

# A Framework for Ontology Usage Analysis

Jamshaid Ashraf

School of Information Systems, Curtin University, Perth, Western Australia  
jamshaid.ashraf@gmail.com

## 1 Research Problem and Motivation

The Semantic Web (also known as the Web of Data) is growing rapidly and becoming a decentralized social and knowledge platform for publishing and sharing information. In the early days of the Semantic Web (1999-2006), research efforts of the community were centered around knowledge representation; thus, most of the research work was focused on building ontologies (ontology engineering), developing formal languages to represent them (ontology language), methodologies to evaluate and evolve ontologies (ontology evaluation and evolution (OE)), and logic for reasoning with them. As a result of this, even though ontologies were being developed but their instantiation was inadequate to provide the actual instance data needed for the evaluation and analysis of the developed ontologies. In order to overcome this issue, test data was often used to perform the above tasks [1]. However, in the recent past, the focus has shifted towards publishing data either with little or no use of ontologies [2]. This shift in focus is credited to the Linked Open Data (LOD) Project which has published billions of assertions on the Web using well known Linked Data principles. Because of this, the research focus has shifted from knowledge-centered to data-centered and is now settling down at the point where domain ontologies are being used to publish real-world data on the Web. This trend promotes consistent and coherent semantic interoperability between users, systems and applications. In this regard, several domain ontologies have been developed to describe the information pertaining to different domains such as Healthcare and Life Science (HCLS), governments, social spaces, libraries, entertainment, financial service and eCommerce.

According to the PingTheSemanticWeb.com which maintains a list of namespaces used in RDF documents, there are around 1810 namespaces (URIs) of ontologies/vocabularies currently being used on the Web (as of 12th Jan. 2012). However, while there are several ontologies being used, there is no formal approach to **evaluate, measure, and analyse the use of ontologies on the Web**. Such a study is very important to (a) *make effective and efficient use* of formalized knowledge (ontology) available on the web, (b) provide a *usage-based feedback loop* to the ontology maintenance process for a pragmatic conceptual model update, and (c) provide erudite *insight on the state of semantic structured data*, based on the prevalent knowledge patterns for the consuming applications. In the absence of such analysis, even though we may have some techniques to deal with information overload, but they may not provide us with efficient knowledge synthesizing techniques that are important to fully realize the potential of distributed knowledge space.

The presence of thousands of ontologies and billions of triples representing real-world data, now, provides the perfect foundation to carry out empirical studies to analyse the use of ontologies. Based on the aforementioned discussion, the aim of this PhD research is to develop a semantic framework for measuring and analyzing ontology usage known as an Ontology USage Analysis Framework (OUSAF). The proposed framework will be equipped with a set of metrics that measures qualitative and quantitative aspects of ontology usage, providing with important analysis on various detailed factors.

