Provenance Descriptions using the OWL functional syntax in Protégé

Mrinal Pandey , Rajiv Pandey, Manuj Darbari

Abstract*: The extension of the current web also known as the semantic web provides for techniques that facilitate interoperation and interconnectivity amongst applications. There is huge amount of data present on the web however computers cannot interpret the meaning of data and cannot retrieve useful information from the web by making useful decisions from this data. Thus, semantic web provides solution to these problem by embedding meta-tags to the web of data [1]. This enables software agents to interpret the meaning of information present on the web and retrieve meaningful information from the web. The embedding of meta-tags is done by Ontologies [2]. Ontologies in semantic web serve as a data model that can be used to describe concepts and relationships present with respect to those concepts in a particular domain. Although, the semantic web ontologies help the pedagogical agents to easily find and retrieve information present in the web, yet they provide no measures to facilitate the trustworthiness off information retrieved by these ontologies. As such Questions like when, how, where, by whom was a piece of data item developed still remain a question of concern. Several trust metrics and models have been proposed by researchers regarding trust assessment in semantic web Ontologies, however no standard prescribed form is present to implement and embed the same. In our paper thus we have proposed to show the implementation and embedment of trust using the concept of provenance i.e. by tracking the lineage. A discussion and implementation of provenance is provided using the PROV-DM data model in a University People Program Ontology. We have used the PROV-DM as it is the first standard model proposed by world wide web consortium and serves as a bases to easily, create, and embed provenance assertions in Ontologies.

***Index Terms*: Semantic Web, Ontology, Trust, Provenance**

I. INTRODUCTION

The amount of information being gathered from various telecommunication medium is increasing day by day.This is facilitated by the current web which allows us to share and connect smart devices amongst themselves through a variety of mediums. The current internet is primarily a publishing medium that allows data to be made available for public or personal consumption. It, however does not provide for mechanisms to allow this content to interact with each other. only hyperlinks, by the means of which browsers can navigate from one page to another.

The current web does not provide for mechanisms to enable web pages to interact with the each other and understand each other's content [3]. Recent researches [4] have found that an average of 32000 pieces of information are associated with each user this includes that data gathered from emails, web pages, documents and other sources. Many personal Information management tools [4][5][6] are used for organizing and retrieving information, however there is still huge dearth of quality information management and retrieval tools. The available tools usually provide for keyword-based search as opposed to semantically enabled search [7]. The Semantic web provides solutions to these problems by adding meta-tags to pages. Semantic Web was first proposed by the creator of the Worldwide Web, Tim Berners-Lee. According to Tim-Berners Lee semantic web aims at "creating Web which not only links documents together but also provides for technologies that recognize the meaning of the information in those documents thus providing for effective and efficient results." Semantic web is thus considered an extension of the current Web that provides common structure to the web pages by means of semantic tags thereby allowing easy recognition by external software agents as well as humans [8]. Ontologies play a crucial role in developing knowledge base systems. Ontologies are beneficial for two main reasons in developing knowledge base systems [9]. It allows for a more structured and methodical design of the knowledge bases; and it also facilitates distribution and reuse [10]. Ontologies, though help in sharing of knowledge across various domains and also act as a good mechanism to facilitate semantic search, yet they lack the element of trust. Several Researches [11] have shown that common search criteria such as creation and modification are not remembered correctly just about 50% of the time, more so the title that is the most basic criteria for searching any document is just partially correct 47% of time and totally incorrect 20% of time. Contemporary research on trust provides several mathematical trust models to the trust, however practical trust models are very few [12]. Trust can be incorporated by incorporating the element of provenance in our Ontology. In the following section we thus describe the definition of trust and discuss its relationship with provenance.

Revised Manuscript Received on June 08, 2019.

Mrinal Pandey, Research Scholar, Amity Institute of Information technology, Amity University Uttar Pradesh, Lucknow, India.

Rajiv Pandey, Assistant Professor, Amity Institute of Information technology, Amity University Uttar Pradesh, Lucknow, India.

