Domain Ontology Usage Analysis Framework

Jamshaid Ashraf [1],Maja Hadzic [2]

Digital Ecosystem and Business Intelligence (DEBII)
Curtin University, Perth, Western Australia
[1]jamshaid.ashraf@gmail.com
[2]m.hadzic@curtin.edu.au

***Abstract*— The Semantic Web (also known as Web of Data) is growing fast and becoming a decentralized knowledge platform for publishing and sharing information. The web ontologies pro- mote the establishment of a shared understanding between data providers and data consumers, allowing for automated information processing and effective and efficient information retrieval. The majority of existing research efforts is focused around ontology engineering, ontology evaluation and ontology evolution. This work goes a step further and evaluates the ontology usage. In this paper, we present an Ontology USage Analysis Framework (OUSAF) and a set of metrics used to measure the ontology usage. The implementation of the proposed framework is illustrated using the example of GoodRelations ontology (GRO). GRO has been well adopted by the semantic e-commerce community, and the OUSAF approach has been used to analyse GRO usage in the dataset comprised of RDF data collected from the web.**

I. INTRODUCTION

The Semantic Web (also known as Web of Data) is growing fast and becoming a decentralized knowledge platform for publishing and sharing information [1]. Machines are the main actors in the semantic web, interacting with the information that is represented in machine process- able format enabling auto- matic information discovery. The semantics are added to the web of data through the use of ontologies, thereby allowing machines to interpret the domain knowledge formally conceptualized by web ontologies. The semantic web continues to expand with the vast amount of RDF data that is semantically annotated using vocabularies and ontologies available on the web [2].

Numerous ontologies are being developed and used to publish information on the web. Swoogle[1], a semantic search engines, has an index of 10,000 ontologies and Ping The Semantic Web[2] Web2 has listed around 1442 known namespaces used in the documents available in their repository. These two sources give an indication of the number of ontologies available, but does not provide any detail on how those ontologies are being used or the co-usability of different ontologies that exist in the semantic web. Despite having several ontologies, only a few of them are well populated [3].

Ontologies, often, are developed using an ontology development process based on a certain methodology [4] and thus the end product is the ontology document comprising of concepts, relationships, attributes and axioms. Ontology, being an engineering artifact can be viewed as a product developed based on certain requirements (i.e. modelling domain of interest). On- tologies are evaluated before they are used or reused, and are usually discussed in literature within the research area of Ontology Evaluation [5], [6] Ontology evaluation techniques are useful during the ontology engineering process [7], They are also very helpful to end users in evaluating and deciding which ontology best meets their requirements. Ontology evaluation is used for verifying its correctness against requirements, validating its conceptualized model of the real-world and assessing it from the end users perspective [7]. Ontology Evaluation and different overlapping areas such as ontology evolution [8] and ontology change management [9], investigate the ontology while it is being developed or after it has been developed, but do not consider how the ontology is being used by the end users.

To the best of our knowledge, we have not encountered any particular discipline which discusses a systematic approach to-analysing the usage and adoption of a particular ontology on the web in real-world settings. Therefore, there is a need to develop a systematic approach for evaluating and analysing the particular ontology usage, and its adoption and uptake by different users on the semantic web. In other words, there is a need to provide an insight into the prevailing structure of ontology and understand the patterns available, and understand the actual use while also considering the intended use.

As the first study along this line, in this paper we present a framework for conducting a domain ontology usage analysis. To make the analysis reflect a real-world setting, we build the dataset comprising of RDF data collected from more than a hundred different websites using domain ontology as the common denominator. We also propose a set of metrics to measure the ontology usage both qualitatively and quantitatively. This allows us to understand the depth and breadth of particular domain ontology adoptions and the structured data patterns available in the web of data. For the evaluation of the framework, GoodRelations [10] ontology is considered since it conceptualises the web e-commerce domain.

The rest of the paper is organized as follows. In Section 2, we briefly outline the motivation for ontology analysis and its usefulness for different stakeholders. Then in Section 3, we describe the ontology usage analysis framework and discuss its components. In Section 4, after describing common terminologies used in this paper, we introduce the new metrics and measures used for the usage analysis. Section 5

[1]http://swoogle.umbc.edu
[2]http://www.pingthesemanticweb.com

provides implementation details including dataset collection and experiment results. Section 6 presents a review of the related literature relevant to ontology usage analysis. Section 7 concludes with a discussion and possible future work.

II. MOTIVATION

The burgeoning of RDF data and the adoption of ontologies have produced the need to evaluate and understand the current adoption and implementation of web of data. Obtaining a pragmatic view of the current implementation, and analysing the use of ontologies, will help to provide the feedback loop to all the stakeholders of the semantic web community. This includes data publishers, data consumers and ontology developers.

