A Generic Implementation Framework for Measuring Ontology-Based Information

Yinglong Ma [1,2], Xiaolan Zhang [1] , Beihong Jin [2] , Ke Lu [3]

[1] *School of Control and Computer Engineering, North China Electric Power University, #2 Beinong Road, Beijing 102206, China, E-mail: yinglongma@gmail.com*

[2] *State key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, #4 Nansi Street, Beijing 100190, China,*

[3] *University of Chinese Academy of Sciences, Beijing 100049, China,*

Abstract

As a knowledge representation tool, ontologies have been widely applied in many fields such as knowledge management and information integration, etc. Ontology measurement is an important challenge in the field of knowledge management in order to manage the development of ontology based systems and reduce the risk of project failure. This paper proposes a generic implementation framework for stable semantic ontology measurement. Through this framework, an ontology will be measured according to its semantic enriched representation model (SERM). The SERM model of an ontology can be used for stably measuring the semantics of the ontology. Then ontology metrics are integrated into the framework to measure candidate ontologies according to its SERM model. The related experiments are made to show that the framework can effectively measure the semantics of ontologies.

Keywords: Ontology, ontology engineering, Ontology semantics, Ontology measurement

1. Introduction

Ontology engineering is a subfield of knowledge engineering [1,2], and has spurred the development and application of ontologies. As a knowledge representation tool, ontologies provide the shared semantic vocabulary that agrees on domains of interest. They play an important role in the development of semantic-driven knowledge systems [3,4,5,6,7,8]. Currently, plenty of domain ontologies for their ontology based systems have been developed in a manual or (semi-)automatic manner. Ontology engineers often reuse ontologies by evaluating the existing ontologies [10,11].

Measuring ontologies effectively and correctly is the precondition on which the meaningful and useful ontology evaluation can be made. Some of the existing approaches such as the literature [12,13,14,15,16], have been only successfully used to measure the less expressive and explicitly expressed ontologie. However, the ontologies often built in real world have more expressivity, which include many concepts residing in complex concept constructors. If these approaches are used to measure more expressive ontologies, then they will possibly neglect much of the implicit semantic information of ontologies. Meanwhile, ontology knowledge can be represented in a very flexible way. That is, the same semantic knowledge can possibly be represented in different lexical structures. So the same semantic knowledge will

possibly have variable measurement results. This will cause unstable and meaningless ontology measurement. We argue that a generic ontology measurement methodology should be suitable for stably measuring both less and more expressive ontologies.

In this paper, we propose a generic implementation framework for stably and automatically measuring the semantics of ontologies. The framework is designed and implemented in a modular manner. An ontology transformation approach is adopted to generate the semantic enriched representation model (SERM) of an ontology. Furthermore, an automatic approach is used to collect the measurement entities based on SERM models. At last, based on the collected measurement entities, ontology metrics can be integrated into the framework for performing ontology measurement tasks.

This paper is organized as follows. Section 2 introduces the related work. Section 3 is the preliminaries about ontology representation and ontology measurement. In Section 4, we give an overview of our ontology measurement framework in a modular manner. Section 5 gives a semantic enriched representation model (SERM) for ontologies, and discusses the process of generating SERM models including four steps. In Section 6, we discuss the uniqueness of SERM model and stability of ontology measurement.In Section 6, we summarize and collect the measurement entities based on SERM models by analyzing the types of measurement entities. Section 7 introduces how to integrate ontology metrics based on SERM models. In Section 9, we make the related experiments to show the effectiveness of our framework. Sections 10 is the conclusion and future work.

2. Related work

Most of existing ontology measurement frameworks [12,14,15,16] mainly consider ontology structure and seldom measure ontological semantics. Ensan and Du proposed four metrics NSLD, NMLD, NSED and NMED, which concentrate on measuring the semantics of ontologies for modular ontologies[24]. Oh et al. [25] also proposed some ontology cohesion and coupling metrics for modular ontologies.

Zhang et al. discussed some ontology complexity metrics [26]. The stability of measurement and the semantic measurement of ontologies were discussed in our previous work [17,22,23]. We proposed four ontology cohesion metrics, where metrics NMIS and AVAI are used to measure semantically inconsistent ontology [17,22]. In the literature [23], we first gave a formal definition about stability of ontology measurement. However, whether the existing ontology metrics can be integrated into a generic framework for measuring the ontologies with different expressivity remains to be deeply discussed, while this paper focuses on this issue.

A few representation models of ontologies were proposed for graphical visualization of ontologies. Rudolph et al. proposed an ordered binary decision diagram (OBDD) method [27]. The generated OBDD diagram is a generalization of binary decision trees. It is difficult for ontology engineers to intuitively understand in practice. The emerging semantic link model [28] is a description of semantic relations among objective existences, and pursues semantic richness rather than correctness.

To our knowledge, there is no generic implementation framework which works well for stably measuring the semantics of both less and more expressive ontologies.

