Versioning OWL Ontologies using Temporal Tags

Punam Bedi, and Sudeep Marwha

Abstract—Ontologies play an important role in semantic web applications and are often developed by different groups and continues to evolve over time. The knowledge in ontologies changes very rapidly that make the applications outdated if they continue to use old versions or unstable if they jump to new versions. Temporal frames using frame versioning and slot versioning are used to take care of dynamic nature of the ontologies. The paper proposes new tags and restructured OWL format enabling the applications to work with the old or new version of ontologies. Gene Ontology, a very dynamic ontology, has been used as a case study to explain the OWL Ontology with Temporal Tags.

Keywords—Frame and slot Versioning, OWL, Ontology Versioning, Semantic Web.

I. INTRODUCTION

WITH the increasing popularity of the semantic web among the researchers, the semantic web applications are being conceived in the variety of domains in different streams of the society. For the development of semantic web applications, knowledge representation in the form of ontologies becomes the core issue of research. Researchers worked together to create standards, tools and languages to speed up the development of ontologies. A single ontology is often developed by different groups, may be separated geographically and continue to evolve over time. Due to the distributed and dynamic nature of the web, the knowledge component of these applications changes more rapidly than the other components of the applications. Ontologies being the knowledge representation technique in semantic web applications need to handle this change. Research in ontology engineering nowadays is more focused on ontology management problems rather than ontology creation and ontology formalism. Noy [4] [5] [10] has listed and classified the core issues in the ontology management. These include maintain libraries of ontologies, import and reuse ontologies, translate ontologies from one formalism to another, provide support for ontology versioning, specify transformation rules between different ontologies and versions of the same ontology, merge ontologies, align and map between ontologies, extract semantically independent parts of the ontology, support inference across multiple ontologies, and support query across multiple ontologies.

In [9] ontology has been defined with the versioning and backward compatibility by keeping each version in different file. In a business world, organizations tend to keep the applications running for years once they reach to stable version as they have invested huge money in building and deploying them. Web is dynamic, huge and a vast resource with millions of applications and people interacting with each other. Keeping multiple versions of a single ontology simultaneously and allowing the dependent applications to choose from the version they depend upon causes some serious problems. First, there will be many of the version files existing at a single moment and they will grow in number very rapidly. Second, it creates confusion to choose from the different versions for the dependent applications. Third, someone has to put resources in terms of money, space, time and efforts to maintain all these versions. Fourth, dependent application requires at least some programming effort each time the application is synchronized with the latest version of ontology. Synchronization with latest version is needed as the knowledgebase applications can not provide good results if they do not update their knowledgebase.

OntoView [8] includes structural comparison of source ontologies. The system identifies different types of changes between versions of the same concept and allows its users to enlist a conceptual description of how the concept has changed. However, if the concept name has changed, OntoView does not attempt to determine that the concept with the new name is just an image of old concept.

This paper focuses on the problems related with ontology versioning, ontology aligning and mapping and specifying transformation rules between different versions of the same ontology. These problems need to address issues such as identifying ontology versions, specifying explicitly logs of changes between versions and determining a set of additional ontology changes that each change specified by the user incurs. Identifying an ontology version is required as one needs to fetch a particular version that suit the requirements of an application or to analyze the differences between a set of versions against the novelty of concepts or usage of different versions. In our approach, we propose that identification of different versions will be carried out very easily as all these
versions are stored in a single file and proposed changes in RDF attributes will ensure that at any point of time the particular OWL ontology version will be generated on the fly for the desired user or application. This is possible by deploying a web service that extracts a particular version from the temporal tagged ontology rather than keeping a static OWL file at the server. The service sends the extracted ontology to the requesting semantic web application. Given the decentralized nature of ontology development, logs of changes may not always be available so it becomes all the more difficult to align and map different concepts or determining additional ontology changes that each change incurs in the ontology. Introduction of new tags and the modified structure of the OWL file tackle this problem very effectively. Comparing our approach with the existing systems of managing ontology versioning such as OntoView, we found our approach becomes very easy and efficient as it introduces the changes in the RDF attributes that forces the ontology developers to take care of these issues while developing and updating the ontology rather than the identifying changes after the development and release of each version.