Manuj Darbari, Associate Professor, Department of IT, BBDNITM, Lucknow, India

II. PROVENANCE AS A PARAMETER FOR TRUST

A. *Review Stage*

Trust is a fundamental factor enabling the interaction of people with each other; hence it is natural that trust be researched with respect to relationship between applications and agents [13]. A trustable ontology enables systems with different models to share confidence relationship amongst each other and also the information on how this confidence relationship has been formed. This is thus a crucial parameter in today's digital business, where different organizations with differing framework must co-operate, communicate and network to utilize various services [13]. Several definitions of trust have been proposed by researchers over years. For. Eg. Diego Gambetta and Audun Josang propose that trust is a level degree to which an agent can consider that another agent will responsibility and successfully perform a specific set of action[14][13]..Another definition of trust describes trust to be the credibility associated with a passionate entity that it will behave without any malicious intent, while a third definition describes trust to be the credibility associated with a rational entity that it will forbid any malicious manipulation by the passionate entity[15][13].Provenance [16] also known as lineage [17], pedigree or inception attempts, however attempts to track the history of a data item. It also attempts to capture the history of all those people that are involved in processing the data item. Provenance is receiving substantial attention in various research areas[18][19].Provenance has gathered substantial attention due the crucial role it serves in real life applications like healthcare[20],banking and finance and life sciences[21][22][23].Provenance thus acts as a parameter for trust measurement because it provides answers to questions like-

- Who generated a piece of data item?
- When was it generated?
- What are the sources from which it is derived?
- What are the procedures used to derive it?
- Thus, in our paper we aim to provide answers to these questions using the PROV-Dm data Model. We have tried to show the implementation of the PROV-DM data model in a university people
- Ontology for tracking provenance.

III. USE OF SEMANTIC WEB AND ONTOLOGIES IN THE EDUCATION

The greatest advantage of the sematic web is its ability to retrieve information by interpreting the meaning of the phrase provided in the search criteria. This allows for retrieval of meaningful information and sorting out of irrelevant details. This is made possible because of the Semantic Web. The semantic web allows "the users to find dependencies between tagged information using reasoned rules and governing tools also called as ontologies which provide logic and structure to the information that is present in semantic web pages"[24].Several Ontologies like Edu Onto and Onto Edu have already been developed to provide for LMS and other operations of the University. These Ontologies-

- Allow students to effectively store, retrieve and locate informative content on the web.
- Provide for effective research by aggregating information from different and relevant sources with the help of meta-tags, as opposed to keyword-based search sources.
- Aid in the construction of new learning material by semantically tagging these materials thereby allowing the creation of new learning from existing ones.
- Help in the creation of Personal Learning Networks [24]
- Help in Personal Educational Administration-Semantic web ontologies help describe institutional courses and degrees semantically, this will intern allow the components of education available across institutions. Description of courses using semantically enabled tags will allow students to get courses and experiences from several institutions on the same platform [25]

The above points clearly state the advantages of semantic web and Ontologies in the education domain, however the trustworthiness of information gathered from these ontologies is a crucial issue that needs to be addressed. In the following section we discuss the ways to handle this issue by the means of Provenance assertions.

IV. OUR WORK

Provenance as described before helps us in embedding the component of trust in our Ontology. The procedure for describing provenance is stated below.
The methodology that we have applied to develop this Provenance model is action Research methodology which is stated clearly by the **Fig 1** provided below-

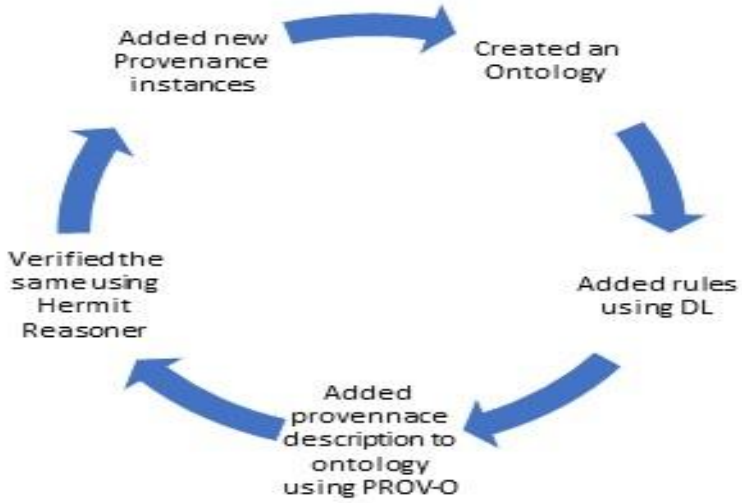

Fig 1:Action Research Methodology for Adding Provenance to University Ontology.