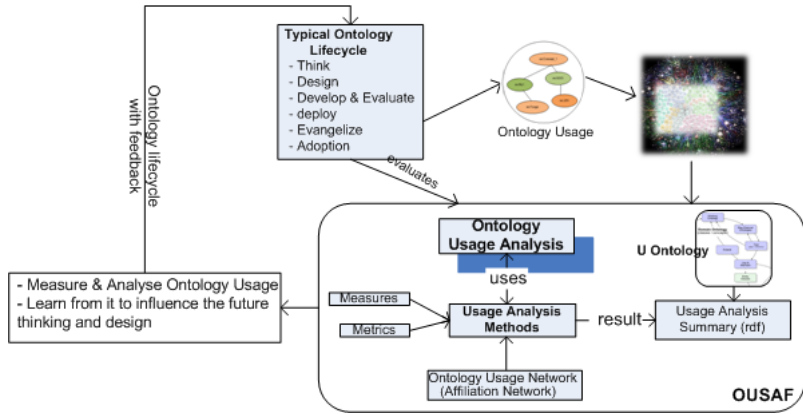
## 2 State of the Art

Related work in this area can be classified into two areas: first, work which focuses merely on knowledge (ontologies) and second, work which addresses RDF-data-related issues such as management, quality and usefulness. Pertinent to knowledge management and closely related to this research work is Ontology Evaluation (OE), which validates and verifies the developed ontology to measure the extent to which it serves and fits the intended purpose. Existing OE approaches focus mainly on evaluating the developed ontology and do not provide insight into how the ontology is utilised and adopted by its eventual users. As reported in [3], despite the fact that at present there are thousands of developed ontologies, very few of them are well populated and widely used in real-world applications. On the data-centered side, researchers have focused on evaluating the nature, quality and patterns of the RDF data published in response to the LOD project. For example in [4], while evaluating the quality, noise and inconsistency present in RDF data, the authors have provided guidelines for both publishers and consumers to assist in improving the quality and usefulness of data.

## 3 Methodology and Approach

The key step of the framework is to define the set of metrics to measure the usage of the ontologies and data and identify the emerging knowledge patterns. The metrics and measures that will be proposed in this research will evaluate the dataset from two aspects, the first of which is semanticity in which the presence and use of terminological knowledge in the dataset will be looked at by defining quantitative measures. The second aspect is the structurality whereby the knowledge base is conceived as an Ontology Usage Network, modeled using an affiliation network to understand and measure applicable social network properties such as centrality, degree and association (see Figure 1). The theoretical exploration and development of the framework is divided into the following sub-stages:

**Construction of RDF dataset:** In order to conduct an empirically grounded study, it is important to collect the real data that instantiate the ontologies on the Web. For increased effectiveness, the dataset is required to have two essential characteristics: 1) it should comprise real-world data collected from the Web, and 2) data should be collected during different time intervals to capture the changes and trends in knowledge patterns over time.



**Fig. 1.** Ontology Usage analysis Framework and Ontology Lifecycle with feedback loop

**Define metrics and measurements:** The framework supports empirical analysis from two perspectives: the ontology perspective and the RDF data perspective. From the ontology perspective, ontology is considered as an engineering artifact (ontology document) to characterize the components defined in a document such as vocabulary, hierarchal and non-hierarchal structure, axioms and attributes. From the RDF data perspective, we analyse the RDF triples to understand the patterns and the structure of the data available in the dataset. Metrics developed in OUSAF are grouped into three categories, namely, concept-centric metrics, relationship-(object property) metrics and attribute-(data property) metrics.

- *Concept-Centric Metrics:* In concept-centric metrics, the structure of each concept is considered in order to determine its importance within the ontology. Concept Richness, Concept Usage and Concept Population are potential metrics to measure the concept-centric metrics. These metrics help in measuring the concept instantiation and the information available with these instances.
- *Relationship and Attribute Centric Metrics:* Relationship-centric metrics aim to measure the relationship, attribute richness and usage of concepts in a knowledge base. To achieve this, metrics such as Relationship Value, Relationship Usage, Attribute Value and Attribute Usage are employed.

Furthermore, to increase the utilization of ontology usage, the analysis results will be conceptualized using Ontology Usage Ontology (U Ontology), as shown in Figure 1.

**Structural Properties of the Ontology Usage Network:** The aim of this step is to analyse the topological properties of the ontology usage network by extending the affiliation network model. We propose the construction of an on-

tology/vocabulary usage network as a bipartite network in which nodes are divided into two sets with links (considered as edges in our model) only between the nodes of different sets. A bipartite network provides the representation of our usage model and helps us to study general purpose network properties [5]. To topologically analyse the domain ontology usage in the web of data, several structural properties such as centrality (whether a network has a 'center' point or points), reciprocity (whether ties look identical from either end), density (how many potential ties in a network actually exist), and reachability (how many ties it takes to go from one 'end' of a network to the opposite 'end') will be studied to capture the topological aspect of ontology usage.

**Evaluation:** The proposed OUSAF framework will be evaluated by accessing Semantic Web data using the knowledge patterns generated by the usage analysis. Furthermore, the ontology usage summary will be used to auto-generate the prototypical queries to access the relevant information from the knowledge base and observe the precision and recall.