Semantic web applications require a framework [5] that is independent of the ontology versioning and do not need previous ontology versions to exist simultaneously along with the latest version.

We discuss our approach with the very popular Gene Ontology [2] as a case study. In the current research environment, where new genome sequences are being rapidly generated, and where comparative genome analysis requires the integration of data from multiple sources, it is especially germane to provide rigorous ontologies that can be shared by the community. We have chosen the Gene Ontology because it is perhaps the most dynamic ontology with over few thousands of new additions each year. The Gene Ontology consortium releases a new version of the ontology monthly. Daily versions in OWL format can also be downloaded which are generated from the Concurrent Versions System (CVS).

The rest of this paper is organized as follows: section 2 discusses the knowledge representation using frames, temporal frames and slot versioning. Section 3 describes knowledge representation for the web using Ontologies, and versioning of OWL ontologies. Section 4 deals with the problem of ontology versioning and our approach of using temporal tagged ontologies by extending OWL to solve it. A case study of Gene Ontology is given in section 5 in which transformation of existing OWL Gene Ontology, to the temporal tagged OWL Gene Ontology is described. Section 6 concludes the paper with merits of the approach.

II. KNOWLEDGE REPRESENTATION USING FRAMES

A frame is basically a structure for holding various types of knowledge. Frames are given names, with the presumption that the knowledge contained within a particular frame is in some way interrelated. Many authors and speakers may refer to frame-like structures as units, objects, concepts, schemas or entities. Frames are structured ways of representing descriptive information & they provide a natural mapping for the kind of knowledge that is centred around one concept or object. The organization allows efficient searching because there is immediate access to relevant information. Frame structure provides a natural method for representing hierarchies of information, thus allowing the inheritance of values. Although frames were conceived independently of the object-oriented paradigm they are in fact consistent with it, and provide an excellent demonstration of their power. Indeed frames are capable of representing both specific and general knowledge, and are capable of accommodating both descriptive and prescriptive computations.

Frame Based Systems have many advantages like immediate access to relevant information, easy to include default information and detect missing values etc. Frames can be used easily with Production rules, thereby facilitating partitioning, indexing and organizing production rules of a system. The implicit hierarchy available in frame taxonomies also permits hierarchical segmentation of rules. All these merits of the frame allow it to be the basis for knowledge representation technique called Ontology for the web.

A. Temporal Frame Representation

The inherent capability of Frame Based System for structuring knowledge has been the motivation for enhancing its capabilities by adding various dimensions to it. In some situations, it is not appropriate to discard the old information. The temporal (time) dimension [3] is added keeping these situations in mind. Knowledge in these temporal frame systems has the period of validity, which is attached to either frames or its slots or both. If the period of validity is attached to frames, it is called frame versioning and if attached to its slots, it is called slot versioning. The Tframe system [1] is based on slot versioning in contrast to frame versioning in order to achieve efficiency in terms of time and space. The system based on slot versioning is better because frame versioning approach has high degree of redundancy owing to duplication of the entire frame, especially when the changed portion is relatively small compared with the unchanged portion.

III. KNOWLEDGE REPRESENTATION USING ONTOLOGIES

Ontologies are structures or models of known knowledge. Ontology is a partial and explicit account of a conceptualization. The degree of specification of the conceptualization, which underlies the language used by a particular knowledge base, varies in dependence of our purposes. An ontological commitment is thus a partial semantic account of the intended conceptualization. In OWL ontology, concepts are arranged in hierarchical format with each concept is represented by a node in the hierarchy. An OWL class having various properties and relationships with the other classes represents each node. Relating it with the frames and slots, a class in ontology is based on the frame and
A. Versioning of OWL Ontologies

