2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Applications of Ontology in Management of Information Asset

Shing-Han Li
Department of Information Management,
Tatung University,
Taipei, Taiwan
e-mail: shli@ttu.edu.tw

Kuo-Chung Wang
Department of Information Management,
Tatung University,
Taipei, Taiwan
e-mail: jasoneric.3@yahoo.com.tw

Abstract—After being computerized, enterprises import various information assets to enhance competitiveness. In order to make management of information asset more efficient, reduce the difficulty of auditing and make professional knowledge of this domain to reach the goal of share and reuse through the concept and technique of semantic web, the research imports the methodology of knowledge engineering and combines auditing regulations composed by SWRL to build an information-asset knowledgebase system. The methodology of this research can divide into eight steps. The first six ones build professional knowledge of information asset to ontology according to OWL, RDF and technique of semantic web combined with the analyzing of information asset. The last two not only combine regulations in SWRL and utilize information-asset knowledgebase system built by Protégé, but also integrate database and import inference engine to implement. The methodology can assist auditors to know the status of information asset well at any moment. It can also support auditors to discover the problems apropos in efficiency. In addition to detecting harmful individual of information asset, it also can verify that all information assets are authorized and dealt with appropriately. Besides, it can make professional knowledge of this domain to reach the goal of share and reuse through the concept and technique of semantic web.

Keywords: *Information Asset, Information Security, Ontology, Knowledge Base*

I. INTRODUCTION

According to "2006 CSI/FBI Computer Crime and Security Survey," the enterprises generally pay much attention to cognitive education of information security, nearly 80% (77%) emphasizing the announcement of information security policy, the rest, internet security (76%) and the training of information security management (72%). In addition, the survey discovers that more than 60% enterprises still have losses in business owing to inappropriate information security management, such as improper use of resources, leakage of data and negligence of inner management. Information asset is an important part of information management. It can be probed in two aspects. So far as software, the enterprise could just use traditional audit and control to manage physical assets in the past. It not only wasted manpower and time, but also cost a great deal of money to buy expensive software. If we cannot control and manage staff in using computer effectively, it may cause the abuse, like downloading, installing illegal software, browsing illegal websites or taking up time to do personal works. If unworthy employees steal the confidential information of corporation, it not only debates the efficiency of enterprises, but also puts them in the risk to go out of business; In terms of hardware, it will be a difficult problem if there are problems in the employees' computers but MIS cannot control in time. How to save manpower, time and manage efficiently is also an arduous one. Therefore, the research collects the information about information asset and issues a monographic study, aiming at the questions above.

Consequently, the objective of this research is to combine the high-speed operation ability and the methodology of knowledge engineering to bring up an information asset management system. We expect to construct the knowledge of information asset to ontology by the technique of semantic web, and time data collection to reach the goal of knowing well the use statuses of employees' computers, comprehending the statuses of hardware devices and software in order to decrease the inconveniency of information asset management. For the sake of raising the reliability of audit, the model in this research is also to organize information asset and match the audit rules as to make the estimation hard err, and to find the harmful employees' computer to enterprise by semantic inference. Furthermore, in order to enable auditors easy to comprehend and apply, the users beyond the field to understand, this research establishes the information asset knowledge base which can help the system comprehend the meaning of users' queries, simplify the miscellaneous of audit rules and reach the purpose of the share and reuse of professional knowledge.

II. RELEVANT BACKGROUND

A. Semantic Web

Semantic Web is an extension of the prevalent Web where information is given well-defined meaning, better enabling computers and people to process in cooperation. Resources on the Web are given explicit meaning using markup language to make up a metadata layer in addition to the current information pool. The Web Ontology Language (OWL) [2] provides rules for defining knowledge structures, i.e., ontology, in order that instances of knowledge can be created based on the common structures. OWL is an extension to the Resource Description Framework (RDF) [1] by adding vocabularies used to define the knowledge structures instead of the tree structures in XML documents. The Semantic Web framework can be summarized as providing a metadata layer in content-interoperable

978-0-7695-3762-7/09 $26.00 © 2009 IEEE
DOI 10.1109/IIH-MSP.2009.40

languages, mainly in RDF, of which intelligent or automatic services can be made by machines based on the layer. The typical architecture for managing the metadata layer, for example, KA2 [7] and Sesame [4], consists of an ontology-based knowledge warehouse and inference engine as the knowledge-based system to provide intelligent services at the front end, such as conceptual search and semantic navigation for user to access the content as summarized in Figure 1. In the backend is the content provision component, consisting of various tools for creating metadata from unstructured, semi-structured and structured documents. In this paper, we employ an annotation tool to annotate existing resources in order to create metadata in RDF. In addition, we can use RDF editor and wrapper to create metadata in RDF.