## 4 Initial Results and Proposed Benefits

To understand the nature of the structured (RDF) data published on the web, the domain ontologies used and their co-usability factor, the use of semantic web technologies and plausible reasoning (how much implicit knowledge is inferable), in [6] we considered the e-commerce dataset (called GRDS) by crawling the web. The latest version of the dataset comprises of 27 million triples collected from approximately 215 data sources. In [6], we performed empirical analysis on the dataset to analyse the data and knowledge patterns available and found that a small set of concepts (lite ontology) of the original model is, in fact, used by a large number of publishers. We also learnt that, with current instance data, there is not much for RDFS reasoner(s) to infer implicit knowledge due to the invariant data and knowledge patterns in the knowledge base. Based on the visibility obtained, in [7] we have proposed a framework and metrics to measure the concepts and property usage in the dataset by keeping in view their richness within the ontological model. In [8], the ontology usage framework is used to extract the web schema, based on the ontology instantiation and co-usability in the database. The web schema represents the prevailing schema, providing the structure of data useful for accessing information from the knowledge base and building data-driven applications. Based on the research done so far and the initial results obtained, we are confident of the benefits that can be realized with the implementation of OUSAF, which include: (a) assisting in building Semantic Web applications to offer rich data services by exploiting the available schema level information and assisting in providing an improved context driven user interface and exploratory search [9]. based on auto discovery of explicit and implicit knowledge; (b) enabling client applications to make expressive queries to the Web by exploiting the schema patterns evolving through the use of ontologies; and (c) empirical analysis of domain ontology usage, as shown in Figure 1,

that provides the feedback loop to the ontology development life cycle. Knowing the sub-model of the original ontology, which provides information about usage and adoption, will assist the ontology developer to pragmatically refine, update and evolve the conceptual model of ontology.

## 5 Conclusion and Future Work

The objective of this research is to design and implement a semantic framework to evaluate and analyse the usage of domain ontology. From a high level view, given a domain ontology and dataset, we would like to analyse the usage of ontology qualitatively and quantitatively, its co-usability factor with other ontologies and plausible reasoning. In future work, I will construct the dichotomized one mode ontology-by-ontology co-usability matrix to have the ontology linkage graph. Furthermore, the usage patterns represented through U Ontology will be further used to develop the access layer for the web-of-data (see Figure 1).

## References

1. Tao, J., Ding, L., McGuinness, D.L.: Instance data evaluation for semantic web-based knowledge management systems. In: HICSS, pp. 1–10. IEEE Computer Society (2009)
2. Jain, P., Hitzler, P., Yeh, P.Z., Verma, K., Sheth, A.P.: Linked data is merely more data. In: AAAI Spring Symposium: Linked Data Meets Artificial Intelligence. AAAI (2010)
3. Ding, L., Zhou, L., Finin, T., Joshi, A.: How the semantic web is being used: An analysis of foaf documents. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences - Track 4, vol. 4, pp. 113–120. IEEE Computer Society, Washington, DC (2005)
4. Hogan, A., Harth, A., Passant, A., Decker, S., Polleres, A.: Weaving the pedantic web. In: Linked Data on the Web Workshop (LDOW 2010) at WWW 2010 (2010)
5. Borgatti, S.P., Everett, M.G.: Network analysis of 2-mode data. *Social Networks* 19(3), 243–269 (1997)
6. Ashraf, J., Cyganiak, R., Oriain, S., Hadzic, M.: Open ebusiness ontology usage: Investigating community implementation of goodrelations. In: Linked Data on the Web Workshop (LDOW 2011) at WWW 2011, Hyderabad, India, March 29 (2011)
7. Ashraf, J., Hadzic, M.: Domain ontology usage analysis framework. In: SKG, pp. 75–82. IEEE (2011)
8. Ashraf, J., Hadzic, M.: Web schema construction based on web ontology usage analysis. In: JIST. Springer (2011)
9. Tvarožek, M.: Exploratory search in the adaptive social semantic web. *Information Sciences and Technologies Bulletin of the ACM Slovakia* 3(1), 42–51 (2011)