3. Representation and Measurement for Ontologies

Ontology knowledge can be represented in some standard web ontology languages such as Web Ontology Language (OWL) recommended by W3C[18]. OWL includes the three inter-related parts: OWL Lite, OWL DL and OWL Full. They have different capabilities of semantic expressivity. From the viewpoint of structural semantics of ontologies, ontology can be regarded as a set of intramodule relationships, which is denoted SIR.

Definition 1. (**Intramodule Relationships, IR**) An intramodule-relationship (IR) refers to a binary relation between ontology elements, which is one of the following forms as follows.
— Relations between classes, `owl:subClassOf(A,B)`, `owl:equivalentClass(A,B)` and `owl:disjoint-`

With(A,B), respectively denote subsumption, equivalence and disjointness between Classes A and B.

— Domain or range relation, `owl:domain`(R, A) and `owl:range`(R, B), respectively represent the domain and range of a binary relation R, i.e. Classes A and B.

—Relation between properties, `owl:subProperty-Of`(R,S), denotes the property subsumption between Properties R and S.

— Relation between individuals, R(a, b), represents the association between Individuals a and b by R.

— Membership relation, `owl:type`(a, A), denotes that Individual a is an instance of Class A.

A class (a.k.a concept) is either atomic or anonymous. A complex concept consists of atomic concepts by some constructors such as `owl:equivalentClass`, `owl:unionOf` and `owl:Restriction`, etc. The ontologies with more constructors will have more expressivity, and vice versa. An ontology based on Definition 1 can be regarded as a graph based model, where nodes are concepts or individuals, and edges are IRs.

Measuring an ontology is a process of collecting and calculating the related ontological entities specified by ontology metrics. An ontology metric is generally defined by a formula that shows how the entities in the ontologies to be measured are calculated, and is an indicator of certain characteristics of ontologies such as ontology cohesion, complexity, coupling, and semantic coverage, etc. Entity types that ontology metrics are concerned about generally include classes/concepts, properties, depth/path, fanins, fanouts and partitions, etc.

4. Overview of Generic Framework

An overview of the implementation framework for stable semantic ontology measurement is shown in Figure 1. The framework includes four main components, i.e., generating semantic enriched representation model, entity summarization, integrating ontology metrics, and performing stable semantic ontology measurement. The four main components are sequential, i.e., a component is performed after the previous one is accomplished. In the following, we specifically discuss the framework.
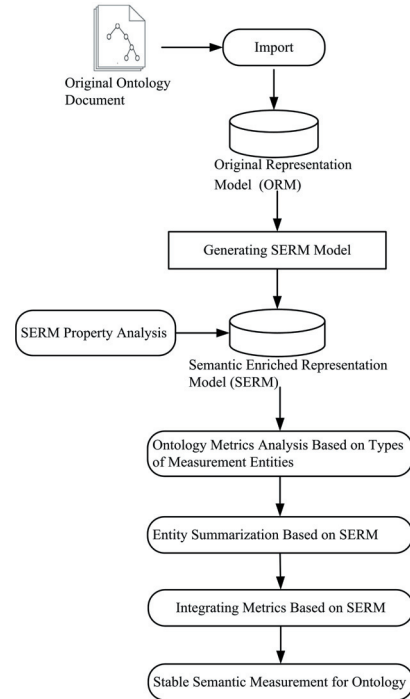

Fig. 1. The overview of framework

An original ontology document can be imported and loaded by some programmable ontology management platform such as KAON2 [19]. It can be built as an original representation model (ORM), which is a graph based model, where nodes are just the explicitly expressed concepts or individuals, and edges are the IRs between nodes.

The component of generating semantic enriched representation model is one of the core components of our framework. It is mainly to build the semantic enriched representation model (SERM) for an ontology by semantic derivation. A SERM model is also a graph based representation model, and is achieved by the four sequential steps such as refining the concepts, extracting IRs, Eliminating IR cycles and avoiding double counting. The details of generating SERM models for ontologies will be discussed in the section 5.

The component Entity Summarization is just to summarize all kinds of entities associated with entity types. It focuses on collecting the specific entities from the generated SERM models of ontologies

rather than their ORM models, where ORM is the original representation model which is used to represent the explicitly expressed knowledge in an ontology. For the future ontology measurement, once a new type of entities is summarized, the entities of the type will be collected into the corresponding sets of entities.

The component of integrating ontology metrics is to implement the specific ontology metrics based on SERM models. The reason why we do like that is to ensure the stability of measuring ontologies. Each integrated ontology metric is just to specify how to calculate the related ontology entities by a formula. Generally, the component needs to unambiguously implement each of the integrated ontology metrics. As such, we can easily collect and calculate the related metric values when we perform the tasks of ontology measurement.

At last, ontology engineers can perform the process of ontology measurement according to the algorithm that an ontology metric defines.

Note that, the framework proposed in this paper is for ontology measurement rather than ontology evaluation. As mentioned in the previous section, ontology measurement is a process of collecting and calculating the related ontological entities specified by ontology metrics. However, generally speaking, ontology evaluation is a process that associates ontology metric values with certain ontology quality characteristics, and then validates the associations between them by many ontology cases. As such, we can explore which metrics should be used as indicators of certain quality characteristics such as complexity, coupling and cohesion, etc. We argue that a uniform and effective ontology measurement methodology is the precondition on which effective ontology evaluation can be made. So this paper concentrates on proposing a uniform and effective framework for ontology measurement rather than ontology evaluation which is beyond the scope of this paper.

