

SEMANTIC ANNOTATION AND REVISION CONTROL

Kiavash Bahreini and Atilla Elci

*Department of Computer Engineering, and Internet Technologies Research Centre
Eastern Mediterranean University, T.R.N.C., Famagusta, via Mersin 10, Turkey*

Keywords: Annotation, Metadata, OWL, Revision Control.

Abstract: Software engineers and programmers will probably find themselves needing to manage multiple versions of their software. This entails, among others, managing source codes, inserting metadata tags for annotation, tracing source changes from current to previous versions, checking respective change logs, retrieving different versions of the source code, etc. The issues are more pronounced for software teams and especially those working in distributed development environments. Similar issues are observed in the case of dealing with OWL files and other enterprise systems documentation resources. It is noted that, although currently not being practiced, ontology-based annotation techniques in revision control can be influential in surmounting many of the problems associated with such issues. These issues and related new approaches on revision control are considered in this paper. We introduce a novel revision control approach based on semantic ontology annotation in distributed environments.

1 INTRODUCTION

This paper concerns source code control, making separate ontologies in distributed environments using OWL (OWL 2004), applying annotation techniques into semantic ontology making, and security model for data and documents sharing based on multilevel security in enterprise systems.

First issue is source code control (SCC) which is only one step in software configuration management (CCC 1984). SCC is especially very important for any software development team working in distributed environment (Sink 2004). Distributed revision control is the other hot topic which software development team deal with the problem of controlling shared documents in enterprise systems. Distributed revision control takes a peer-to-peer approach, as opposed to the client-server approach of centralized systems. Rather than a single, central repository on which clients synchronize, each peer's working copy of the code base is a bona-fide repository (Wheeler 2004).

An open system of distributed revision control is characterized by its support for independent branches, and its heavy reliance on merge operations.

Next important issue, managing separate ontologies in distributed environments using OWL,

involves the related information resources placed in distributed systems on the Web. So, for providing seamless services at any distance, efficient management of the distributed information resources is required (Kim et al 2007).

It is useful to keep some general questions in mind while reading this paper. How can we manage shared documents or resources among individual or distributed ontologies? How can we process text based documents? How can we investigate essential lines which have been changed in text? How can we check the version of source code via ontologies? How can we annotate ontologies using version control? How can we merge contents of ontologies? How can we run queries over ontologies to retrieve some new changes inside source codes? Would it be better to put all of the content of changed-source code inside ontologies or is it better to have just the lines which are changed as compared with old version.

In the Semantic Web, the dispensability will be increased. Among languages for describing metadata and ontology, only OWL can support the dispensability of Web; (Kim et al. 2007) so in this research, we considered using OWL as a semantic language in our distributed ontologies to annotate documents.

Next issue which has been considered for years is talking about metadata based ontology and

one founded answer for this problem was annotation. Annotation is defined as the extra information asserted at a particular point in a document or other piece of information (Annotation 2007). Offering a way to enable semantic annotations that could be easily organized and found is one of the advantages of semantic web (Pereira et al 2006). For convenience, annotations are represented as an extension of the ontology, but they could be implemented by any other means, as they are not a proper part of the domain knowledge (Castells et al 2007). Annotation can be used to overcoming two of the major challenges of the semantic Web initiative which are availability of semantic Web content, and the challenge of ontology based information retrieval (Abrahams 2005). In our approach, annotation is used as a high level tier to work with OWL files on top of the source code files to handle metadata over text (source code), to hold versioning of codes, and to retrieve change history of data.

In this paper, we consider features of revision control for supporting in distributed environment on the semantic web, repository, text processing. Then, we classify features of OWL and ontology for supporting distributed environment on the Web and consider of metadata in distributed systems. Then, we consider, semantic annotation, and finally conclusion and future work.

2 REVISION CONTROL

Several different forms of version controlling have been pursued to solve problems faster by doing slice controlling in the source codes. The main goal of those forms can be described by just relating changes to the metadata files. The first major division of kinds of version controlling is based on revision control over any changed items versus revision control over the whole content of the documents. There is probably a better understanding of the changeable analysis source code of transformation of data inside the source codes than there is of full of control transformations, which is an old technique in the source code control issue.

2.1 Repository

Repository and working folder are two terms in revision control, in which, repository is a place to store the source code and a working folder is a place which each individual developer works on it. All version control systems have to solve the same fundamental problem: how will the system allow

users to share information, but prevent them from accidentally stepping on each other's feet? It's all too easy for users to accidentally overwrite each other's changes in the repository (Ben Collins et al 2007).

2.2 Distributed Revision Control

Distributed source controlling approaches to discovery would require remote methods are regularly synchronized. Because of distributed revision control tasks involve comparing several sources, analyzing and understanding relations between their elements can be cognitively difficult. In this case, we strongly suggest taking advantage of using semantic metadata files.