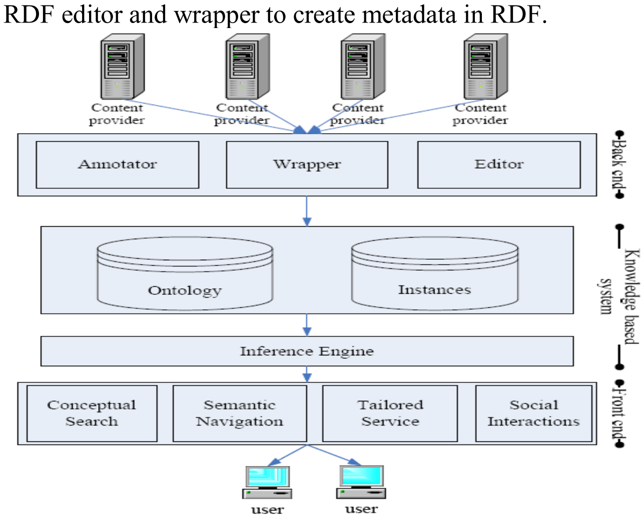

Figure 1. System Architecture for Managing the Metadata Layer of Semantic Web

B. RDF

The conceptual model of RDF is a labeled directed graph [1] where the nodes in the graph resources are identified by using URI [8] or literals representing atomic values of some kinds. The label on the directed arc connecting two nodes, from a resource to another or a resource to a literal, representing the relationship between both ends. An arc, also identified by using URI, connecting two nodes is a triple of the subject-predicate-object, similar to the structure of simple declarative sentence. Thus an RDF document can be taken as a list of triples. A triple states a fact about a subject, a resource on the Web. Since the nodes in RDF are identified using the addressing scheme of the Web, an RDF document can be easily combined with other to form an integrated description of resources on the Web.

C. SWRL

SWRL is a rule language based on ontology [3]. It describes antecedent-consequent relationships among entities in knowledge base using Horn-like rules. A rule, identified by an URI, is composed of the head, i.e., the consequent part and the body, i.e. the antecedent part. Each part consists of a number of atoms. An atom can be C(x), P(x, y), sameAs(x, y), differentFrom(x, y), or other build-in predicates. An atom C(x) holds x, an instance of the class C or data range C. An atom of P(x, y) is true if x is related to y by property P. For same As and differentFrom to be true if x and y are identified as the same or different objects. Details of other build-in predicates can be found in the paper [3]. The classes, data range, properties used in the atoms of rule are selected from the ontology, that is, the OWL schema. An OWL ontology is created to describe the concepts and relationships between concepts in a domain. SWRL is used to further describe the antecedent-consequent relations, so that we can create knowledge bases on the Semantic Web that has not only schema knowledge but also inference capabilities.

III. RESEARCH METHOD

The development of this research establishes information asset knowledge base mainly via six steps brought by "Ontology Development 101"[5], classification of information asset, RDF and OWL. Because only ontology cannot reach the purpose of inference, the research enhances the six steps to eight and uses SWRL to establish related audit rules on related connection of correct concept on meaning.