Data publishers need to know about the structured data usage, patterns and approaches in order to improve the quality, quantity and usefulness of the data. The vocabularies and ontologies being utilized provide the formal structure and schema so that information can be arranged in a consistent and shared manner to augment the non-structured content on the web to a fully or semi-structured content space. The reuse of commonly used vocabularies, based on some imperative usage analysis, will allow a small unified schema to persist in a large number of web resources, which is useful in many ways. RDF triple stores, Reasoners and SPARQL endpoints can implement and provide built-in support for these well used sets of vocabularies (unified schema) to offer efficient services such as interlinking with related entities, a materialized view of implied (inferred) knowledge, and ontology-based indexing.

Data consumers and semantic web client applications need to know the popular and populated data structures in order to access information efficiently. This helps in building data knowledge-driven applications [11] based on the ontology model used to describe data.

Ontology developers need to understand the sub-model of the original ontology model which has been more widely adopted and used, and refine the existing ontology model accordingly [12]. The well used sub-model with more instantiation can be used during the ontology engineering process to validate and verify the new version of domain ontology. The aforementioned points are the preliminary arguments for the usefulness of ontology usage analysis. As the research community progresses in this area, we will experience serendipitous discovery of its usefulness.

III. FRAMEWORK

In our previous work [13] we looked at one of the most widely used domain ontologies, GoodRelations, and analysed its usage on the web. After realizing the significance of vocabulary/ontology usage analysis, we have proposed a framework (see Figure 1) to conduct similar ontology usage analysis for any domain ontology and investigate it adoption, implementation and uptake by the end users.

The main role of the Ontology Usage Analysis Framework (OUSAF) is to support empirical analysis of RDF data on the web with focus on domain vocabularies and ontologies. The framework supports empirical analysis from two perspectives: one from the ontology perspective and the other from the RDF data perspective. From the ontology perspective, we consider the ontology 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 so as to understand the patterns and the structure of the data available in the dataset.

The analysis of ontology usage on the web is different from assessing the quality of ontology and ontology evaluation. In the following paragraph, we informally describe Ontology Usage Analysis (OUA) and discuss those aspects in which it differs from its adjacent areas like ontology evaluation, ontology maintenance and ontology evolution.

OUA is different in many ways from ontology evaluation [14]in spite of having overlap. To understand the difference, let us first discuss the informal definition of OUA and then compare it with the definition of ontology evaluation in the context of an ontology development framework. The OUA analyses the use of ontology on the web in a real-world setting by measuring its usage, its usefulness and commercial advantages. Even though no formal definition of ontology evaluation is available in the literature, it is commonly referred to as a set of tools and methods to compare, validate and rank similar ontologies [15], [16]. Also, ontology evaluation is often used within the context of a single ontology. For example, to evaluate a newly designed ontology for its structure, content, coverage, etc. Ontology evaluation and other ontology quality approaches are important; however, their emphasis is more on guaranteeing that what is built will meet the requirements (ontology developers view) and that the final product (ontology artifact) will be as error free as possible. Therefore, in some ontology engineering methodologies, ontology evaluation is a built-in process while in others it is considered as an independent component [14]. On the other hand, OUA focuses on the post implementation assessment scenario where actual utilization of a particular ontology in the semantic web context is observed and its adoption, co-use and reuse is analysed. OUA focuses on the instantiated structured data on the web-based on domain ontology. For this reason, OUA can be viewed as a separate and independent activity from ontology development. OUA can even be considered as a post-implementation process and a part of ontology maintenance or ontology evolution.

Now we will consider the overlap between OUA and ontology evolution. The emphasis of OUA is to understand and measure the ontology (vocabulary) usage in terms of its population, semantic relationship between different concepts, conformance with linked data principles and possible inferencing depending on the axioms of ontology. Ontology evolution, which is closely related with ontology change and versioning, covers the change management process to keep the ontology artifact up-to-date and increase its effectiveness and usefulness. Ontology evolution is defined in [17] as the ‘timely adoption of an ontology to the arisen changes and the consistent management of these changes’.The sources of change

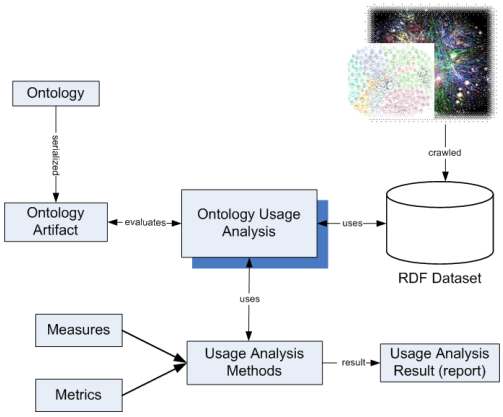

Fig. 1. Ontology Usage Analysis Framework

that trigger ontology evolution are explicit requirements or the result of some automatic change discovery method. Nevertheless, the current approaches [18] ignore an important source of information: the actual utilization of ontology on the web. Actual utilization needs to be analysed using metrics and measurements to qualitatively and quantitatively describe usage. This is the main objective of this paper.