In distributed revision control model on the Internet, there exist many aspect like: repository replication, repository backup, access controls, supporting multiple repository access methods, customizing subversion experience, branching and merging, entries file, metadata files etc.

2.3 Text Processing

Text processing is an important task in revision control. Our aim in this investigation did not involve making of metadata using text processing algorithm to retrieve source code changes. Thus, it can be concluded that known text search processing would suit this goal. This subject has been discussed in great depth in others' research.

3 ONTOLOGY AND METADATA

Metadata and ontology will play important roles for advanced information retrieval systems. Metadata is data about data describing the content, quality, condition, and other characteristics of data. And metadata can represent semantic relations between information resources. Ontology defines vocabulary and represents relations between terms for a specific domain (Kim et al 2007).

OWL has rich expressive power and is considered as the next standard language for describing metadata and ontology. In the Semantic Web, the dispensability will be increased. Among languages for describing metadata and ontology, only OWL can support the dispensability of Web (Kim et al 2007).

All ontologies in our model are distributed in the network. So, for providing seamless services without concern for distance, integration of distributed ontologies is required. If we want to get

programmers codes, we should use integrating techniques over ontologies on the network. Special relations between classes or properties in the ontologies should be expressed. Some features of OWL are used to support the integration of distributed ontologies are: `disjointWith`, `equivalentClass`, and `equivalentProperty`. Many writers (e.g. YounHee Kim, YongWook Kim) modified the storage part of Sesame to support distributed ontologies written in OWL. Their structure consists of four tables on RDB. For further info the reader is directed to Kim et al, 2007 paper.

4 DATA INTEGRATION

One of the greatest challenges facing the distributed revision control today is the integration of data and resources. Two approaches for integrating applications are commonly used today: point-to-point and services bus integration (Microsoft Journal 2006a). We simply can expand the above approaches to realize our goal of data integration for revision control over enterprise systems.

In the first approach, one direct link is created to establish a connection between two nodes while message passing is used in the second approach. When a new node add to enterprise architecture, managing a complete integration between nodes exponentially increase in the cost of each node in turn (Microsoft Journal 2006a).

5 SEMANTIC ANNOTATION

Semantic annotation is a specific metadata generation and usage schema aiming to enable new information access methods and to enhance existing ones (SWT 2006). Annotation is extra information asserted with a particular point in a document or other piece of information (Annotation 2007).

In our approach, either we use annotation over OWL files or source code we can apply some metadata over codes already tagged or make new tags for any other changes in source files. Some highlighting text and posting sticky notes on our metadata documents or web pages and share these with other people would be very useful. We simply can insert many tags inside our metadata files between these two tags `<owl:AnnotationProperty rdf:ID="">` and `</owl:AnnotationProperty>` which were declared to use in OWL files to annotate data. Some criteria should apply in OWL files to use

annotation in our model in source codes. First of all, author(s) information e.g. name and surname, then many fields like date and time of implementation of document, line number, changed line numbers, size of the file before and after changes, owner of file, number of history records exist in OWL files related to this file etc. The sample code lines below are prepared to show in which way we are able to apply annotation tags in our model.

```
...
<owl:AnnotationProperty rdf:ID="Smith">
<rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#
DatatypeProperty"/>
<rdfs:domain
rdf:resource="#LowLevelProgrammer"/>
<rdfs:range>
<owl:DataRange>
<owl:oneOf rdf:parseType="Resource">
<rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema
ma:string">Smith</rdf:first>
<rdf:rest
rdf:resource="http://www.w3.org/1999/02/22-
rdf-syntax-ns#nil"/>
</owl:oneOf>
</owl:DataRange>
</rdfs:range>
</owl:AnnotationProperty>
...
```

In a nutshell, semantic annotation is about assigning to entities and relations in the text links to their semantic descriptions in ontology. Most importantly, automatic semantic annotation enables many new applications: highlighting, semantic search, categorisation, generation of more advanced metadata, smooth traversal between unstructured text and formal knowledge. Semantic annotation is applicable to any kind of content-web pages, regular (nonweb) documents, text fields in databases, video, etc (SWT 2006). Since using OWL files in semantic web environment are going to be standard and approximately all applications like semantic web programs, semantic web sites etc are going to use them, so we prefer introducing OWL files as our standard and use them as metadata and annotation.

6 CONCLUSIONS

Investigating the security impact of semantically-enhanced OWL-based programs opens a new research area in OWL database security. Ontologies are used extensively in XML-based applications to improve data exchange in decentralized environments. We discussed that these new technologies might lead to undesirable data disclosure. Ontologies and OWL-based tools can

facilitate inference attacks on large, publicly available XML documents.