Step 1. Determine the domain and scope of the ontology
Step 2. Consider reusing existing ontologies
Step 3. Enumerate important terms in the ontology
Step 4. Define the classes and the class hierarchy
Step 5. Define the properties of classes—slots
Step 6. Define the facets of the slots
Step 7. Use SWRL to establish audit rules
Step 8. Import the inference engine

In this sectiuon we focus on the construction of knowledge structures from the information assets and our target is represented in the standard languages of the Semantic Web, namely, OWL, RDF and SWRL. The construction task can be divided into eight steps: building the domain ontology, in RDF and OWL, by consulting the information assets, and then analyzing the rules in audit and forming rules in SWRL. For the first six steps of the task, we employ "Ontology Development 101"[5] employed by the Protégé project [6] to the construction of ontology. After identifying the domain and scope in the initial step, we realize that there is no existing ontology for our purpose. Therefore we create one for ourselves. We then create a list of important terms in the information assets, which contains potential concepts and relationships in the domain schema. Part of the list of terms is taken as the classes. In step 4, we identify the parent-child class relationships with the method of "Bottom-Up" in them as well, which forms a class hierarchy. We use Protégé as the ontology editor to record what we have obtained so far. The next two steps of the list of terms is considered as the relationships in the domain and is entered as properties of classes. After the class hierarchy and slots in classes are created, we can enter instances for respective classes. For the remaining two steps of the task, we carefully examine each rule in order in common audit and analyze the rule parts of the condition and the conclusion part of the rule to find out the objects and properties in them. Then we formulate rules in SWRL according to the common audit rule. Since the objects and relationships in each rule have been extracted and built in domain schema, the corresponding rule in SWRL therefore is constructed from the terms in the schema. To create SWRL rules, we add an SWRL plug-in into Protégé and facilitate the GUI provided to edit the rules. At last step, we import domain ontology and the audit rules in SWRL into inference engine. The ontology will change into the fact list of Jess, SWRL will change into the rule base of Jess and integrating audit rule and fact to be computerized by Jess. The new facts are probably produced after inference are just recessive knowledge and can be the facts to infer. Besides, these facts also write back into ontology. In the following, we show all the steps to illustrate the process of establishing information asset knowledge base and take examples of the audit rules in SWRL.

Classes and properties in the rules:

- If space of Logical Devices of certain information asset is smaller than the limit set by knowledge base, the caution will be sent out.
- If motherboard_ID of certain information asset the same as the one recorded in knowledge base, it means that motherboard_ID dosenot be changed; on the other hand, the caution will be sent out.

Constructing rules in SWRL:

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- Immediacy(?x) ∧ IDevice(?x, ?y) ∧ swrlb:equal(?y, "C:") ∧ IField(?x, ?z) ∧ swrlb:equal(?z, "% Free") ∧ IPage(?x, ?z) ∧ swrlb:equal(?z, "Logical Devices") ∧ IValue(?x, ?a) ∧ swrlb:lessThan(?a, 0.4) → has_Caution(?x, "check Drive C")
- motherboard_changed(?x) ∧ IIcon (?x, ?y) ∧ swrlb:equal(?y, 42) ∧ IPage (?x, ?z) ∧ swrlb:equal(?z, "Motherboard") ∧ IField (?x, ?a) ∧ swrlb:equal(?a, "Motherboard ID") ∧ IValue(?x, ?b) ∧ swrlb:notEqual (?b, "63-0100-000001-00101111-090805-Alviso$0AAAA000_BIOS DATE: 09/08/05 VER: 202") → has_Caution(?x, "Motherboard changed")

IV. Implementation Based on Semantic Web Development Environment

A. Creating Knowledge Base

The development environment is to carry out the analysis task as described in the preceding section. First of all, we create class hierarchy and add properties to classes. From the analysis of the information asset, we obtain 46 classes as the basic and main vocabulary, which forms a three-level class hierarchy. Part of it is shown in Figure 2(a). Then we add properties to each class in the class hierarchy. The properties are categorized into datatype and object properties, respectively, where the ranges of the former types are atomic data, while the latter are instances of classes. We obtain 25 datatype properties as shown in Figure 2(b).