Figure 1, provides the schematic diagram of the OUA framework known as OUSAF. In its simplest from, OUSAF receives domain ontology as input and analyses the ontology use within the dataset comprised of RDF data collected from the web.

In the following, we briefly introduce the model and terminology used throughout this paper. The notation and terminologies used within this paper are already familiar to the semantic web community [19]. The model of Ontology and Knowledge base used in the metrics and measurements of domain vocabulary usage is primarily based on [20].

RDF Term. Given the set of URI references U, the set of blank nodes B, and the set of literals L, the set of RDF terms is denoted by $RDFTerm{:=}U{\cup}B{\cup}L$. The sets *U*, *B* and *L* are pair-wise disjoint.

RDF Triple. A triple $t := (s, p, o){\in}(U \cup B)\ X\ U\ X\ (U \cup B \cup L)$ is called an RDF triple, where *s* is called subject, *p* predicate, and *o* object.

Ontology Structure. An ontology structure is a 6-tuple $O := \{C, P, A, H_C, prop, att\}$consisting of two disjoint sets C and P whose elements are called concepts and relation identifiers, respectively, a concept hierarchy H_c: H_c is a directed, transitive relation $H_c \subseteq CXC$ which is also called concept taxonomy. $H_c(C_1, C_2)$ means that $C_1$is a sub-concept of C_2, a function *prop*: $P \rightarrow C\ X\ C$, that relates concepts non-taxonomically (The function *dom:*$P \rightarrow C$ with $dom(P) := \Pi_1(rel(P))$ gives the domain of *P*, and *range*:$P \rightarrow C$ with *range(P)*:$= \Pi_2(rel(P))$ gives its range. For *prop(P)* = (C_1, C_2) one may also write $P(C_1, C_2)$. A specific kind of relations are attributes A. The function $att : A \rightarrow C$ relates concepts with literal values (this means *range(A):=STRING*)

Dataset (ontology based metadata). A metadata structure is a 6-tupel *Dataset* $:= \{O, I, L, inst, instr, instl\}$, that consists of an ontology O, a set I whose elements are called instance identifiers (correspondingly C, P and I are disjoint), a set of literal values L, a function $C \rightarrow 2^I$ called concept instantiation (For $inst(C){=}\ I$ one may also write $C(I)$), and a function $inst : P \rightarrow 2^{IXI}$ called relation instantiation (For $instr(P) = I_1, I_2$ one may also write $P(I_1, I_2)$). The attribute instantiation is described via the function $instl : P \rightarrow 2^{IXL}$ relates instances with literal values.

IV. METRICS

In ontology evaluation, different approaches are used to evaluate ontology such as gold standard [21], application based [15], data driven [22] and evaluation done by humans [23]. In application- and data-driven approaches, different metrics and measurement are proposed by several researchers [15], [4]. However, those metrics are not applicable in our case because they measure the quality of stand-alone ontology or compare it with other ontologies to rank them based on their concept coverage. In this study, we consider the dataset comprised of semantic data collected from web of data and measure instantiation and the relationship of a conceptualized domain modelled by ontology. Metrics used in OUSAF are grouped into three categories: concepts metrics, relationship (object property) metrics and attribute (data property) metrics.

A. Concept Metrics

In concept metrics, we first look at the structure of each concept to determine its importance within the ontology. Then we measure its instantiation and the information available with these instances. Concept Richness, Concept Usage and Concept population are frequently used within the context of Concept Metrics.

Concept Richness (CR)

When considering a specific concept in ontology, one need to consider the relationship it has with other concepts and the number of attributes available to describe the instances. This includes the typed binary relationship (non-hierarchical) with other concept and data properties providing attribute values for data description of concept. Formally, the concept richness of a particular concept $CR(C)$ of a given domain ontology is calculated by adding the number of non hierarchical relationships and attributes.

$$CR(C) = |P_C| + |A_C|$$

$CR(C)$of a concept reflects its possible contribution toward providing formal structure to represent the specific view of the domain, conceptualized by the concept. In other words the higher the number of concept richness the richer is the concept in terms of its description. P_C return the number of

object properties of C while A_C returns the number of data properties of C. The value of $CR(C)$ is a positive integer number including zero. This helps us to rank the concepts based on their richness values and analyse the correspondence, if any, between the richness value and the usage and population of the concept in semantically annotated web of data.

Concept Usage (CU)

Concept usage measures the instantiation of the concept in the knowledge base (KB). Here instantiation means the number of unique URI references used to create members of the class represented by concept. In RDF graph, we are referring to the triples in which *rdf:type* predicate is used to create members of a given concept. The concept usage $CU(C)$ is formalized as follows:

$$CU(C) = |t = (s, p, o)|p = rdf : type, o = C|$$

$CU(C)$returns an integer number (zero possible) and helps in measuring the usage of each concept in KB and rank them based on their instantiation.