We proposed some theoretical and practical ways on source code control based on semantic web, annotation techniques, and use of OWL as metadata file via secure connection in enterprise architecture. This approach uses an ontology-aided inference process to identify ontology equivalent information with inconsistent security classification. We represented some solutions to build a system which is able to monitor users' source code, capable of text processing, annotate metadata in OWL files in distributed environments.

As it is mentioned above, our approach helps software architects, project managers, programmers, developers, etc to obtain a secure distributed control over source codes and annotate the whole content of documents in the best manner.

7 FUTURE WORK

Designing and implementing a source code control which can be called knowledge-based SCC (KBSCC) based on semantic distributed ontologies annotation using enhanced multilevel security in enterprise systems is the first goal in the future. Semantic web is going to define the next generation of web, so grafting some ideas from current applications such as web-based programs to the semantic web is useful for future. Making a secure system, and using multiagents or service-oriented architecture are three important issues on the distributed networks; hence, using multiagents approach or SOA for implementing semantic web-based programs via secure connections is our next target to establish a high level and stable system.

REFERENCES

- CCC, (1984), Change and Configuration Control. IEEE Software, July 1984, Volume: 1, Issue:3, On page(s): 112a-112a, Digital Object Identifier: 10.1109/MS.1984.234733.
- Sink, E. (2004), What is source control, http://www.ericssink.com/scm/scm_intro.html .August 26, 2004.
- Wheeler, David A. (2004). Comments on Open Source Software / Free Software (OSS/FS) Software Configuration Management (SCM) Systems. , April 10, 2004; lightly revised May 18, 2005, Retrieved on 2007-05-08, <http://www.dwheeler.com/essays/scm.html>.
- OWL. (2004), Web Ontology Language; W3C Recommendation 10 February 2004. <http://www.w3.org/TR/owl-features/>. Accessed October 11, 2007.
- Kim, YounHee, YongWook Kim, ByungGon Kim, and HaeChull Lim. (2007), IEEE 2007, Management System for OWL Documents in Distributed Environment on the Semantic Web. Advanced Communication Technology, The 9th International Conference. 12-14 Feb. 2007 Volume: 2, On page(s): 1216-1220 Location: Gangwon-Do, ISSN: 1738-9445 ISBN: 978-89-5519-131-8 INSPEC Accession Number: 9551150 Digital Object Identifier: 10.1109/ICACT.2007.358577 Posted online: 2007-05-07 11:28:34.0.
- Annotation. (2007), 9 October 2007, Wikipedia. <http://en.wikipedia.org/wiki/Annotation>
- Pereira Rui G. and Mário M. Freire. (2006), IEEE 2006, SWedt: A Semantic Web Editor Integrating Ontologies and Semantic Annotations with Resource Description Framework. Telecommunications. (2006). AICT-ICIW '06. International Conference on Internet and Web Applications and Services/Advanced International Conference on Publication Date: 19-25 Feb. 2006, On page(s): 200- 200, ISBN: 0-7695-2522-9, Digital Object Identifier: 10.1109/AICT-ICIW.2006.184.
- Castells Pablo, Miriam Fernandez, and David Vallet. (2007), IEEE 2007, An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. Knowledge and Data Engineering, Publication Date: Feb. 2007, Volume: 19, Issue: 2, On page(s): 261-272, Location: Los Angeles, CA, USA, ISSN: 1041-4347, INSPEC Accession Number: 9317010, Digital Object Identifier: 10.1109/TKDE.2007.22.
- Abrahams, B. and Wei Dai. (2005), IEEE 2005, Architecture for automated annotation and ontology based querying of semantic Web resources. Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on Publication Date: 19-22 Sept. 2005 On page(s): 413-417, ISBN: 0-7695-2415-X, INSPEC Accession Number: 8747729, Digital Object Identifier: 10.1109/WI.2005.34.
- Ben Collins-Sussman, Brian W. Fitzpatrick, and C. Michael Pilato. (2007). Version Control with Subversion: For Subversion 1.5: (Compiled from r2880). (pp. 2-7) <http://svnbook.red-bean.com/>.
- SWT. (2006), Semantic Web Technologies: Trends and Research in Ontology-based Systems, John Davies(Editor), Rudi Studer (Co-Editor), Paul Warren (Co-Editor), ISBN: 978-0-470-02596-3, Hardcover, 326 pages April 2006.
- Microsoft Journal. (2006a), The Architecture Journal issue 7, Generation Workflow. Workflow in application integration, <http://msdn2.microsoft.com/en-us/library/bb245667.aspx>, pp. 19-23, Kevin Francis, 2006.