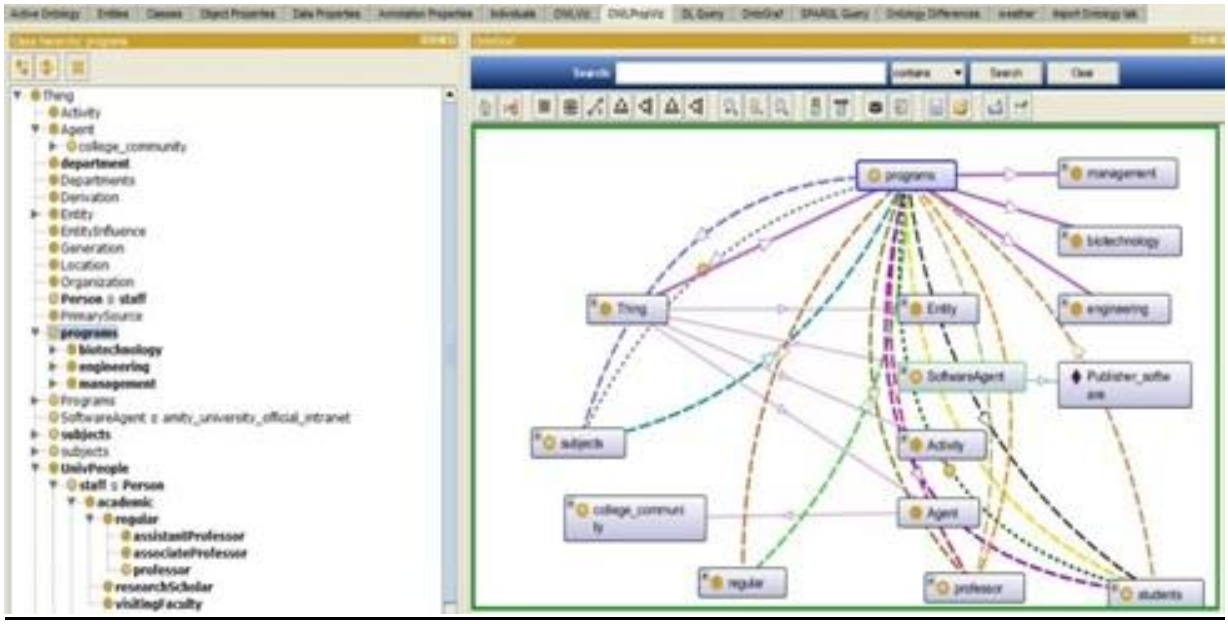

Fig 2: A screenshot of the Ontology is shown Above.

A bifurcation of the classes used in our Ontology and those that aid in provenance are described below

Table I

Classes of University	**Subclasses**
Programs	Biotech, Engineering, Management
Subjects	Core subjects, Foundation subjects
University People	Staff , students

Table II

Classes	**Equivalence classes**
Agent->Person	Staff which includes Academic and Administrative staff.
Post graduate student	MCA, MBA, M.Tech,…..and other post graduate courses
Undergraduate student	B.tech, BCA,BSC,BBA and other under graduate courses
Agent -> software agent	This includes amizoner a software agent that automatically publishes documents
Agent->Organization	This can be any organization on the behalf of which the other agents or the human agents act like any University.

Table III

Classes to Help in embedding provenance	Subclasses
Agent	Organization, person, software agent
Entity	Bundle, Collection, plan
Activity	Communication, Generation, Invalidation, Association, Attribution, Delegation
Influence	Activity influence, agent influence, entity influence

V. WHY OWL FUNCTIONAL SYNTAX

In our approach we have used the OWL functional syntax because the OWL functional syntax helps in converting structural specifications to concrete syntaxes. It is easy to use and understand, though it is not considered as a standard, but it is the best syntax to convert OWL structural specifications to concrete syntax.