Concept Population (CP)

Concept population (CP) calculates all the triples in the KB where concept's instances (URI references) is used to either create relationship with other concepts or provide data description using attributes. CP is different from CU because in CU we consider only the unique instances of type Concept and not the use of those instances in providing information description about resources. In the RDF graph, we consider all the triples that have an instance identifier either as a subject or object of the tripe.

$$CP(C) = |t = (s, p, o)|s = C(I),\ o = C(I)\ or\ L|$$

$CP(C)$measurement returns an integer number (zero possible) reflecting the semantic representation and coverage of the concept. This helps in knowing the prevalent structure available in KB assisting in information retrieval. Concept instance identifier is used to either create a relationship with other concepts or attributes are used to describe entity.

B. Relationship Metrics

In the following, we discuss the metrics defined to measure the relationship and attribute richness and usage in a knowledgebase. We define concepts such as Relationship Value (RV), Relationship Usage (RU), Attribute Value (AV), and Attribute Usage (AU).

Relationship Value (RV)

Relationship value reflects the possible role of the object property in creating typed relationships between different concepts. Object property links the instances of the concepts defined as the domain of this property with the instances of the concepts defined as range of the property.

$$RV(P) = |dom(P)| + |range(P)|$$

$RV(P)$returns an integer number, reflecting the number of concepts in which property can be used to create relationships and provide rich description of concept. A property with higher *RV* reflects its generalization as more concepts can be linked through this property. On the other hand a lower *RV* value conveys property specificity.

Relationship Usage (RU)

Relationship usage calculates the number of triples in a dataset in which object property is used to create the relationship between different concept's instances.

$$RU(P) = |t := (s, p, o)|p = P|$$

The result of *RU* is a positive integer number (zero possible). *RU* is helpful in indexing the properties in combination with *RV* to support efficient information retrieval. It is also helpful in developing knowledge base applications where relevant data is automatically retrieved and presented based on the available data space.

Attribute Value (AV)

Attributes of a concept are the data properties used to provide literal (typed or un-typed) values to the concept instances. AV reflects the number of concepts that have this data property.

$$AV(A) = |dom(A)|$$

Attribute Usage (AU)

Attribute usage measures how much data description is available in KB for a concept instance.

$$AU(A) = |t := (s, p, o)|p \in A, o \in L)|$$

C. Knowledge Base Metrics

In the following we define metrics to measure the ontology population in a dataset in order to analyse the use of domain ontology.

Domain Ontology Population (DOP)

Ontology population measures the amount of structured data available in KB that is annotated using ontology RDF Terms. This includes the concept instantiation and the instance references used for describing resources. The description of resources includes the relationship with other concepts and the attributes values.

$$DOP = C_i(I) + \sum_{i=1}^{|C|} C_i(I) + \sum_{j=1}^{|P|} P_j(I_m, I_n) + \sum_{k=1}^{|A|} A_k(I_m, I_n)$$

Here C is the set of concepts, P is the set of object properties and A is the set of data properties as defined in domain ontology respectively

Domain Ontology Usage (DOU)

Domain Ontology Usage (DOU)measures the use of ontology vocabulary in the dataset. DOU is measured by dividing the DOP by the dataset size (KBS). The size of the dataset is the total number of triples stores in the knowledge base

$$DOU = \frac{DOP}{KBS}$$

The result of DOU is the percentage indicating the coverage of the domain ontology vocabulary in the dataset. A high value of DOU tells us that domain ontology usage is dominant over the knowledge base usage and has better semantic coverage.

V. IMPLEMENTATION

In the following we present the implementation of OUSAF by employing the aforementioned metrics and measures. In order to conduct a credible study, the evaluation of the framework has been carried out a web ontology that enjoys a reasonable adoption and is being used in a real world setting. GoodRelations (GR) [10] is selected as the domain ontology for our study. Its selection is based on its popularity[3] and usage in real world e-commerce websites. The dataset used for the analysis comprises 105 different data sources which have used GR to annotate information on the web.

A. GoodRelations Ontology

The latest version[4] of GR ontology comprises 27 concepts (classes), 49 object properties, 43 data properties and 43 named individuals. From a high level view, the GR model is based on three main concepts, each focusing on a separate aspect of e-commerce domain. These three main concepts are Business Entity, Offering and Product or Service and each one is discussed in details in the following sections. GR ontology is available at http://purl.org/goodrelations/v1 and **gr** is the prefix used in this paper and also in general practice to access vocabulary defined by the GR.

Business Entity: gr:BusinessEntity concept represent a business organization (or any individual) which intends to offer or seek products on the web. The main purpose of this concept is to provide the attributes necessary to describe any business such as name of the company, address, its particular location, vertical industry in which it operates and any other identifier which makes it uniquely identifiable on the web.