5. Semantic Enriched Representation Model

Generating semantic enriched representation model (SERM) for an ontology is one of the core tasks of our framework. A SERM model of an ontology is built mainly by refining ontological concepts, extracting IRs of ontologies and Eliminating IR cycles. Meanwhile, in considering some principles in statistics, double counting in calculating ontological entities should be also avoided. A SERM model of an ontology is generated based on its original representation model (ORM).

Definition 2. (**SERM model**) A SERM model SERM=(N,R,L), is a triple, where N is a set of nodes, R⊆N×N is the set of IRs, and L is a labeling function which assigns each r∈R a relation name of IRs.

5.1. Refining Concepts/Classes

In ORM, nodes are only the explicitly expressed concepts. For less expressive ontologies, especially for the ontologies whose concepts are all atomic, ontology measurement based on ORM is no problem. However, if an ontology includes many complex concepts, then its ORM needs to be extended because ORM does not contain these complex concepts. Considering that OWL is the standard ontology language, we here review all constructors building complex concepts in OWL. According to OWL syntax, OWL has the six forms of class constructors: `owl:Restriction`, `owl:unionOf`, `owl:intersectionOf`, `owl:oneOf` and `owl:complementOf`, as well as `owl:Class`, where `owl:Class` is used to define an atomic class, and the others are for building anonymous classes. We need to traverse all possible constructors in an ontology when we want to refine all the classes in the ontology. By XML parser, we can detect these anonymous classes, and assign each of them a unique name. The naming of anonymous classes is implemented by `_:1`, `_:2`, ⋯, `_:n`. As such, each anonymous class will be explicitly expressed in an ontology, and further added into its SERM model.

Note that some anonymous classes may be nestedly defined by other anonymous classes. For example, Figure 2 is a description of the anonymous class describing "Thing with Ph.D degree and Professor title". The anonymous class is nestedly defined by the two nested anonymous classes, i.e., Class

1 and Class 2. Here, Class 1 represents the concept "Thing_with_Ph.D_degree", and Class 2 represents the concept "Thing_with_Title_Professor". In fact, we proposed an algorithm that can automatically find out all concepts in a recursive manner [20].

```
<owl:Class>
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasDegree"/>
      <owl:SomeValuesFrom rdf:resource="#Ph.D"/>       Class 1
    </owl:Restriction>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTitle"/>
      <owl:SomeValuesFrom rdf:resource="#Professor"/>  Class 2
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
<owl:Class rdf:ID="Professor"/>
<owl:Class rdf:ID="Ph.D"/>
```

Fig. 2. An example of nestedly defined concept

Detecting and recognizing anonymous classes are crucial to ontology measurement because they are indispensable parts of ontology representation. If they are ignored in ontology measurement, then we can not capture semantics of ontologies. Figure 3(a) is the ORM model in Figure 2, and Figure 3(b) is the semantic model of the ontology in fully considering all possible concepts and individuals. An ellipse is to represent a concept, and a dashed arrow with a rectangle represents a binary relation, where the rectangles are the label names of IRs.
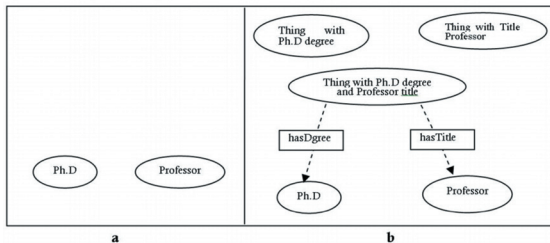

Fig. 3. Two kinds of graph-based ontology representations

5.2. *Extracting Intra-module Relationships (IRs)*

In ontology measurement, IRs are crucial for an ontology to measure its faouts, fanins and depth, etc. Once all classes including atomic and anonymous classes are found out, IRs between them should be also extracted. Especially, those implicit IRs derived from the explicit knowledge should be excavated into SERM. As such, the SERM of an ontology can preserve the semantics of the original ontologies as more as possible.

Extracting of implicit IRs means to excavate the five types of IRs mentioned in Definition 1. However, in considering that some IRs are implicit in an ontology, the first step is to construct a semantic model of the ontology by the existing tools of ontology management and reasoning such as KAON2. The semantic model constructed in these tools possibly has special data formats. Such a semantic model is not suitable for ontology measurement because most of the existing ontology measures are based on graphs associated with ontologies. Then, we need to traverse the associated relations in the semantic model, and extract them to IRs in the corresponding SERM.
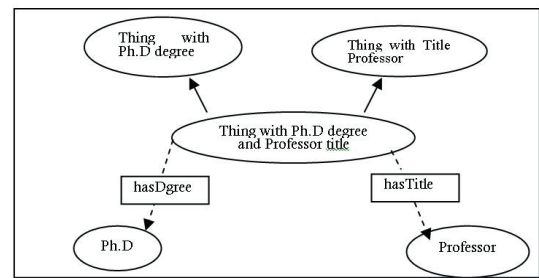

