

Provenance Linking using Bundles in OWL Ontology

Mrinal Pandey
AIIT
Amity Institute of
Information Technology
Lucknow, India

Rajiv Pandey
AIIT
Amity Institute of
Information Technology
Lucknow, India

ABSTRACT

The Semantic web envisioned by Tim Berners lee, provides for intelligent knowledge retrieval. Although in addition to knowledge retrieval there is also a need to make the knowledge so derived as trustworthy. This requires the incorporation of trust or provenance information in the semantic web. Provenance serves as a crucial factor in enhancing the trust ability of the semantic web. This paper aims at the creation of trustable semantic web by creating provenance assertions and provides for verifying the trust ability of these assertions by providing provenance of provenance descriptions for the same. This is shown using Bundles a special data structure required for the linking of provenance bundles. Also, we have tried to illustrate how the provenance descriptions created by one application can be effectively manipulated by other application by the use of these bundles

General Terms

Semantic Web; Provenance; Ontology; Bundles; MentionOf

Keywords

Semantic Web; Provenance; Ontology; Bundles; MentionOf..

1. INTRODUCTION

The semantic web also known as intelligent web, allows for effective and intelligent knowledge retrieval. It provides for technologies and standards that enable the world wide web to be used by both man and machines. The web contains huge amount of data but does not provide for mechanisms for data to be understood by machines. Semantic web however endeavors to do the same. Semantic web aims to solve the problem of adding well-defined meaning to the current Web and aims to device procedures that can be incorporated to the current web so as to add meaning to the information thereby enabling machines to make use of this information so as to achieve interoperability and cooperation with people. To achieve this purpose, it not only requires languages for expressing machine-understandable meta-information along with terminologies like namespaces or ontologies for making the metadata available, but also requires tools and architectures for the same. Also these technologies and languages should be such that they provide for specific syntax and semantics which enables the meta information to be transferred or easily interchanged amongst applications and be easily understood by both man and machines. Also there is an equal need for applications that allow the human users to use the semantic web easily[1]. Mechanisms that provide for annotating, searching, concatenating and retrieving semantic information, together with semantic languages, must also be present[2]. The Semantic Web aims at giving information a "well-defined meaning, better enabling computers and people to work in cooperation" [3] through the definition of ontologies. In order to achieve the objective of converting the

existing web of information to semantic information for easy searching, access and retrieval the following are required-

- Conception of an XML layer, that allows for representation of the structure of data
- Subsequent formation of an RDF layer, for describing the meaning of data
- Establishment of an Ontology for representing the data and the relationships among these data.[2]
- Enhancing trust ability of the meta-data by incorporating the element of trust into it using provenance.
- Consequently querying the metadata by using languages such as SPARQL.
- This can be clearly depicted in figure 1.

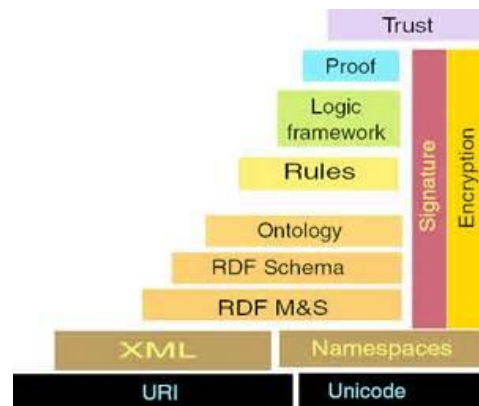


Fig 1. Semantic web Layered Cake[2]

The main advantages of Semantic Web can be summed up as follows[4]:

- Provides for the addition of a conceptual layer thereby allowing for the expansion and description of the current web.
- Provides for machine-readable, interpretable and editable web content
- Providing for the easy creation and searching of semantic assertions and annotations.
- Enhancing search mechanisms by using Ontologies;
- Enabling software agents to use smart data thereby allowing for the simple execution of sophisticated tasks
- Allowing for smooth communication between independent software agents

- Providing for trust implementation in the upper layers of the semantic web.

