internet.[1] Capitalizing on such intrinsic properties of the Internet as simplicity, ubiquity and scalability, Tim Berners-Lee introduced the World Wide Web [1] as a platform for publishing and consuming information on a universal scale. The World Wide Web (also known as the WWW or *Web*), which without a doubt is one of the most significant computational phenomena to date, has revolutionized information sharing by providing a decentralized information platform which has enabled and empowered users to be more interactive and participative, turning each user of the Web into a potential publisher. Being able to publish information which is accessible to anyone in the world with access to the Web for a low cost has resulted in the proliferation of approximately 14.3 trillion web documents[2] containing information on a variety of topics, and creating a huge amount of diversified information.

As a consequence, we are witnessing an inexorable rise in user-generated content that is padded with *metadata* to provide additional (syntactical and structural) information about the content, such as content ownership, provenance-related information, content categorization and labelling. This stage of the evolution of the Web, which is termed Web 2.0 [2], is described '*as a concept that takes the Internet as a platform for information sharing, interoperability, user-centred design, and collaboration on the Web*'. While having such meta-information is useful, its full potential can only be realized if such information from the Web is able to be retrieved for consumption, according to the required needs. To achieve this and take the Web closer to its original envisaged design,[3] the Semantic Web, whose vision seeks to transform the present *Web-of-Documents* into a *Web-of-Data* where the Web forms a global space for seamless knowledge integration is proposed [3]. This global space provides the mechanism to start describing tangible and non-tangible entities such as *people*, *software modules*, *projects*, *concepts*, and *documents* on the Web which are achieved by the core technology stack of the Semantic Web, Linked Data principles with other various components such as Syntax, Identifiers, Taxonomies, and Rules. While each component is important in its own right, in this paper, our focus is on ontologies that enable machines to interpret information in a common way and provide machine-processable meta-information enabling automatic information sourcing, retrieval and interlinking.

Ontologies form the main component of the Semantic Web vision as they provide the semantics for RDF data; that is, they transform data into meaning. In the literature, ontologies are defined as a '*formal specification of conceptualization*' [4]. Ontologies provide a shared and common understanding of the domain thereby enabling people and machines to communicate across heterogeneous distributed application systems. Thus, they specify a machine-readable vocabulary in computer systems, which is then used to infer and integrate knowledge, based on the semantics they describe. Such benefits of ontologies have been realized in various domains such as Healthcare and Life Science (HCLS) [5], governments,[4] social spaces [6,7], libraries [8], entertainment [9], financial services [10] and e-Commerce [11] by developing domain ontologies. As is the case with any information system or product, ontologies being the end product go through different stages of building, namely *engineering*, *evaluation*, *population* and *evolution*. However, one stage of the ontology-building process that has not received due attention is its *usage analysis*, that is, analyzing the exposure which the developed ontologies are receiving from end users. In other words, ontology usage analysis ascertains how the developed ontologies are being used; providing useful insights and information to the Ontology development process. In this paper, we explain the importance of Ontology Usage analysis in the different stages of an Ontology development and implementation before presenting an overview of our proposed Ontology USage Analysis Framework (OUSAF) to measure ontology usage. We briefly discuss the different phases of OUSAF before showing, by using the results from a real-world dataset, how OUSAF assists a Semantic Web user to determine the usage of an ontology from different perspectives.

This paper is organized as follows. Section 2 presents the need for and importance of including ontology usage analysis in the Ontology lifecycle, as well as a justification, by using several motivational scenarios. Section 3 presents the state-of-the-art overview of the existing literature from the viewpoint of ontology usage analysis and

[1] http://www.internetworldstats.com/stats.htm; accessed 16/6/2014.
[2] http://www.factshunt.com/2014/01/total-number-of-websites-size-of.html; accessed 16/06/2014.
[3] http://www.w3.org/History/1989/proposal.html; accessed 10/5/2014.
[4] http://oegov.org/ and http://dvcs.w3.org/hg/gld/raw-file/default/dcat/index.html; accessed 12/4/2014.