Fig. 4. Semantic enriched graph-based ontology representation

Taking Figure 3 as an example, we find that Concept "Thing_with_Ph.D_degree_and_Professor_Title" is not only a subclass of the concept "Thing_with_Ph.D_degree", but also a subclass of the concept "Thing_with_Title_Professor". These implicit subClassOf IRs should be excavated and explicitly expressed in the ontology to construct the semantic enriched ontology representation. Similarly, the other types of IRs in the semantic model can also extracted. What we need to do is to search all labels defined in Definition 1, e.g., `owl:subPropertyOf`, `owl:domain`, `owl:range`, `owl:type`, and so on. For each of the detected IR labels, we first determine their two endpoints (classes or individuals), then transform it into the corresponding IR in SERM model. Figure 4 is the graph based representation after mining of implicit IRs in Figure 3(b), where an arrow is a subClassof IR.

5.3. *Eliminating IR Cycles*

The IRs for class inheritance (a.k.a. subsumption) are what ontology measurement is concerned about because such IRs reflect the semantics of ontologies. Moreover, they are always associated with some ontology measures such as path and depth, etc. A path should not include any cycle. Otherwise, its depth is variable, which is an important cause of impeding the stability of ontology measurement. However, cycles of the subClassOf IRs often exist in ontology representation. So eliminating IR cycles for class inheritance is considered as an important step in this paper, and is necessary to ensure stability of ontology measurement.

Strictly speaking, a cycle of subClassOf IRs is of the form `owl:subClassOf(A, A_1)`, `owl:subClassOf(A_1, A_2)`, $\cdots$, `owl:subClassOf` (A_n, A), where $A, A_1, A_2, \cdots, A_n$ are concepts. The specific treatment to cycles of subClassOf IRs is to detect all cycles of IRs based on the traditional deep first search (DFS) algorithm in Graph Theory. Then, for each cyclic of IRs, we replace each A_i ($1 \leqslant i \leqslant n$) with A. As a result, the cycles of subClassOf IRs are compressed into one concept node A, and A has its multiple equivalent names, i.e., A_i ($1 \leqslant i \leqslant n$).

In fact, the IRs for class equivalence can also cause instable measurement. Formally speaking, Class A is equivalent to Class B, i.e., `owl:equivalentClass(A, B)`, means that both `owl:subClassOf(A, B)` and `owl:owl:subClassOf(B, A)` hold. So the treatment to cycles of equivalentClass IRs is also similar to that to cycles of subClassOf IRs.

5.4. *Avoiding Double Counting*

Double counting is a statistical error which should be avoided in ontology measurement [21]. Double counting of ontological IRs is mainly caused by the transitivity relationships. We can eliminate double counting by deleting the transitivity derived IRs. For examlples, `owl:subClassOf(A, C)` is indirectly obtained by the two related IRs `owl:subClassOf(A, B)` and `owl:subClassOf(B, C)`. If the two IRs are counted, then `owl:subClassOf(A, C)` should no longer be counted because counting `owl:subClassOf(A, C)` means the double counting to the two IRs. This is necessary to make ontology measurement satisfy the basic principle of measurement.

6. Uniqueness of SERM Models

Semantic enriched representation models are proposed for stable semantic measurement of ontologies. According to the process of generating SERM models, we fully consider the semantics of all of the possible elements consisting of ontologies, i.e., concepts, instances, properties/relationships and axioms. Concepts are nodes in SERM models. Instances (if any) also corresponds nodes in SERM models. Ontological properties and axioms are regarded as different types of IRs in SERM models. Each entity in a SERM model for an ontology is strictly in accordance with the underlying semantics of the ontology. Formally speaking, let $\mathscr{O}$ be an ontology, and $SERM_{\mathscr{O}}$ be the SERM model of $\mathscr{O}$ according to the principles of generating SERM models mentioned in Section 5. For any defined element *elem* in $\mathscr{O}$, $\mathscr{O} \models elem$ implies $SERM_{\mathscr{O}} \models \rho(elem)$. So we can obtain the Lemma 1.

Lemma 1. *$SERM_{\mathscr{O}}$ is a semantic-preserving representation model of Ontology $\mathscr{O}$.*

We can further find that the process of generating SERM models is monotonic. So we can obtain the Lemma 2.

Lemma 2. *The process for generating the SERM model for Ontology $\mathscr{O}$ is terminable if and only if there is no element in $\mathscr{O}$ for the SERM model.*

The classes defined in an ontology are finite. The relationships between these finite classes are obviously finite. So the process of generating SERM models can always find out all defined elements in an ontology and add them into its SERM model until there is no elements for SERM models. So the process for generating the SERM model for Ontology $\mathscr{O}$ must be terminable.