Semantic web thus is regarded as an extension of the current World Wide Web that provides for technology stacks and languages that help to generate information that has well defined meaning and can be easily used by both man and machines. The World Wide Web Consortium (W3C) recommends XML, XML Schema, RDF, RDF Schema and Web Ontology Language (OWL) as standards and tools for the implementation of the Semantic Web[5].

2. ONTOLOGIES IN SEMANTIC WEB

The Semantic Web aims at giving information a "well-defined meaning, better enabling computers and people to work in cooperation" [6] through the definition of ontologies. [3]. Ontologies are the key components that allow for effective knowledge representation in the Semantic Web. Ontologies provide for a data model that helps in the effective mapping and representation of a set of concepts related to a domain and also provide for proper representation of the relationships between that exist between those concepts within a domain. Ontology also helps in reasoning the relationships that exists between various objects present in a Domain [7]. Ontologies can be easily queried using languages such as SAPRQL. Contemporary research caters to the development of Semantic web tools; however, the work done using the tools, like RDF Editors [8]- [9] or Ontology Editors [10] [11] has been done on the lower layers of the semantic web. The upper layers of the semantic web which cater to trust and proof have still remained un-touched. In this paper therefore we have worked on the proof and the trust layer of semantic web by introducing the concept of provenance and incorporating it in our University people program ontology. We have also demonstrated how provenance descriptions can be effectively altered by various entities and agents and how provenance of provenance can be maintained in a document. Thereby aiding in the trust ability of the document

3. PROVENANCE AND ITS USE IN SEMANTIC WEB

Trust has gathered a great deal of focus with respect to semantic web in the past years. Researchers have proposed many trust Models [12][13][14], Belief Functions [15], Data Quality Algorithms, Trust Inference Algorithms [16], to assess the trust ability of semantic web documents. Many researchers have proposed the use of graph theory to assign values to trust and then consequently to infer the degree of trust from these values. But establishing trust model is difficult using only the approach of graph theory or "one-time I think this person is great" value. Here we therefore propose the concept of trust using the concept of provenance provided in the PROV-DM data model proposed by W3c[17]. Provenance also referred as pedigree or lineage refers to the information about the origins of data and consecutively the transformations applied to derive the data. Researchers have proposed different definitions of provenance with respect to different domains. Provenance is crucial as it can be used for assessing the quality of data, determining its audit trail, defining attribution and replication procedures for the data [18]. The paper demonstrates the creation of provenance using the concept of entity, agent and activity. Figure 2 gives a brief overview of entity, agent and activity's we have demonstrated the way to link provenance descriptions to each other using the concept of Bundles and MentionOf relation. For this purpose, we have used the tool named Prov-Store[23] for the same.

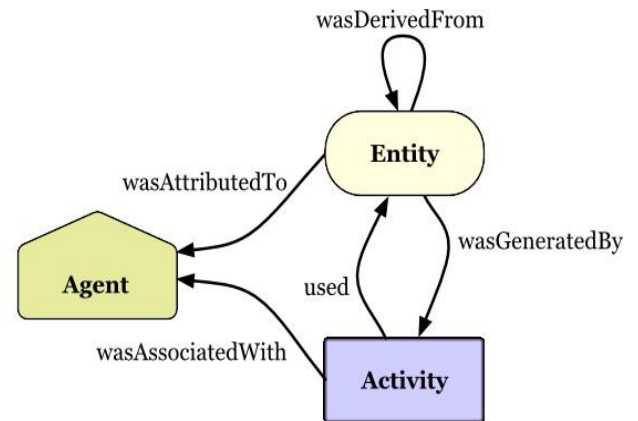


Fig 2.-Relationship between entity,agent and activity[24].