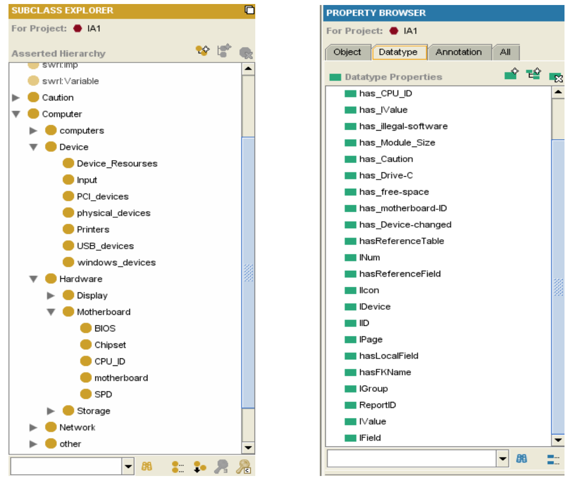

Figure 2. Class Hierarchy, and Datatype Properties

After the domain hierarchy is formed, we therefore can extract rules by analyzing the common audit situation. The condition and conclusion parts of rules are classes and properties selected from the domain hierarchy. We use protégé SWRL plug-in as the testing environment when we construct the rules. From the analysis on the common audit situation we can obtain audit rules in SWRL as shown in Figure 3.

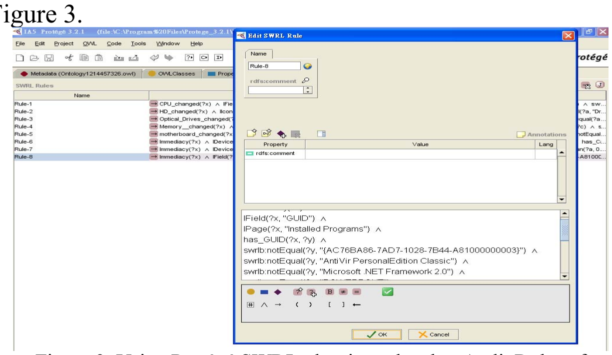

Figure 3. Using Protégé SWRL plug-in to develop Audit Rules of Information Assets

B. Services Based on the Metadata Layer

We have used Protégé as the development environment to create the knowledge base of audit rules of information assets as the metadata layer of the Semantic Web. We then conduct implementation of services based on the metadata layer. First of all, we employ the query interface of Protégé to test the conceptual search and semantic navigation on the content of the metadata layer. We choose the Class and Properties tabs for navigating the content of metadata layer according to the semantic structure. We click the Query tab of Protégé to perform conceptual search. On the query interface, we select the classes of interest first, and then further specify the constraints by filling relational expressions of properties. Figure 4 shows the result of a query instance for finding out that "Immediacy09 has the record that the one of legal installed software in information assets has an unique GUID." We use the Jess inference engine installed in the protégé environment to test the inference capabilities inherent in the SWRL rule base. First, user creates the metadata of an audit rule in information assets by using the ontology editing facility of Protégé. This metadata can be exported in RDF format that is used as the input to the SWRL inference engine. The inference engine loads in the knowledge base and SWRL rule base, and obtains new facts based on them as summarized in Figure 5. Taking for an example, there is an information asset with the following five property-value pairs: “IField: CPUID Revision”, “IPage: CPUID”, “IGroup: CPUID Properties”, “IIcon: 80”, “has_CPU_ID: 000006D8h”. With these property-value pairs, the inference engine tests the property-value with the audit rules and concludes that the caution is the type of “CPU_changed” , and it will be sent out. The result is shown in Figure 6.