On the other hand, it is impossible that the same ontology $\mathscr{O}$ has multiple different SERM models. Assume that Ontology $\mathscr{O}$ has two SERM models, respectively denoted $SERM_{\mathscr{O}}$ and $SERM'_O$. We further assume that there exists an element α such that

$SERM_{\mathscr{O}}$ satisfies α, and $SERM'_{\mathscr{O}}$ does not satisfy α, i.e., $SERM_{\mathscr{O}} \models \alpha$ and $SERM'_{\mathscr{O}} \not\models \alpha$. Then there exists the element α that is an element in $\mathscr{O}$ but is not generated for $SERM'_{\mathscr{O}}$. This means that the process of generating $SERM'_{\mathscr{O}}$ has not been terminated according to Lemma 2. Thus, $SERM'_{\mathscr{O}}$ is only an intermediate model rather than a SERM model. So we can conclude the following theorem.

Theorem 3. *The SERM model of an ontology is unique.*

The uniqueness of final semantic derived models guarantees that the measurement results between different ontologies should be comparable, and hence such ontology measurements are stable. We therefore can obtain the conclusion.

Theorem 4. *Ontology measurement based on SERM models is stable.*

7. Entity Summarization

The measurement entities for ontologies are just the elements consisting of ontologies. The difference between measurement entities and ontological elements is that they possibly have different granularities. Measurement entities are fine-grained if they are the basic element of ontologies, e.g., concepts/classes, properties, axioms and instances. Coarse-grained measurement entities are the intrinsic constructs of ontologies, which consist of different kinds of basic ontological elements.

By the extensive analysis of most of the existing ontology metrics, we summarize the measurement entities of the existing ontology metrics [12,13,14,15,16,17,24,25], most of which mainly include concepts/classes, properties, class inheritance, axioms, instances, fanouts, fanins, paths and partitions. In these types of measurement entities, classes, properties, class inheritance, axioms and instances belong to the fine-grained types of measurement entities because they are the basic elements consisting of ontologies. In contrast, fanouts, fanins, paths and partitions are coarse-grained types of measurement entities because they themselves consist of some fine-grained types of measurement entities.

Let $\mathscr{O}$ be the measured ontology, and $SERM_{\mathscr{O}}$=(N, R, L) be its final semantic derived

model. The sets of the different kinds of measurement entities in $SERM_{\mathscr{O}}$ are represented as follows.

The set of classes: $SC_{\mathscr{O}}=\{v|v \in N$ and v is a class node$\}$.

The set of leaf classes: $SLC_{\mathscr{O}}=\{v|v \in N$ and $\nexists v'((v',v) \in R \wedge L((v',v))=$owl:subClassOf$\}$.

The set of properties: $SP_{\mathscr{O}}=\{p|p \in R\}$.

The set of instances: $SI_{\mathscr{O}}=\{v|v \in N$ and v is an instance node$\}$.

The set of class inheritance: $SCI_{\mathscr{O}}=\{s|s \in N \times N \wedge L(s)=$owl:subClassOf$\}$.

The set of ontological axioms: $SA_{\mathscr{O}}=\{s|s \in SC_{\mathscr{O}} \times SC_{\mathscr{O}}\}$.

The set of fanouts w.r.t class C: $SFO_C=\{f|f \in SC_{\mathscr{O}}$ and $L((f,C))=$owl:subClassOf$\}$.

The set of fanouts w.r.t class C: $SFI_C=\{f|f \in SC_{\mathscr{O}}$ and $L((C,f))=$owl:subClassOf$\}$.

8. Integrating Ontology Metrics into Framework

Based on these sets of measurement entities, we can integrate the ontology metrics related to basic types of measurement entities into our framework. First, the entities that ontology metrics are concerned about should correspond to the entities in the SERM model of the measured ontology. Second, the related formula and algorithm should be also translated so that these ontology metrics could work in SERM models. For example, we use the metric Inheritance richness (*ir*) proposed by Tartir et al. [14] for measure an ontology $\mathscr{O}$. Metric *ir* is defined as $ir = \frac{\Sigma_{C_i \in \mathscr{C}}|H^c(C,C_i)|}{|\mathscr{C}|}$, where $|H^c(C,C_i)|$ is the number of subclasses (C) for a class (C_i) and the divisor $(|\mathscr{C}|)$ is the total number of classes. If we want to perform stable semantic ontology measurement, then the stable semantic measurement w.r.t *ir* for $SERM_{\mathscr{O}}$ is easily obtained by the corresponding equation: $ir = \frac{\Sigma_{v_i \in SC_{\mathscr{O}}}|SFO_{v_i}|}{|SC_{\mathscr{O}}|}$.

By Section 7, we can directly summarize some basic measurement entities for ontologies. These entities types can be also directly expressed by SERM models for ontologies. However, some coarse-grained measurement entities such as paths and partitions, are not directly expressed. Additional procedures need to be used to obtain the coarse-grained

entities for ontologies. For example, a partition of an ontology, *de facto*, is a subontology of the ontology. Because our ontology measurement is based on SERM models of ontologies, a partition of an ontology is a maximal sub-SERM model of the ontology.

Definition 3. (Sub-SERM model) Let S=(N, R, L) and S'=(N', R' L') be two SERM models. If S' is a sub-SERM model of S, denoted S'⊆S, if and only if S'⊆S, R'⊆R and ∀r(r ∈ S'∧r ∈ S ⟹ L'(r)=L(r).