4. PROVENANCE IN UNIVERSITY PEOPLE PROGRAM ONTOLOGY

Ontologies are "vocabularies of representational terms, classes, relations, etc with agreed upon definitions ,in the form of human readable text and machine enforceable, declarative constraints on their well-formed use"[19]. An Ontology would only be effective on incorporating the element of trust. The element of trust requires the embedding of provenance assertions. The following section provides the code of University People Program Ontology constructed in Protégé using OWL. We have first shown the incorporation of simple provenance instances into the given Ontology and consecutively have made an attempt to link the provenance descriptions using Bundels and Mentionof relation. It has been aimed to map the Ontology individuals to the components of Entity, Activity and Agent of the PROV-DM for the purpose of embedding provenance information of the same. A snippet of the University People Program is as follows [20].

```

<ObjectPropertyAssertion>
<ObjectProperty IRI="#teaches"/>
<NamedIndividual IRI="#Rajiv"/>
<NamedIndividual IRI="#mobilecomputing"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
<ObjectProperty IRI="#isTaughtBy"/>
<NamedIndividual IRI="#mobileComputing"/>
<NamedIndividual IRI="#Rajiv"/>
</ObjectPropertyAssertion>
  
```

The above code states that Rajiv which is an individual is linked to mobile computing through the object property teaches & vice versa through object property 'isTaughtBy'[20]. Upon mapping the above OWL ontology snippet to Entity, Agent and activity, Rajiv is known to be an Agent who performs the activity of Authoring the end semester reports of various students for the subject Mobile computing, thereby generating the Entity known as report 1. The provenance assertions for the above scenario can be defined in the following form –

```

prefix ex <http://www.example.com/>
entity(ex:report1,[prov:type="report"],ex:version=1)
  
```

```
entity(ex:report2,[prov:type="report"],ex:version=2)
agent(ex:rajiv)
agent(ex:Mrinal)
activity(ex:ac1)
activity(ex:ac2)
wasAttributedTo(ex:report1,ex:rajiv)
wasAttributedTo(ex:report2,ex:Mrinal)
wasDerivedFrom(ex:report2,ex:report1)
wasGeneratedBy(ex:report1,ac1,2012-0524T10:30:00+01:00)
wasGeneratedBy(ex:report2,ac2,2012-0524T10:30:00+01:00)
endDocument[21]
```

However the questions regarding the authenticity of the reports so generated also need to be answered. This requires provenance of provenance to be maintained. Also the reports which are entities can be altered by other individuals which are natives of class professor and thus are themselves agents .The problem of expressing provenance of provenance can be catered to by using the bundles data structure provided by the PROV-DM data model. Bundles are named set of Provenance descriptions and allow the provenance of provenance to be defined as they can themselves be expressed and declared as entities[22]. However in scenarios where a particular entity is manipulated by multiple agents or multiple applications the provenance description created by ones party should be linked to provenance descriptions provided by another party. Thus the requirement of stitching provenance descriptions can be addressed by using a special relation called Mention which would provenance descriptions created in one bundle to be linked to provenance descriptions provided in other bundles. Thus in the following section we first describe the creation of bundles and further describe how these bundles can be related to each other.