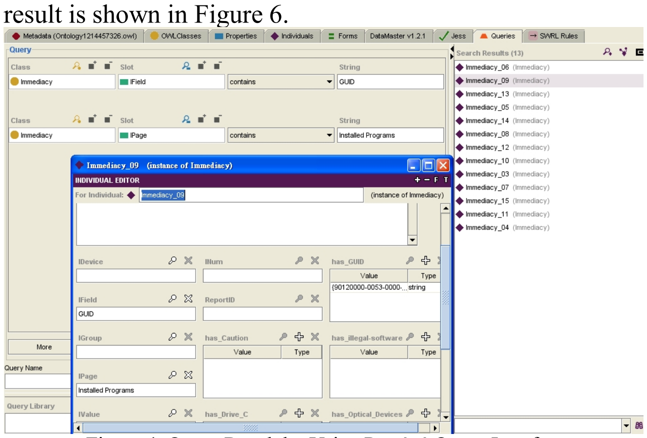

Figure 4. Query Result by Using Protégé Query Interface

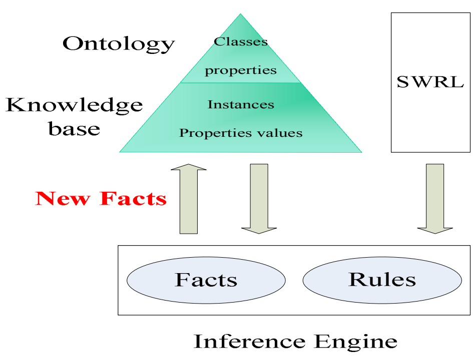

Figure 5. Architecture of SWRL Inference

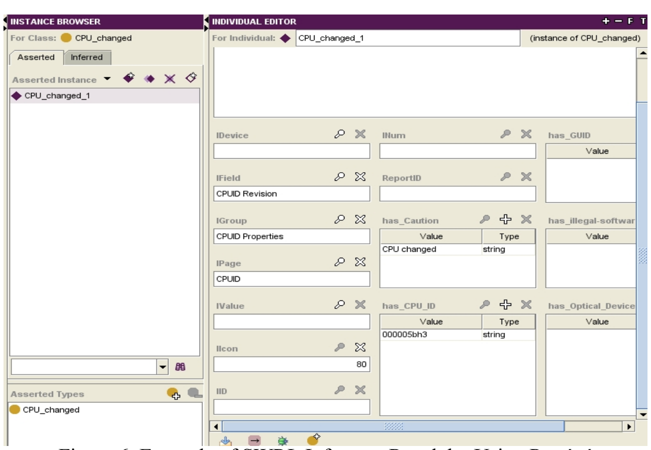

Figure 6. Example of SWRL Inference Result by Using Protégé

V. Conclusion

Our research establishes the domain knowledge of information assets to ontology and constructs the metadata of audit rules by the concept and technique of semantic web. The metadata is represented by using the standard Semantic Web architecture and is implemented by using open sources. The metadata consists of domain schema in OWL and instances in RDF, and rules in SWRL. Besides, the research depicts the architecture of information assets and common audit rules, and implements a model knowledge base system eventually. This research provides the ontology with inference-logic as kernel and combines with audit rules in SWRL to import into inference engine. The methodology not only help auditors to control and manage information assets in efficiency, but also can look up the information users need quickly and accurately. And then detecting or investigating the individual which is harmful to enterprise in time to make sure that all operations conform to security policy. Through the methodology which integrates the technique of semantic web with the collection of information assets records, it can raise the achievements of auditors and decrease the management cost of enterprise. Our research adopts OWL DL to present knowledge. It has more abundant descriptive vocabulary and hierarchical framework, and can present complete domain knowledge. Besides, it can infer based on class to produce recessive knowledge. In brief, it can make share and reuse of knowledge easier via the concept and technique of semantic web. In addition to the purposed audit rules and ontology of information assets, it also can apply to other domain knowledge and related rules, and then display the particularities of share and reuse of ontology.

Acknowledgment

This research has been supported by the National Science Council, Taiwan, under grant no. NSC96-2416-H-036-001.

References

[1] B. McBride. RDF Primer, W3C Recommendation 10 February 2004. Available at http://www.w3.org/TR/rdf-primer/

[2] D. L. McGuinness and F. van Harmelen (eds.). OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004. Available at http://www.w3.org/TR/owl-features/

[3] I. Horrocks, et al. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission 21 May 2004. Available at http://www.w3.org/Submission/SWRL/

[4] J. Broekstra, A. Kampman, F. van Harmelen. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In The Semantic Web – ISWC 2002, volume 2342 of Lecture Notes in Computer Science, pp. 54-68. Springer. 2002.

[5] Natalya F. Noy and Deborah L. McGuinness., “Ontology Development 101: A Guide to Creating Your First Ontology,” Stanford Medical Informatics Technical Report SMI-2001-0880, 2001. http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf

[6] Protégé project. Available at http://protege.stanford.edu/

[7] R. Benjamins, D. Fensel, S. Decker, and Gomez-Perez. KA2: building ontologies for the Internet: a mid term report. International Journal of Human Computer Studies. pp. 687-712, 1999.

[8] T. Berbers-Lee. Uniform Resource Identifier (URI): Generic Syntax. RFC3986. 2005. Available at http://gbiv.com/protocols/uri/rfc/rfc3986.html.