For the sake of simplification, if S=(N, R, L) is a SERM model, then we use S.N and S.R to obtain the sets N and R of S, respectively.

Definition 4. (Partition in SERM models) PT is a partition of the SERM model S if and only if the followings hold:
— PT⊆S,
— if ∀PT'⊆S such that PT.N∩PT'.N≠ ∅, then PT'⊆PT.

It is not difficult to find that Definition 4 gives us a procedure to obtain a partition of a SERM model. In fact, we also can easily obtain the set of all partitions of the SERM model.

Another coarse-grained measurement entity type is path. A path of a SERM model is a sequence of class inheritance from the root class to a leaf class. From the definition of SERM models, we cannot directly obtain the procedure of getting a path in SERM models, so we need to deeply refine paths in SERM models such that we can integrate some ontology metrics related to paths into our framework.

Depth of ontologies is an important characteristic of ontology quality. However, It rests on the measurement to paths of ontologies. In considering that ontology metrics needs to be integrated into our framework, the depth of an ontology is just the depth of its SERM model. The depth of a SERM model refers to the depth of the paths with the maximum number of nodes. The depth of a path is the number of nodes in the path. However, it is possible for an ontology or its SERM model to have multiple sequences of class inheritance from the root class to a leaf class. That is, there are possibly multiple paths from the root class to a leaf class.

Definition 5. (Path of SERM model) Let SERM$_\mathcal{O}$=(N, R, L) be the SERM model of an on-

tology $\mathcal{O}$, and RC∈N be the root class of $\mathcal{O}$. The ith path to leaf class C: $P_C^i=\{v_1, v_2, \cdots, v_m\}$, where $v_1, v_2, \cdots, v_m \in SC_\mathcal{O} \wedge v_1 =RC\wedge v_m =C\wedge(v_k, v_{k+1}) \in$ R for all $1 \leqslant k \leqslant m-1$ }.

Definition 6. (Depth of path) The depth of the path P=$\{v_1, v_2, \cdots, v_m\}$ is denoted |P|, which is the cardinality of P.

Definition 7. (Depth of SERM model) Let S be a SERM model, and SP be a set of all paths from the root class to leaf classes. The depth of S is denoted depth(S), where depth(S)=$\max_{i=1}^{|SP|}\{p_i\}$, where p_i is the ith path in SP. is the cardinality of SP.

Definitions 5, 6 and 7 provide us a feasible procedure to obtain the paths and their depth. The related algorithms are also easily developed so that we can integrate ontology metrics related to paths and depth into our framework.

9. Experiments and Analysis

Current ontology measurement is made based on ORM models which cannot perform stable semantic ontology measurement for more expressive ontologies, so our experiment analysis concentrates on two aspects for validating the effectiveness of our methodology. On one hand, for those ontologies with less expressivity, especially for the ontologies in which all classes are atomic, we need to discuss whether the proposed framework works well for less expressive ontologies. On the other hand, for more expressive ontologies, we need to compare whether SERM models can excavate more entities for stable semantic ontology measurement as opposed to ORM models.

9.1. Measurement Effectiveness for Less Expressive Ontologies

In an ontology in which all classes are atomic, there is no anonymous class (i.e, complex concept). And thus, there is nothing to do in the step of refining classes because refining classes is just to explicitly find and name anonymous classes in the ontology to be measured. Furthermore, because there is no complex concept in the ontology, we cannot extract

any implicit IRs residing in complex concepts. So its ORM and SERM models are same. As for the step of eliminating IR cycles, there is a little difference between ORM and SERM. If there are the IR cycles in an ontology, then cycles still exist in its ORM model. In contrast, these cycles will not exist in its SERM model. In the case, if the ontology metrics related to path and depth are applied for measuring the ontology, then the ontology measurement based on ORM model possibly does not work. However, the ontology measurement based on its SERM model still work well. If no cycle exists in the ontology, its ORM and SERM models are also same. The step for avoiding double counting is similar.

By the analysis above, our SERM based framework will work well for less expressive ontologies.

9.2. Comparison between ORM and SERM Models for More Expressive Ontologies

An intuitive experiment is that we compare them by using the same ontology metrics respectively based on ORM models and SERM models. In order to do that, we further need to consider the two issues. The first is that how to select ontology metrics because there are many ontology metrics proposed. The second is that what are the data sets for testing ontologies.

9.2.1. Data Sets and Selection of Ontology Metrics

We randomly collected eight testing ontologies by using the Swoogle search engine [9]. The testing ontologies are as follows: Wine, Person, miniTambis, swrc, Terrorism, publication, univ-bench and GlycO.