5. EXPRESSING PROVENANCE OFPROVENNACE AND LINKINGACROSS BUNDLES

Bundles- Bundles provide for provenance of provenance description. Bundles also help in impelmenting the concepts of specialization and generalization by means of provenance in an Ontology. Concepts expressed as sepcialized and generalized versions allow for the linking and stiching of provenance generated by various agents through the concept of Mention.Below we have mapped the simple provenance definitions provided above as bundles thereby ensuring the trustability of the creator of the document.[23]

```
Document
prefix ex http://www.example.com/
prefix Mrinal <http://example.org/Mrinal/>
prefix Rajiv <http://example.org/Rajiv/>
wasGeneratedBy(Rajiv:bundle4,-,2012-024T10:30:00+01:00)
```

```
wasGeneratedBy(Mrinal:bundle5,-,2012-05
25T11:15:00+01:00)
entity(Rajiv:bundle4, [prov:type='prov:Bundle'])
entity(Mrinal:bundle5, [prov:type='prov:Bundle'])
agent(ex:Rajiv)
agent(ex:Mrinal)
wasAttributedTo(Mrinal:bundle5, ex:Mrinal)
wasAttributedTo(Rajiv:bundle4, ex:Rajiv)
bundle Rajiv:bundle4
prefix ex http://www.example.com/
prefix Mrinal <http://example.org/Mrinal/>
prefix Rajiv <http://example.org/Rajiv/>
entity(ex:report1, [prov:type="report", ex:version=1])
wasGeneratedBy(ex:report1, -, 2012-05-24T10:00:01+01:00)
endBundle
bundle Mrinal:bundle5
prefix ex <http://www.example.com/>
prefix Mrinal <http://example.org/Mrinal/>
prefix Rajiv <http://example.org/Rajiv/>
wasGeneratedBy(ex:report2, -, 2012-05-25T11:00:01+01:00)
entity(ex:report2, [prov:type="report", ex:version=2])
entity(ex:report1bis)
mentionOf(ex:report1bis, ex:report1, Rajiv:bundle4)
wasDerivedFrom(ex:report2, ex:report1bis, -, -, -)
endBundle
endDocument[23]
```

The above code creates two bundels named bundle 5 and bundle 4 both the bundles act on the common entity called report 1 which is the report that has been generated by the agent rajiv showing the progress of a student in the subject mobile computing however as the final report of a students can be generated by appending the progress of the student in other subjects. Thus the attributes of report1 can be appended or altered by other individuals of the class professor as happens in the case of the agent Mrinal which appends the contents of the report and thus generates the new report of the student namely report2.Also the use of Bundles provides for effective management and maintenance of provenance of provenance thus making the source generating the provenance trustworthy in itself. The figure provided below provides provenance descriptions of the various agents, activities and entities in the Ontology along with the document metrics of the same[23].