Version Support in OWL is very limited and can be described at two levels viz., at ontology level and at class or property level. At ontology level, owl:priorVersion property links to a previous version of the ontology being defined and can be used to track the version history of an ontology. Ontology versions may not be compatible with each other. Within an owl:Ontology element, we use the tags owl:backwardCompatibleWith and owl:incompatibleWith to indicate compatibility or the lack thereof with previous ontology versions. At class or property level, owl:versionInfo is provided. As opposed to the previous three tags, the object of owl:versionInfo is a literal and the tag can be used to annotate classes and properties in addition to ontologies. For many purposes, doing version tracking at the granularity of an entire ontology is not enough. Maintainers may wish to keep version information for classes, properties, and individuals - and even that may not be sufficient. The incremental nature of class expressions in OWL implies that one ontology may add restrictions to a (named) class defined in another ontology, and these additional restrictions themselves may require version information. OWL Full provides the expressive power to make any sort of assertion about a class, i.e. that it is an instance of another class, or that it (and not its instances) has a property and a value for that property. This framework can be used to build ontology of classes and properties for tracking version information. The OWL namespace includes two predefined classes that can be used for this purpose: owl:DeprecatedClass and owl:DeprecatedProperty. They are intended to indicate that the class or property will likely be changing in an incompatible manner in a forthcoming release. The versions are mostly maintained by version control systems such as CVS that are developed for keeping track of the code of traditional software.

IV. VERSIONING OWL ONTOLOGIES USING TEMPORAL TAGS

OWL class contains rdf:Id or rdf:Resource tag and all the properties of the class has rdf:Resource tag for naming and identification which in turn becomes their globally unique identification when seen in combination with the URI of the XML document. Relating the OWL with frames and slots, we introduce the concept of frame and slot versioning in OWL. The existing standard of OWL contains only an annotation property for capturing the version of the ontology. When the value of a property of a class has changed or name of the property has changed between two versions, the slot versioning is used to capture the change. According to slot versioning, only the version of changed property is created and inserted above the existing latest version in the same OWL file. When the class name or the intrinsic attribute of the class has changed then we use the frame versioning and the whole is inserted above the existing version in the same OWL file. The unique identity of different concepts are maintained by modifying the format of rdf:Id and rdf:Resource value. The values of these attributes are appended to the corresponding version number separated by the delimiter @. We propose to introduce two new tags in the underlying RDF viz. rdf:Validity and rdf:Timestamp. These two attributes will be used along with the rdf:Id and rdf:Resource to identify the changes that occur in subsequent versions of the ontology. The rdf:Validity attribute will have one of the four possible values viz. Always True, Latest, Past or Deleted. The rdf:Timestamp attribute will be used to represent the time when the concept has changed. It will be of the format YYYY-MM-DDThh:mm:ssTZD as specified in ISO 8601. These two attributes will be used to identify the same concepts of the different versions of the ontology. At any point of time the concepts of latest version of the ontology will have the value of rdf:Validity set to Always True or Latest and the value of rdf:Timestamp attribute is set to release date and time. If rdf:Validity is Always True then it means the concept is marked as universal truth and can not be changed in new versions. The Always True value is for the basic concepts, changing which the ontology itself become so much different than the previous version that it will be better to term it as new ontology. If the concept has value Latest then that it will be treated as the concept of the latest version of the ontology. Latest value should be unique among the different versions of the same concept. The value of Rdf:Timestamp attribute will be used to extract a particular version on request of the user/application. The changed concept will be added as a new line having all other attribute same as the older one except for the rdf:Id / rdf:Resource, rdf:Timestamp, rdf:Validity and the updated information. The rdf:Timestamp will contain a new date and time while rdf:Validity will have Latest as its value. The rdf:Resource and rdf:Id will be appended by a @ and the number of the version. This is necessary to maintain the uniqueness of the resource or id. The same name before the @ signifies that the new added concept is essentially the next version of the existing concept. The ontology will still keep all the concepts including the latest as well as the past. The new concepts in the ontology can be added by choosing the appropriate location with the Latest as the value of rdf:Validity and current date and time as the value for rdf:Timestamp. The concepts that are obsolete and need to be removed can be deleted from the latest version only by changing the value of the rdf:Validity as Deleted. The concept will remain there in the older version with its existing values of rdf:Validity, rdf:Timestamp and rdf:Resource and rdf:Id tags. The non deletion of the deleted concept of the latest version in the older versions will keep the older versions compatible with the applications that are running on it. The concepts that are added after the latest release will contain the date and time at which they are incorporated in the ontology. This value will be modified to the release date and time when the next version will be released. By that time the concepts will be present in the ontology as valid changes for the next version.