Offering: Offering is the pivotal concept in GR ontology. This concept allows the business entity to describe a particular offer that it would like to make or seek on the web. Offering can include one or more products with a price specification describable in any possible currency. If the offering requires warranty promises eligible customers of the offer, shipment and shipment charges, acceptable methods of payment, such supplementary details can be attached with the offer description..

Product or Service: The third main concept is Product or Service (gr:ProductOrService). As mentioned earlier, an offer can include one or more products (or services) and is usually described using one of the three possible subclasses of this main (abstract) class. GRs main focus is to cover the conceptual model of offering rather than being a product ontology. However, the gr:ProductOrService concept can be used to describe any product and has different data properties allowing the user to describe lightweight product ontology.

B. Dataset and Data Collection

To analyse the usage, usage patterns and uptake of GR ontology in general and by the e-commerce community in particular, we collected data from different data sources to generate GoodRelations Dataset (GRDS). We first identified the potential data sources. A minimum criterion was that the GR ontology is used to describe the offering or company (Business Entity) or both.

Different semantic search engines such as Sindice[5] and Watson[6] which index RDF documents, have been used to obtain a list of potential data sources. We looked at the URIs of the graphs returned by search engines to extract the URL of websites, containing semantic web data represented in RDF data model. Almost all of the sources have sitemap.xml files available to allow search engines crawlers to access web pages and build their indexes for regular searches. Since we were interested in accessing the web pages which have RDFa, we firstly build list of URLs which has RDFa snippet and then used REST based web services (http://any23.org/ and http://www.w3.org/2007/08/pyRdfa/) to parse RDFa snippets from HTML documents and generate RDF graphs. We loaded these RDF graphs to a triple store to perform our investigation on GR marked dataset. From RDF data management point of view, named graphs are used to group all the triples of one data source under a unique named graph URI which allowed us to query the dataset vertically as well as horizontally.

C. Analysis

In order to understand the domain ontology usage and prevalent schema available in the dataset, metrics discussed in Section 4 are used by OUSAF to analyse the usage. Firstly, we look at the concept distribution in the knowledge base and rank them based on their population. For some measures, SPARQL queries were sufficient to obtain results. Nevertheless, for other complex number processing, Java based Sesame API[7] was used to the perform desired analysis.

[3]PingTheSemanticWeb.com has ranked GoodRelations second to FOAF as the most widely used ontology.

[4]The latest version was updated on Nov 26, 2010 and is the model used & referred in this paper and work.

[5]http://www.sindice.com

[6]http://watson.kmi.open.ac.uk/WatsonWUI/

[7]http://www.openrdf.org/

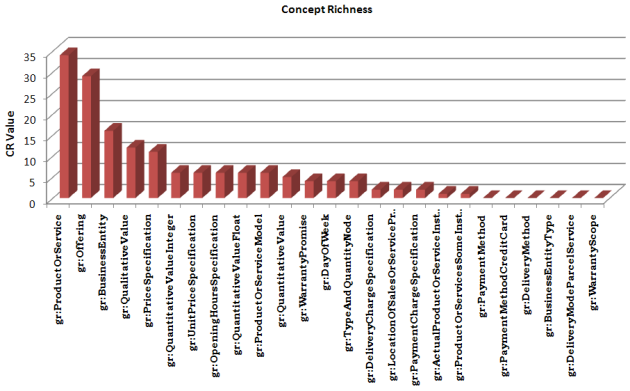

Fig. 2. Concept Richness in GoodRelations Ontology

1) Concept Usage Analysis: By using the CR, CU and CP metrics and GRO as domain ontology, the following usage statistics were obtained from the dataset.

Figure 2 displays the GR concepts in descending order based on their richness value. Few concepts have a high value which means that these concepts have several object and data properties available in the conceptual model, providing rich set of properties for semantic annotation. There are also concepts having only one property, providing minimum structure information about the concept.

There is a small caveat in the figure that needs some explanation. There are axiomatic triples available in the GR ontology, particularly sub-sumption axioms which allow the inference of new knowledge from a knowledge base that is not explicitly stated. According to [19] , subclass inherits all the properties from its superclass. Therefore, under the RDF Semantics, one can also include (add) the richness value of superclass (super concept) to the value of its subclass. For brevity, we did not consider the subsumption behaviour and have restricted this study to the dataset repository without turning on the inference engine feature.

Next we will look at the usage of each concept in the dataset. This provides an estimate about the number of typed entities available in the knowledge base. To understand the relevance, if there is any, between the concept richness and concept usage, we normalized the quantitative measures of *CR* and *CU*.