VI. PROVENANCE USING OWL FUNCTIONAL SYNTAX IN UNIVERSITY PEOPLE PROGRAM ONTOLOGY

The Code and snapshots provided below show implementations of provenance in University. The code snippet provided below states that Nidhi, an individual of class professor and the coordinator of BCA-1 is also an agent who is involved in various creation activities like Creation _of_University_10_year_Progress_Report and Creation_of_IQAC_report_for_Department this shown by the provenance instance prov#wasAssociatedWith, further IQAC_report is a new entity that is derived from the University_10__year_progress_Report as is described by the prov#entiy and prov#wasDerivedFrom and as was created by the Generation activity specified by the prov#wasGeneratedBy. Further Nidhi, belongs to Amity_University which is an oragnizational Agent, hence it can be described by the prov instance pov#actedonBehalfof. The verification of the provenance instances so implemented has been done using Hermit Reasoner as shown in the Fig 3 stated below. Further Fig 4, Fig 5 and Fig 6 provided below also shows the relationships of activities, entities and agents along with the various entities that were involved in the generation of new entities from Original one.

The figures are generated in OWLViz showing individuals and their provenance instances.

Code snippets[26]

```
Class Assertion (: BCA: BCA-1)
ObjectPropertyAssertion(:hasCoordinator :BCA-1 :Nidhi)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#was
AssociatedWith>
:Creation_Of_University_10_Year_Progress_Report :Nidhi)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#was
AssociatedWith>
:Creation_of_IQAC_report_for_Department :Nidhi)
ClassAssertion(<http://www.w3.org/ns/prov#Entity>
:IQAC_report)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#was
AttributedTo> :IQAC_report :Nidhi)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#was
DerivedFrom> :IQAC_report
:University_10__year_progress_Report)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#was
GeneratedBy> :IQAC_report
:Creation_of_IQAC_report_for_Department)
Declaration(NamedIndividual(:Nidhi))
ClassAssertion(:assistantProfessor :Nidhi)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#acte
dOnBehalfOf> :Nidhi :amity_university)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#was
AttributedTo> :University_10__year_progress_Report
:Nidhi)
Declaration(NamedIndividual(:amity_university))
ClassAssertion(<http://www.w3.org/ns/prov#Organization>
:amity_university
```