International Scholarly and Scientific Research & Innovation 1(3) 2007 688 scholar.waset.org/1999.4/12579
The introduced tags rdf:Validity and rdf:Timestamp and changed format of the ID and Resource tags are required as it becomes very efficient to extract the older versions from the master ontology. The OWL ontologies with temporal tags can be created from the OWL ontologies using an algorithm given below.

Algorithm: Creation of OWL Ontologies with temporal tags

Phase-I : Transforming OWL Ontology to OWL Ontology with Temporal Tags
Input : OWL Ontology
Read an existing OWL ontology
Insert the rdf:Validity and rdf:Timestamp tags in every rdf:Resource or rdf:Id statement. Value of rdf:validity is set to “Latest” and value of rdf:Timestamp is set to current date and time.
Get the version number from the user or search it in version annotation property and save it in Current_Version.
Append the Current_Version after the value of rdf:Id and rdf:Resource tags with # as delimiter.
Output the ontology as a new file.

Updating in the ontology concepts mean change in the class names, slot names, slot values, addition of classes, deletion of classes, addition of properties, deletion of classes, deletion of properties, change in the value of properties. Change in classes position in the hierarchy with respect to the other classes will be treated independently by addition of class at new position and deletion of class at the present position. The statement in this algorithm means the block of statements if it is a container statement like class and a single line statement if it is just a property. The rdfs:Comment, rdfs:Label and other tags having neither rdf:Id nor rdf:Resource attribute are taken from the latest version and older values are deleted as they don’t affect the semantics of the ontology.

Phase-II : Updating the Temporal Tagged OWL Ontologies
Input: Temporal tagged OWL Ontology from phase-I
For any subsequent change in the concept in Ontology do
Identify the concept and its location that has changed
Check the rdf:Validity tag value for each version of the concept
If it is “Always True” then
Exit and print not a valid change
If it is “Latest” then
If the task is to update the concept then
Change the value of rdf:Validity to “Past”
Insert the new statement describing the updated concept before the current statement
Set rdf:Validity of new statement to “Latest” and rdf:Timestamp set to current date and time.
Append the current date and time after value of the rdf:Id or rdf:Resource of new statement with # as delimiter.
Exit
If the task is to add a new concept then
Insert the new concept statement/statements at the desired location with its rdf:Id or rdf:Resource appended with the current date and time.
Set the rdf:Validity to “Latest” and rdf:Timestamp to current date and time.
Exit
If the task is to delete the concept then
Set the rdf:Validity to “Deleted” and don’t change the rdf:Timestamp value.
Exit
If it is “Past” then
Exit
If it is Deleted” then
Exit
End loop
End Loop

The algorithm matches the value of the rdf:Validity tag, rdf:Timestamp and version number appended after the resource with the information given for requested version. The version can be extracted using the given version number as well as the latest version up till the given date and time.

V. TRANSFORMING OWL GENE ONTOLOGY TO OWL GENE ONTOLOGY WITH TEMPORAL TAGS – A CASE STUDY

The Gene Ontology (GO) [2] project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO collaborators are developing three structured, ontologies that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.

A. Gene Ontology with Temporal Tags

The concept of temporal tagged ontologies is applied to the Gene Ontology to make it version independent and a particular ontology version can be generated on the fly depending upon the requirement of the user or application. We applied the first phase of creation of OWL ontology using temporal tags algorithm for transforming the Gene Ontology. According to algorithm, the ontology is read, construct by construct from top to bottom and transformation happens for each construct containing either the rdf:Id and rdf:Resource tag. The rdf:Validity and rdf:Timestamp tags are introduced in them and value of rdf:Resource and rdf:Id are modified by appending version number to them. The sample output is shown in table 1 wherein first column contains original Gene Ontology and second contains the transformed one. At the end of the first phase the value of rdf:Validity is set to Latest for all the constructs and a default version number 1.0.0 is appended to the rdf:Id and rdf:Resource values. Current date and time is set for the rdf:Timestamp.