Figure 3displays both *CR* and *CU* for each concept to help us spot the relevance between these two measures. After careful analysis of Figure 3, one can extract the following findings:

- A small part of the ontology is widely used. It is quite evident from the figure that only a small part of the ontology is being used by the end users and the majority of the concept are rarely or never instantiated.
- Concepts with higher richness value also have large instantiation. The concepts with higher CR value have better instantiation compared to those concepts with lower CR value. This finding cannot be generalized because the dataset does not comprise RDF data randomly extracted from the web; rather, a selection criterion was used to populate the knowledge base.
- Generalized concepts have fewer instantiations compared with their specialized concept. We note that the specialized concepts have more entities defined than their super classes. This may become less relevant in terms of information retrieval if the underlying knowledge base provides an RDFS-based reasoning support. For example, rdfs9 rule of RDFS entailment rules [17] says **IF** (*uuu rdfs:subClassOf xxx* **AND** *vvv rdf:type uuu*) **THEN** (*vvv rdf:type xxx*). When RDFS reasoning is available, any query directed to retrieve the instances of the superclass will also include the instances instantiated by its subclasses.

To understand the overall distribution of data and the conceptual coverage of the model, we also investigated the population of each concept by querying the dataset and measuring the use of concept instances in describing entities.

Figure 4 shows the concepts' distribution in the dataset and their relevance to the other ontology concepts. This figure provides an overview of the ontology usage and the trend and patterns available in the knowledge base. In an attempt to draw a comparison between different measurements, we normalized the value based on their maximum value found in the result set. However, this has biased the concept population values which have very few triples using the concepts instance identifier. For example the *CP* value of *gr:PaymentMethodCreditCard* before normalization is 90 which means there are 90 triples in the dataset where the instance identifier of this concept is used. But after normal- ization, due to higher standard deviation, the small values are close to zero. Therefore, this must be taken into consideration when interpreting the chart depicted in Figure 4.

The fundamental rationale is that the higher the number of properties of a concept, the higher is the possibility of this concept being used in the semantic annotation. A concept with a large number of properties provides a wider choice to the end users for describing entities. To a certain extent, this rationale holds true in our study. The concepts with a high *CR* value as well as high *CU* and *CP* values appear on the left in Figure 4 .

In the preceding sections, we have focused on the concept (classes) and investigated their usage focusing on the relevant domain conceptualized by domain ontology.

We have proposed similar metrics for measuring relationship and attribute usage however, these result are not presented in this paper and will be part of our future work.

VI. RELATED WORK

Metrics and measurements for evaluating web resources have been used from the very early days of the web [21]. Different metrics have been proposed for ontology evolution and ontology evaluation, and include quality assessment of the ontology. Several researchers have proposed different