As mentioned in Section 7, ontology metrics can be classified according to their relevant types of measurement entities. Measurement entities can be classified into two categories, i.e., fine-grained and coarse-grained measurement entities. In the fine-grained measurement entities, classes are the most important entities because the other measurement entities rests on classes. So we select the ontology metrics related to classes, number of classes (NOC), as a representative in fine-grained measurement entities. For coarse-grained measurement entities, all

of them deal with the set of IRs of class inheritance, e.g., fanouts, fanins, and path/depth. We select the ontology metric , average depth inheritance tree of leaf nodes (ADIT-LN), as a representative in fine-grained measurement entities. It is related to paths. The definitions of the two metrics refer to the literature [12,13,14,15,16]. Formally, based on entity summarization of SERM models, the formulae of the two metrics w.r.t an ontology $\mathscr{O}$ are as follows.

$$NOC(\mathscr{O}) = |SC_{\mathscr{O}}| \qquad (1)$$

$$ADIT-LN(\mathscr{O}) = \frac{\sum\limits_{C \in SLC_{\mathscr{O}}} depth(C)}{|SLC_{\mathscr{O}}|} \qquad (2)$$

9.2.2. Experiments and Analysis

By using these testing data sets, their ORM and SERM models are respectively built for each of the testing ontologies. Then, we use the two ontology metrics to respectively measure their NOC and ADIT-LN values w.r.t the two models. The measurement results of the two representation models of each of the ontologies are shown in Table 1. Each measurement value in Table 1 is of the form $\frac{X}{Y}$, where X is the ADIT-LN value, and Y is the number of classes/concepts.

Table 1. Comparison of ORM and SERM models

	SERM Model	ORM Model
Wine	$\frac{3.44}{272}$	$\frac{1.27}{76}$
Person	$\frac{3.71}{26}$	$\frac{2.71}{21}$
miniTambis	$\frac{3.47}{270}$	$\frac{1.61}{182}$
SWRC	$\frac{3.34}{105}$	$\frac{2.56}{105}$
Terrorism	$\frac{3.23}{28}$	$\frac{2.5}{21}$
Publication	$\frac{3.82}{24}$	$\frac{2.82}{13}$
Univbench	$\frac{3.79}{51}$	$\frac{2.53}{43}$
GlycO	$\frac{8.06}{496}$	$\frac{8.35}{370}$

Through the ontology measurement w.r.t ADIT-LN, we find the following facts:

1) The ontology measurement based on SERM Models can find more classes than ORM Model ac-

cording to the NOC values in the table. That's because some implicit classes or nested anonymous classes can be found. As opposed to SERM models, the ontology measurement based on ORM models only find those classes that are explicitly defined by some labels such as `owl:Class`. The implicit relations between the newly found classes can be excavated and form a more enriched semantic network than ORM models. Obviously, SERM models can represent more complete semantic information, which will be helpful in stably measuring the semantics of ontologies.

2) A key problem found in the table is that the two groups of ADIT-LN values between SERM models and ORM models are not linear. This means that there is no obvious correlation between the two groups of values. That is, if we want to predict the quality of ontologies, the measurement results based on ORM models and SERM models will reflect variable predictions. A question is further identified, the prediction of which model we can trust? Considering that classes are one of the key components of an ontology, we believe that an ontology cannot be accurately measured if the number of classes of the ontology is estimated too low just based on ORM models. What causes is that the effectiveness of evaluation in the future could be influenced because of the difference of measuring ontological entities. As opposed to ORM models, our SERM based framework can find ontology measurement entities and their associate relations as more as possibly, and theoretically ensure the uniqueness of data model of ontology measurement. There is no doubt that the prediction based SERM models is more credible for predicting the semantic quality of ontologies.

What is noted that ontology measurement based on SERM models requires ontology reasoning through which implicit information can be excavated and form a stable semantic model of ontologies to be measured. However, semantic reasoning inevitably increases the processing time for ontology measurement. Especially for those very large volume of data sets (e.g., ontology knowledge more than 100MB), reducing the processing time for ontology measurement is still challenging.

10. Conclusion

The main contributions of this paper are as follows. 1) We presented an implementation framework for stable semantic measurement of ontologies, and designed and implemented the framework in a modular manner. 2) We proposed SERM models for ontologies, and discussed how to generate the SERM model for an ontology. The uniqueness of SERM models and stability of ontology measurement are theoretically analyzed. 3) We also made the related experiments to illustrate the effectiveness of our framework and compare the difference between the two kinds of models. Our framework for stable semantic ontology measurement is generic and suitable to measure both less and more expressive ontologies.

The future work includes the following aspects. 1) We will exploit the optimization algorithms of semantic reasoning for reducing the processing time of stable semantic ontology measurement. 2) An important application for ontology measurement is to select from ontology libraries the candidate ontologies that are most similar to request ontology by measuring the dis/similarity between them. Another work is to compare ontology similarity by stably measuring the semantics of ontologies for selecting ontologies and enhancing the design quality of domain ontologies.

References

1. Mizoguchi, R. 'Tutorial on ontological engineering: part 3: Advanced course of ontological engineering', New Generation Computing, vol.22, no.2, pp.198-220(2004) .

2. Gomez-Perez, M., Fernandez-Lopez, M., and Corcho, O. Ontological Engineering: With Examples from the Areas of Knowledge Management, E-commerce and the Semantic Web, Springer(2004).

3. Razmerita C. 'An Ontology-Based Framework for Modeling User Behavior - A Case Study in Knowledge Management', IEEE Transaction on Systems, Man and Cybernetics - Part A, Vol.41, No.4, 772-783(2011).