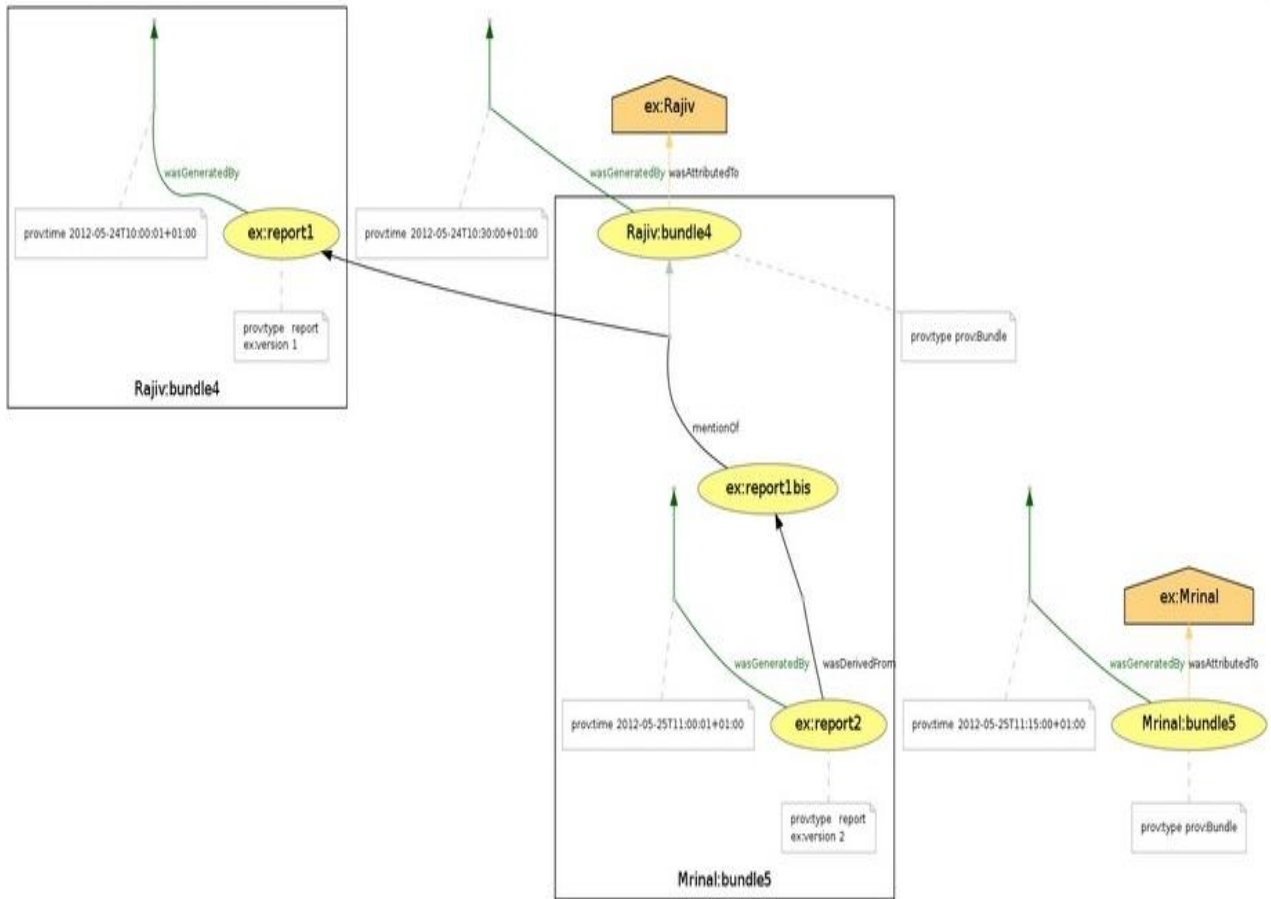


Fig 3: Document showing the relationships of various entities across bundles.[23]

| Document Metrics | | × |
|---------------------|--|----|
| Nodes | | 7 |
| Edges | | 4 |
| Components | | 3 |
| Diameter | | -1 |
| MFD | | |
| entity → entity | | 2 |
| entity → activity | | 0 |
| entity → agent | | 1 |
| activity → entity | | 0 |
| activity → activity | | 0 |
| activity → agent | | 0 |
| agent → entity | | 0 |
| agent → activity | | 0 |
| agent → agent | | 0 |

Fig 4: Document metrics demonstrating the relationship between entities agents and activities.[23]

6. CONCLUSION

Provenance also the lineage of a document is a crucial aspect in determining the trust ability of a document. However provenance of provenance must also be ensured in order for the document to be rendered trustworthy. This can be effectively done by the use of Bundles data structures. A brief illustration illustrating the use of Bundles has been provided

in the paper. In addition aiming at provenance of provenance there must also be ways to stitch provenance definitions provided by one party to be used by another party. The same has been done by the use of Mentionof relation in the above example. It can thus be concluded that Mentionof provides a clear understanding of the association of bundles with each other and provides a model as to how provenance descriptions can be altered by various parties.