The output of phase I was fed to the phase II algorithm for maintaining any subsequent updating of the concepts. Table 2 shows a temporal OWL ontology class that has changed over time. The class has got an another rdf:Resource and owl:Restriction. The addition of these causes a change in the structure of the class, due to the change in the underlying
frame structure. These types of updates need frame versioning and are very frequent in the Gene Ontology as different users for the use of new applications enrich the existing classes. The updated and old classes exist simultaneously in the ontology for the seamless integration of temporal tagged OWL ontology with the existing applications.

### Table I

<table>
<thead>
<tr>
<th>OWL Gene Ontology</th>
<th>Transformed OWL Gene Ontology with Temporal Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;owl:Class rdf:ID=&quot;GO_0005952&quot;&gt;</td>
<td>&lt;owl:Class rdf:ID=&quot;GO_0005952@1.0.0&quot; rdf:Validity=&quot;Latest&quot; rdf:Timestamp=&quot;01:02:2006 #05:05:00&quot;/&gt;</td>
</tr>
<tr>
<td><a href="">rdfs:label</a>cAMP-dependent protein kinase complex&lt;/rdfs:label&gt;</td>
<td><a href="">rdfs:label</a>cAMP-dependent protein kinase complex&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td>&lt;!-- unlocalized protein complex --&gt;</td>
<td>&lt;!-- unlocalized protein complex --&gt;</td>
</tr>
<tr>
<td>&lt;rdfs:subClassOf rdf:resource=&quot;#GO_0005941&quot;/&gt;</td>
<td>&lt;rdfs:subClassOf rdf:resource=&quot;#GO_0005941@1.0.0&quot; rdf:Validity=&quot;Latest&quot; rdf:Timestamp=&quot;01:02:2006 #05:05:00&quot;/&gt;</td>
</tr>
</tbody>
</table>

At the end of the first phase the value of rdf:Validity is set to Latest for all the constructs and a default version number 1.0.0 is appended to the rdf:Id and rdf:Resource values. Current date and time is set for the rdf:Timestamp.

### Table II

<table>
<thead>
<tr>
<th>Phase II OWL Ontology with Temporal Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;owl:Class rdf:ID=&quot;GO_0005952@1.1.0&quot; rdf:Validity=&quot;Latest&quot; rdf:Timestamp=&quot;03:09:2006#10:30:00&quot;/&gt;</td>
</tr>
<tr>
<td><a href="">rdfs:label</a>cAMP-dependent protein kinase complex&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td>&lt;rdfs:comment rdf:resource=&quot;<a href="http://www.w3.org/2001/XMLSchema#string%22%3E">http://www.w3.org/2001/XMLSchema#string&quot;&gt;</a> An enzyme complex, composed of regulatory and catalytic............. &lt;/rdfs:comment&gt;</td>
</tr>
<tr>
<td>&lt;!-- protein complex --&gt;</td>
</tr>
<tr>
<td>&lt;rdfs:subClassOf rdf:resource=&quot;#GO_0005952@1.0.0&quot; rdf:Validity=&quot;Latest&quot; rdf:Timestamp=&quot;01:02:2006 #05:05:00&quot;/&gt;</td>
</tr>
</tbody>
</table>

Table III shows another class that has become obsolete in the new version. The class just contains rdfs:label and rdfs:comment in the new version but was having an owl:Restriction in the old version.