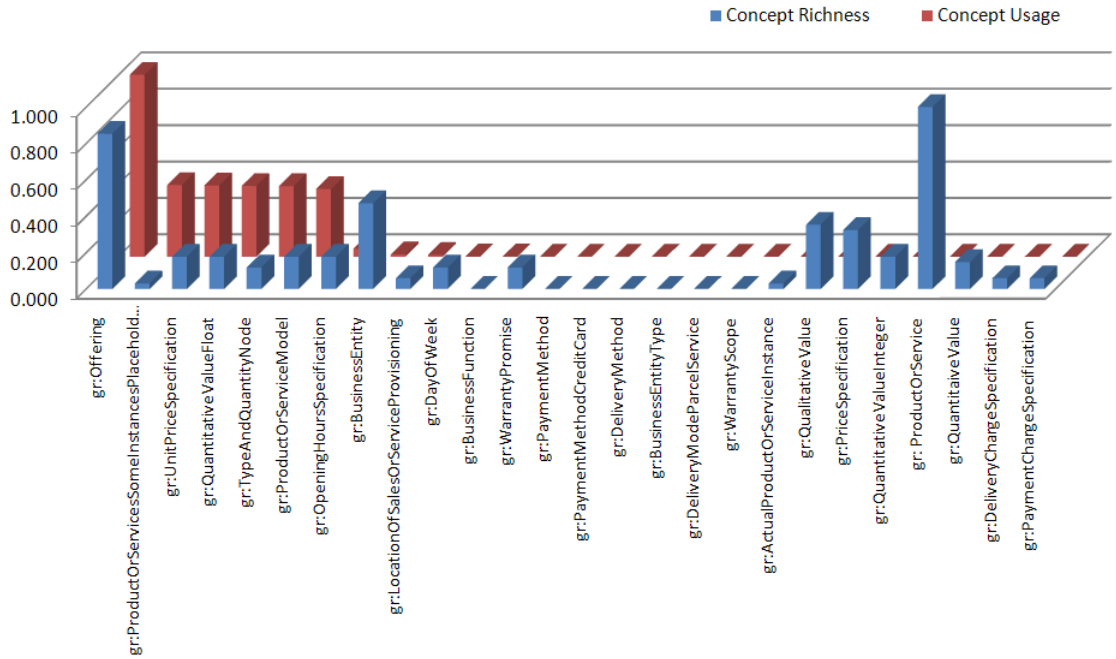

Fig. 3. Concept Richness in GoodRelations Ontology

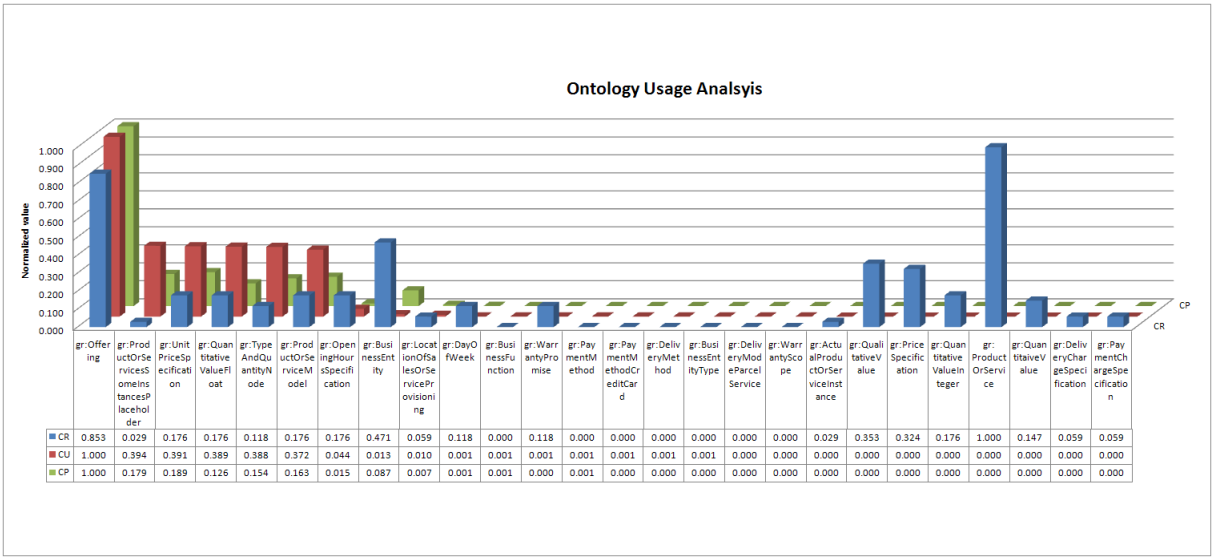

Fig. 4. Ontology Usage Analysis

measures to evaluate ontology, few of which have the same name; however their purpose and interpretations are different. For example, in [15] the authors have used a metric called Class Richness (which can be interpreted as Concept Richness) but the purpose of this metric is quite different from the one we explained in this paper.

In the following, we briefly discuss the related work done on the analysis of semantic web data on the web with or without considering the ontology model.

In [3], authors have analysed the social and structural relationship that exist in the FOAF documents currently published on the semantic web. Their study was focused on understanding the social patterns available and analysing their role in social structures. They also presented details about the other namespaces used in the dataset which can be useful in exploring the use of other vocabularies. This work was concerned with evaluating the FOAF documents and related social structures, but did not provide any framework to conduct a similar analysis on datasets from different domains such as e-commerce, health etc.

In [4], three types of evaluation approaches are introduced to assess the ontology from functional, usability-based, and

structural point of view. The usability-based approach is a data-driven approach which is used to evaluate the ontology by measuring how 'fit' the ontology is in providing the conceptual representation for the entities available in the data (corpus).

In [15], authors present a framework and a OntoQA tool that implements a number of metrics such as richness, connectivity, fullness and cohesion to evaluate an ontology. The metrics used in OntoQA are interesting, although their actual usefulness in not well known. The empirical analysis was carried out on a very small dataset which by no means reflects the actual instantiation. Additionally, the main focuses of the study are ontologies which normally account for 1% of the RDF data on the web.

In summary, the abovementioned work analyses the structural, semantic and taxonomical aspects of ontology while the associated metrics is designed to accommodate ontology-centric parameters. Little or no work has been done on evaluating actual datasets using those ontologies.

VII. CONCLUSION AND FUTURE WORK

In this paper, we described and illustrated the Ontology USage Analysis Framework (OUSAF) and the set of metrics used to measure and understand the actual usage of ontological model in the semantic web. The empirical study was conducted on a dataset comprising of the semantic e-commerce data currently available on the web. GR ontology usage was evaluated within this dataset with the help of OUSAF. We noticed that a very small part of the ontology is being used on the web and several concepts are not being used at all. Also, we noticed that there exists a relationship and correspondence between different measures, such as between concept richness and concept usage and population.

In this paper, we have presented only the concept centric metrics and we plan to include relationship (object properties) and data properties in the framework evaluation as a part of our future work. Additionally, we hope to conduct this research on a larger scale and include a larger dataset in our study. We also intend to automate the process to a certain extent and assist the end users to conduct ontology usage analysis. We also hope to ascertain what it is that makes a segment of the ontology well-used compared with the other less-utiliszed part of the ontology.