7. REFERENCES

- [1] Haibo Yu, Tsunenori Mine, Makoto Amamiya,2012 "Balance: A key factor for the evaluation of semantic web applications"Proceedings of IIAI International Conference on Advanced Applied Informatics,2012.
- [2] .SaleenaJ, Ms.B, Salini, Ms.M.G., Venkateswaran, Dr.Siva,2010. A Semantic Approach to Construct a Knowledge Portal for E-Learning Using Ontology".Proceedings of 2010 4th International Conference on Distance Learning and Education (ICDLE)
- [3] Berners-Lee T., Hendler I, Lassila ,2001.The Semantic Web. Scientific American, 284(5), pp 34-43 (2001)
- [4] Pereira, R. G. , and Freire, M. M.,2005. "Semantic Web", in Encyclopedia of Multimedia Technology and Networking, M. Pagani (ed.), Idea Group, Inc., 2005.
- [5] Pereira, Rui G and Freire, Mário M. ,2006.SWedt: A Semantic Web Editor Integrating Ontologies and Semantic Annotations with Resource Description Framework" Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006)
- [6] Lopez, Gustavo, Leon Claudia ,2012.Impact of certification, evaluation and time in determining trust on the social semantic web.Proceedings of 2012 International Symposium on Computer, Consumer and Control
- [7] Gardner, Michael,2007. Using Ontologies to Support the Sharing of Learning Resources, Last modified 16 Jan, 2007 Published 16 Jan, 2007.
- [8] RDFedt tool. URL: http://www.jan-winkler.de/dev/d_rdfc.htm
- [9] RDF Instance Creator (RIC) tool. URL: <http://www.mindswap.org/~mhgrove/RIC/RIC.shtml>
- [10] Protégé tool. URL: <http://protege.stanford.edu/>.
- [11] OntoEdit tool. On-To-Knowledge web site. URL: <http://www.ontoknowledge.org/tools/ontoedit.shtml>
- [12] O'Hara , Kieron, Alani, Harith., Kalfoglou, Yannis, and Shadbolt Nigel,2004. Trust Strategies for the Semantic Web. Workshop on Trust, Security, and Reputation on the Semantic Web at ISWC2004, Hiroshima, Japan, November 2004. <http://eprints.ecs.soton.ac.uk/10028/01/ISWC04-OHara-final.pdf>
- [13] Almendra Vinicius da Silva, Schwab Daniel,2006. 2nd Semantic Web Policy Workshop (SWPW'06). Trust Policies for Semantic Web Repositories. http://www.l3s.de/%7Eolmedilla/events/2006/SWPW06/programme/paper_14.pdf
- [14] Artz Donovan, Gil Yolanda,2007. A survey of trust in computer science and the Semantic Web: Web Semantics: Science, Services and Agents on the World Wide Web archive. Volume 5 , Issue 2 (June 2007) table of contents. Pages 58-71 . Year of Publication: 2007. ISSN:1570-8268
- [15] Ding Li, Finin Timothy. Weaving the Web of Belief into the Semantic Web. Submitted to WWW2004. <http://ebiquity.umbc.edu/v2.1/get/a/publication/74.pdf>
- [16] Kuter Ugur and Golbeck Jennifer. 2007.A New Algorithm for Trust Inference in Social Networks, using Probabilistic Confidence Models. Proceedings of the Twenty-Second National Conference on Artificial Intelligence (AAAI-07). Vancouver, British Columbia, July, 2007. www.cs.umd.edu/~golbeck/downloads/Sunny.pdf
- [17] PROV-Overview URL: [-https://www.w3.org/TR/prov-overview/](https://www.w3.org/TR/prov-overview/)
- [18] PROV-DM-URL:<https://www.w3.org/TR/2012/WD-prov-dm-20120503/>
- [19] Xu Guoyan , Wang Zhijian,2010.Data Provenance Architecture based on Semantic web Services,Proceedings of 2010 Fifth IEEE International Symposium on Service Oriented System Engineering.
- [20] Rajiv Himanshu,2014 "Evaluating the need of PROV in OWL", *ACEEE, IJRTET*, 2014
- [21] Pandey Mrinal, Pandey Rajiv,2014.Analysis of provnace data model stack for OWL Onology relevance.Proceedings of Parallel Distributed confrence on Grid computing.
- [22] PROV-Links:URL<https://www.w3.org/TR/prov-links/>
- [23] PROV-STORE-URL:<https://provenance.ecs.soton.ac.uk/store/account/login/>
- [24] Online:<https://www.w3.org/TR/prov-o/>
- [25] Online:<https://www.w3.org/2001/09/06-ecdl/slide17-0.html>