### Table III

<table>
<thead>
<tr>
<th>Phase II Output for an Obsolete Class with Frame Versioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;owl:Class rdf:ID=&quot;GO_0000067@1.1.0&quot; rdf:Validity=&quot;Latest&quot; rdf:Timestamp=&quot;03:03:2006#05:05:00&quot;/&gt;</td>
</tr>
<tr>
<td><a href="">rdfs:label</a>DNA replication and chromosome cycle&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td>&lt;rdfs:comment rdf:resource=&quot;<a href="http://www.w3.org/2001/XMLSchema#string%22%3E">http://www.w3.org/2001/XMLSchema#string&quot;&gt;</a> OBSOLETE (was not defined before being made obsolete). &lt;/rdfs:comment&gt;</td>
</tr>
<tr>
<td>&lt;!-- cell cycle --&gt;</td>
</tr>
<tr>
<td><a href="">owl:Restriction</a></td>
</tr>
<tr>
<td>&lt;owl:ObjectProperty rdf:about=&quot;#part_of&quot;/&gt;</td>
</tr>
<tr>
<td><a href="">owl:inProperty</a></td>
</tr>
<tr>
<td>&lt;owl:someValuesFrom rdf:resource=&quot;#GO_0000670@1.0.0&quot; rdf:Validity=&quot;Past&quot; rdf:Timestamp=&quot;03:09:2005#05:05:00&quot;/&gt;</td>
</tr>
<tr>
<td><a href="">owl:Restriction</a></td>
</tr>
<tr>
<td>&lt;!-- intracellular --&gt;</td>
</tr>
<tr>
<td><a href="">rdfs:subClassOf</a></td>
</tr>
<tr>
<td><a href="">owl:Class</a></td>
</tr>
</tbody>
</table>
Semantic web applications that are dependent on the old version will have problems due to the change in the OWL class but in case of Temporal tagged OWL class, the application will still continue to work fine. Fig. 1 shows a portion of the class hierarchy in Gene Ontology. The classes of the latest version are depicted in white and classes of the past version are in gray. The class GO_0007275@1.0.0 has a sub class GO_0009835@1.0.0 that has changed to GO_0009835@1.1.0 between the two versions taken in the case study. Due to the change, the sub classes of GO_0009835@1.0.0 should become the subclasses of the latest version GO_0009835@1.1.0. Since, the sub classes of GO_0009835@1.0.0 contains only a slot or resource that points to its base class, slot versioning is applied here and the rdfs:subClassOf tag is replicated with its resource pointing to the new version. The new version of resource is depicted in white and old version in gray, similar to the classes. The new version of the class GO_0009835@1.1.0 is appeared under another class GO_0048608@1.0.0. The dotted line in the figure depicts that there are many other classes present in between that are not shown in the figure. Other classes shown in Fig. 1 are not expanded to show their resources, as they have not changed between these versions.

One can judge from the code of temporal tagged ontology presented in Table I, Table II and Table III that the temporal tagged ontologies will be having larger size due to the presence of extra tags and due to the repetition of constructs for each version. The code has also become less readable because of the replication. We argue that both the size and readability of OWL files are having very low effect on the performance of OWL ontologies because these ontologies are primarily designed for the semantic web applications usage and particular OWL ontology version can always be generated back from them.

CONCLUSION

In this paper an approach providing consistent and dynamic ontology support to knowledgebase applications is presented. The approach combines the concepts of temporal frame and slot versioning with the ontology to create temporal tagged ontologies with embedded versioning. We also propose to enhance the existing OWL to enable the creation of temporal tagged OWL ontologies. The two new tags i.e. rdf:Validity and rdf:Timestamp are introduced and a scheme is presented for the value of the rdf:Id and rdf:Resource tags for making the temporal tagged ontologies consistent with the non-temporal ontologies. OWL temporal tagged ontologies are designed for the machine consumption and not for human readability as it includes repetition of concepts with even minor changes. These ontologies will be of great use for the semantic web applications and will make them independent of the ontology versions. This means that the semantic web applications developer can devote more attention to the application logic and agent behaviors development without worrying the ongoing changes in the domain knowledge. The dynamic behavior of the temporal tagged ontologies will allow the application to use the most recent concepts of the ontology without sacrificing the stability of the application.

REFERENCES