REFERENCES

[1] P. Mika, "Ontologies are us: A unified model of social networks and semantics," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no. 1, pp. 5–15, 2007.

[2] R. C. Chris Bizer, Anja Jentzsch, "State of the LOD cloud (http://www4.wiwiss.fu-berlin.de/lodcloud/state/)," 2011.

[3] L. Ding, L. Zhou, T. Finin, and A. Joshi, "How the semantic web is being used: An analysis of foaf documents," in *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4 - Volume 04*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 113.3–. [Online]. Available: http://portal.acm.org/citation.cfm?id=1042435.1042928

[4] Y. Sure, S. Staab, and R. Studer, "On-to-knowledge methodology (otkm)," in *Handbook on Ontologies: International Handbook on Information Systems*, S. Staab and R. Studer, Eds. Springer, 2004, pp. 117–132.

[5] A. Gmez-Prez, "Evaluation of ontologies." *Int. J. Intell. Syst.*, vol. 16, no. 3, pp. 391–409, 2001.

[6] A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann, "A theoretical framework for ontology evaluation and validation," in *Proceedings of SWAP 2005, the 2nd Italian Semantic Web Workshop, Trento, Italy, December 14-16, 2005*, *CEUR Workshop Proceedings*, 2005.

[7] M. Sabou, V. Lopez, E. Motta, and V. Uren, "Ontology selection: Ontology evaluation on the real semantic web," in *Proceedings of the Evaluation of Ontologies on the Web Workshop, held in conjunction with WWW'2006*, 2006.

[8] P. Plessers, O. D. Troyer, and S. Casteleyn, "Understanding ontology evolution: A change detection approach," *Journal of Web Semantics*, vol. 5, no. 1, pp. 39–49, 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1229198

[9] M. Klein, "Change management for distributed ontologies," Ph.D. dissertation, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, 2004.

[10] M. Hepp, "Goodrelations: An ontology for describing products and services offers on the web," in *Knowledge Engineering: Practice and Patterns*, ser. Lecture Notes in Computer Science, A. Gangemi and J. Euzenat, Eds. Springer Berlin / Heidelberg, 2008, vol. 5268, pp. 329–346.

[11] N. Guarino, "Formal ontology and information systems." IOS Press, 1998, pp. 3–15.

[12] P. Haase, J. Völker, and Y. Sure, "Management of dynamic knowledge," *Journal of Knowledge Management*, vol. 9, no. 5, pp. 97–107, Oct. 2005.

[13] J. Ashraf, R. Cyganiak, S. O'Riain, and M. Hadzic., "Open ebusiness ontology usage: Investigating community implementation of goodrelations," in *Linked Data on the Web Workshop (LDOW2011) at WWW'2011*, I. the Proceedings of the Linked Data on the Web WWW2011 Workshop (LDOW 2011), Ed., Hyderabad, India, 29 March 2011.

[14] J. Brank, M. Grobelnik, and D. Mladenić, "A survey of ontology evaluation techniques," in *Proc. of 8th Int. multi-conf. Information Society*, 2005, pp. 166–169.

[15] S. Tartir, I. B. Arpinar, M. Moore, A. P. Sheth, and B. Aleman-Meza, "Ontoqa: Metric-based ontology quality analysis," in *Proceedings of IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, 2005.

[16] H. F. Ferula Patrick, "Ontology evaluation , htt://www.sti-innsbruck.at/fileadmin/documents/applied-onto-eng201011/ontology-evaluation-methods.pdf," 2010.

[17] P. Haase and L. Stojanovic, "Consistent evolution of owl ontologies," in *Proceedings of the Second European Semantic Web Conference*, A. Gomez-Perez and J. Euzenat, Eds., vol. 3532. Heraklion, Crete, Greece: Springer, 2005, pp. 182–197.

[18] L. Stojanovic, A. Maedche, B. Motik, and N. Stojanovic, "User-driven ontology evolution management," in *EKAW '02: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management*. London, UK: Springer-Verlag, 2002, pp. 285–300.

[19] P. Hayes, "Rdf semantics," 2 2004. [Online]. Available: http://www.w3.org/TR/rdf-mt/

[20] A. Maedche and V. Zacharias, "Clustering ontology-based metadata in the semantic web," in *13th European Conference on Machine Learning (ECML'02) 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02)*, T. Elomaa, H. Mannila, and H. T. T. Toivonen, Eds., Helsinki, Finland, 2002.

[21] X. Polanco, "Concepts, measures and indicators in the web analysis," T. T. de Obtencin de Indicadores de Produccin Cientfica, Ed., Madrid, March 3-5.

[22] C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks, "Data driven ontology evaluation," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, 2004.

[23] A. Lozano-Tello and A. Gomez-Perez, "Ontometric: A method to choose the appropriate ontology," *Journal of Database Management*, vol. 15, no. 2, pp. 1–18, APR-JUN 2004.