4. Sicilia, J., Sicilia, M.A., Sonchez-Alonso, et al. 'Knowledge representation issues in ontology-based clinical Knowledge Management systems', International Journal of Technology Management, vol.47, no.1-3, pp.191-206(2009).

5. Raymond Y.K. Lau, Chapmann C.L. Lai and Yuefeng Li. 'Fuzzy Ontology Mining and Semantic Information Granulation for Effective Information Retrieval Decision Making'. International Journal of Computational Intelligence Systems, vol.4, no.1,pp.54–65, (2011).

6. Jordi Conesa, Santi Caballe, David Ganon and Josep Prieto. 'Exploiting the Semantic Web to Represent Information from On-line Collaborative Learning'. International Journal of Computational Intelligence Systems, vol.5,no.4, pp.653–667, (2012).

7. Souza, H., Moura, A. and Cavalcanti, M. 'Integrating Ontologies Based on P2P Mappings', IEEE Transaction on Systems, Man and Cybernetics - Part A, Vol.40, No.5, pp.1071-1082(2010).

8. Alexiev, D., Breu, M., de Bruijn, J., et al. Information Integration with Ontologies: Experiences from an Industrial Showcase, John Wiley & Sons, (2005).

9. Li, D., Finin, T., Joshi, Y., et al. 'Swoogle: A Search and Metadata Engine for the Semantic Web', Proceedings of 13th ACM Conference on Information and Knowledge Management (CIKM), pp.652-659, (2004).

10. Simperl, E. 'Reusing ontologies on the Semantic Web: A feasibility study', Data & Knowledge Engineering, vol.68, no.10, pp.905-925(2009).

11. Gaeta, M., Orciuoli, F., and Paolozzi, S., et al. 'Ontology Extraction for Knowledge Reuse: The e-Learning Perspective', IEEE Transaction on Systems, Man and Cybernetics - Part A, Vol.40, No.4, pp.789-809(2011).

12. Orme, A., Yao, H., and Etzkorn, H. 'Indicating ontology data quality, stability, and completeness throughout ontology evolution', Journal of Software Maintenance and Evolution: Research and practice, vol.19, no.1, pp.49-75, (2007).

13. Gangemi, C. Catenaccia, M. Ciaramita, J., et al. 'Qood grid: A metaontology based framework for ontology evaluation and selection', Proceedings of WWW2006 Workshop: Evaluation of Ontologies for the Web (EON2006)(2006).

14. Tartir, S., Arpinar, I., Moore, M., et al. 'OntoQA: Metric-based ontology quality analysis', Proceedings of IEEEWorkshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources(2005).

15. Orme, A, Yao, H., Etzkorn, L. 'Complexity metrics for ontology based information', International Journal of Technology Management, Vol.47, No.1-3 pp.161-173(2009).

16. Lozano-Tello A. and Gomez-Perez A. 'OntoMetric: A method to choose the appropriate ontology', Journal of Database Management, vol.15, no.2, pp.1-18(2004).

17. Ma, Y., Wu, H., Ma, X., et al. 'Stable cohesion metrics for evolving ontologies', Journal of Software Maintenance and Evolution: Research and Practice, vol.23, No.5, pp.343-359(2011).

18. W3C. OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features/, (2004).

19. Motik, B. Reasoning in Description Logics using Resolution and Deductive Databases. PhD Thesis, University of Karlsruhe, Karlsruhe, Germany, (2006)

20. Ma Y., Liu S., Jin B., Xu G.. Inconsistent ontology revision based on ontology constructs. Expert Systems with Application, vol.37(10), 7269-7275 (2010).

21. J. Lawler and B. Kitchenham. "Measurement Modeling Technology", IEEE Software, vol.20, no.3, pp.68–75, 2003.

22. Ma Y., Jin B., Feng Y.. Semantic oriented ontology cohesion metrics for ontology-based systems. Journal of Systems and Software, vol.83, no.1, pp.143-152, 2010.

23. Ma, Y. 'Towards Stable Semantic Ontology Measurement' Proceedings of International Semantic Web Conference (ISWC2010), pp.21-24(2010).

24. Ensan F., and Du, W. 'A Metric Suite for Evaluating Cohesion and Coupling in Modular Ontologies', Proceedings of 2010 Conference on Modular Ontologies (WoMO 2010)(2010).

25. Oh, S., Yeom, H., Ahn, J. 'Cohesion and coupling metrics for ontology modules', Information Technology and Management, vol.12, no.2, pp.81-96(2011).

26. Grau, B. Motik, G. Stoilos, I. Horrocks. 'Completeness Guarantees for Incomplete Ontology Reasoners: Theory and Practice', Journal of Artificial Intelligence Research, vol.43, pp.419-476(2012).

27. Rudolph S., Krotzsch M., Hitzler P. "Description Logic Reasoning with Decision Diagrams: Compiling SHIQ to Disjunctive Datalog", Proceedings of International Semantic Web Conference (ISWC2008), pp.435-450 (2008).

28. Zhuge H. Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning. IEEE Trans. Knowl. Data Eng, vol.21, no.6, pp.785-799 (2009).