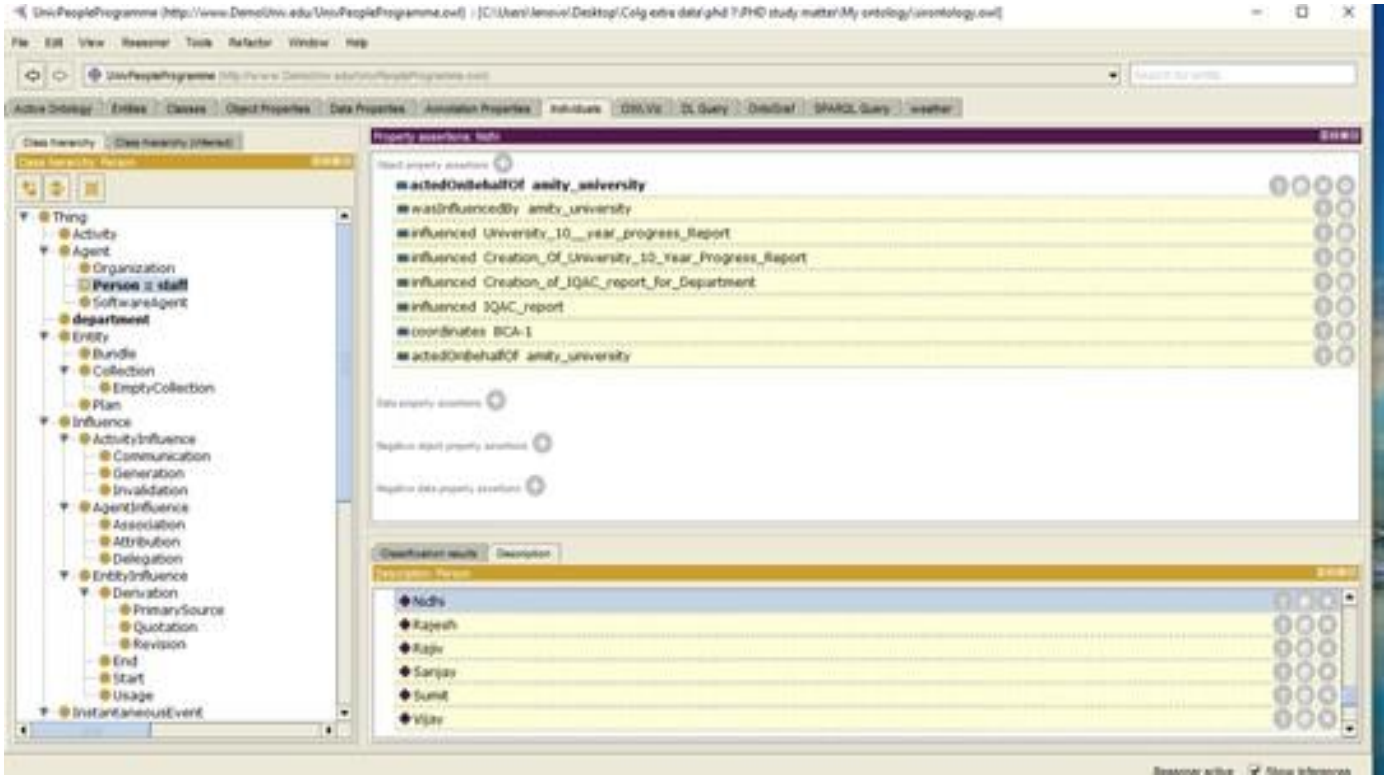

Fig 3: Verification of Provenance using Hermit reasoner

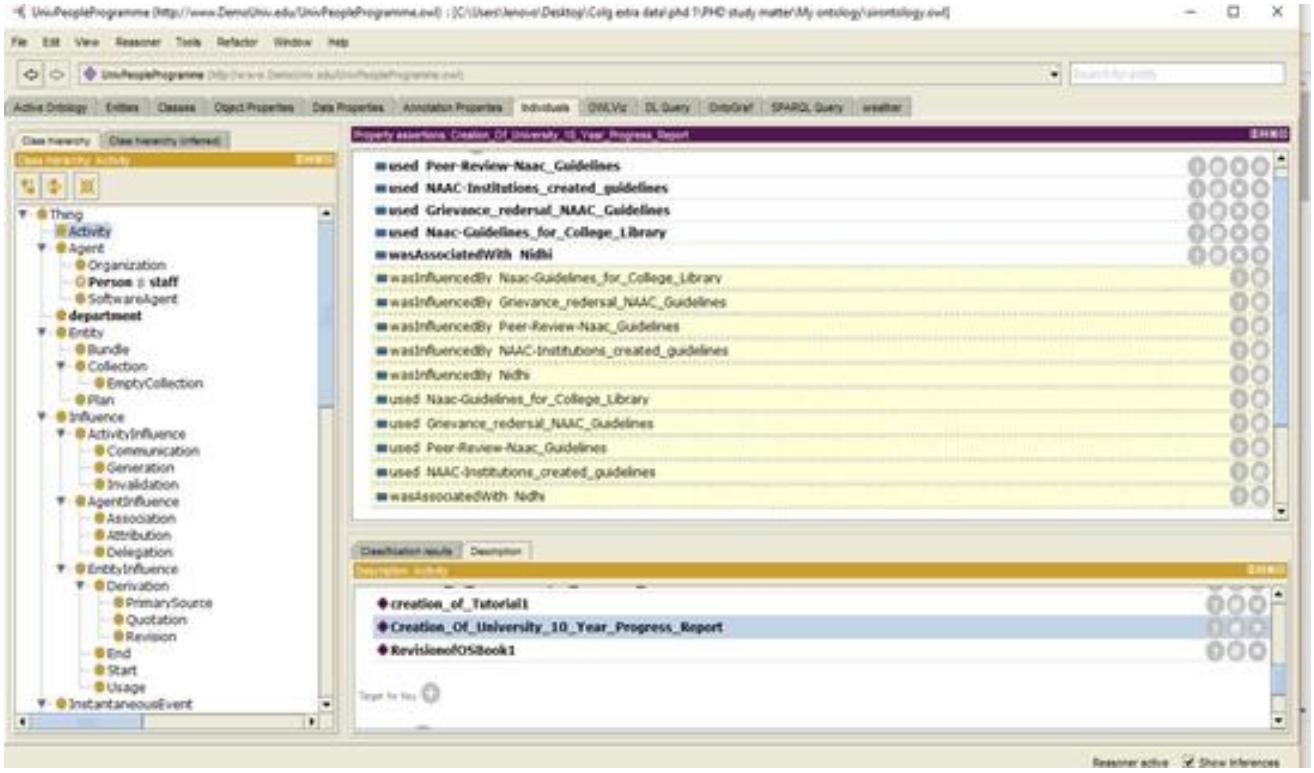

Fig 4: Verification of Provenance instance using Hermit reasoner

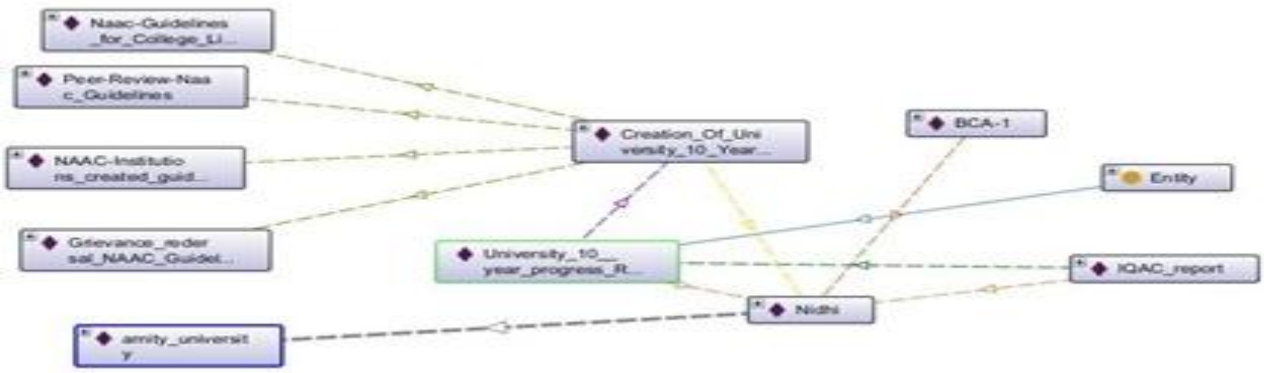

Fig 5: Fig showing the relationships between Entities, Agents and Activities

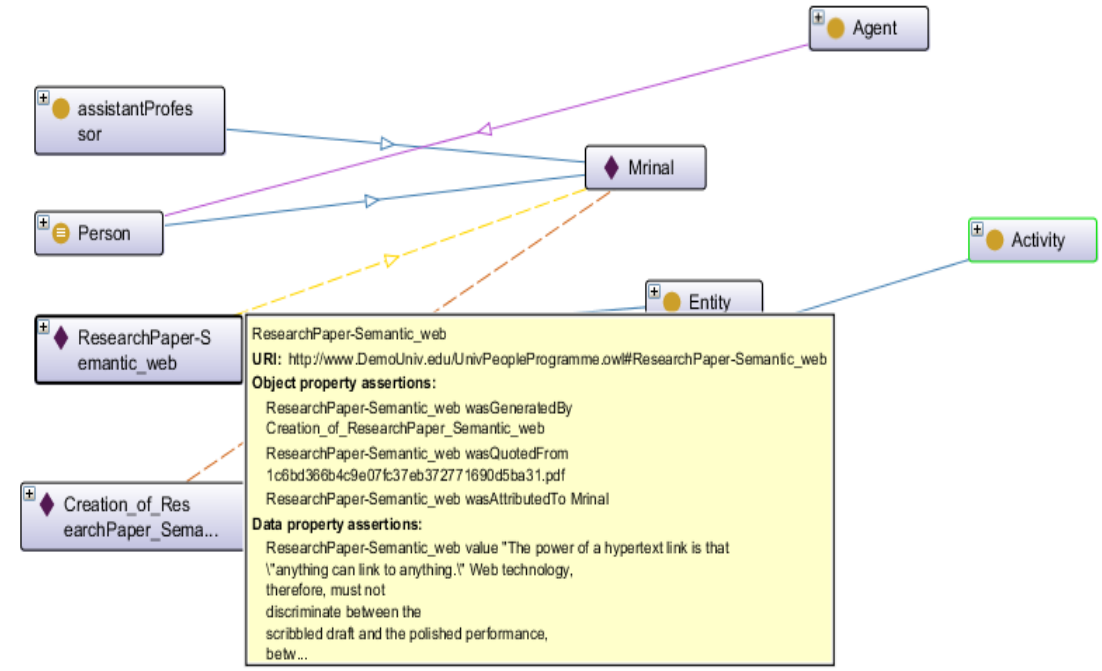

Fig 6: Fig showing the relationships between Entities, Agents and Activities

Example of Activity influence, starting at time and ending at time

In another example we track the provenance associated with an activity tracking the provenance information of the agent that created the activity , the time at which the activity started and ended. This is shown by the prov#wasInfluencedBY , prov#Startedattime and prov#EndedAttime. The same can be shown by the figure and the reasoned inferences shown in the below given Fig 7. In the example shown below we have shown the same by giving the example of the activity CreationOfStudentAggreagate. As this activity is taken out by the Agent Sanjay, an individual of class Professor thus we have used the prov assertion prov#wasInfluencedB and have also shown the time at which it started and ended.

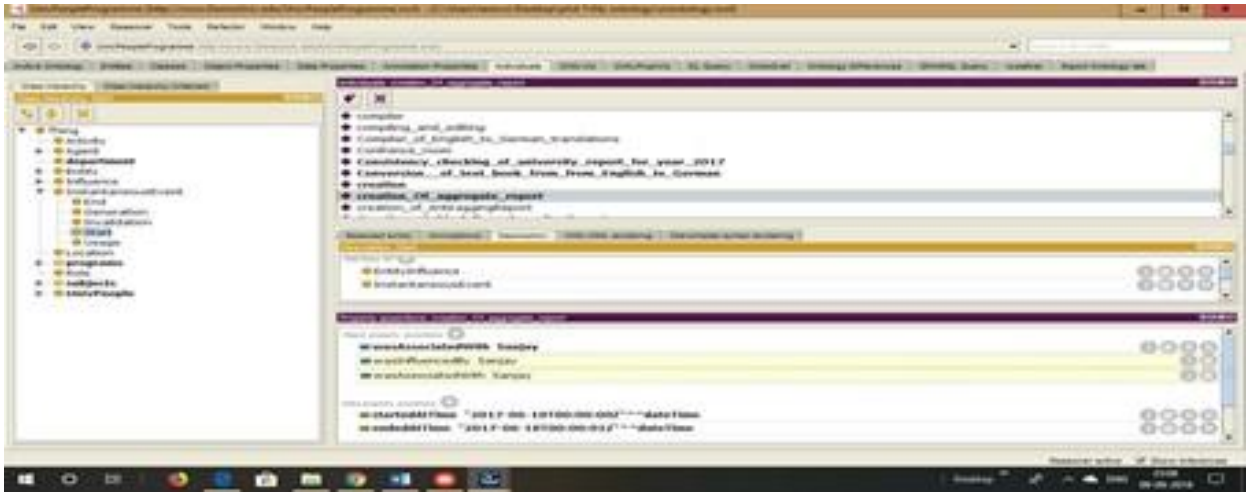

Fig 7: Inferencing for Activity influence, started at and ended at time of provenance using Hermit reasoner

Example of software agent

In the example of the Ontology provided below we have tried to assert the fact that Software's too can act as agents thus. The fig 8 provided below states that Publisher, a type of software agent is involved is involved in the creation of the entity new_facts_on_university_website ,this is done by the activity Creation_of_new_facts_on_university_website. The provenance assertions for the same have been shown using prov#SoftwareAgent, prov#Entity, prov#Activity, prov#wasAttributedTo, prov#wasAssociatedwith. The Fig 8[26] provided below has been generated using OWLVIZ in Protégé.

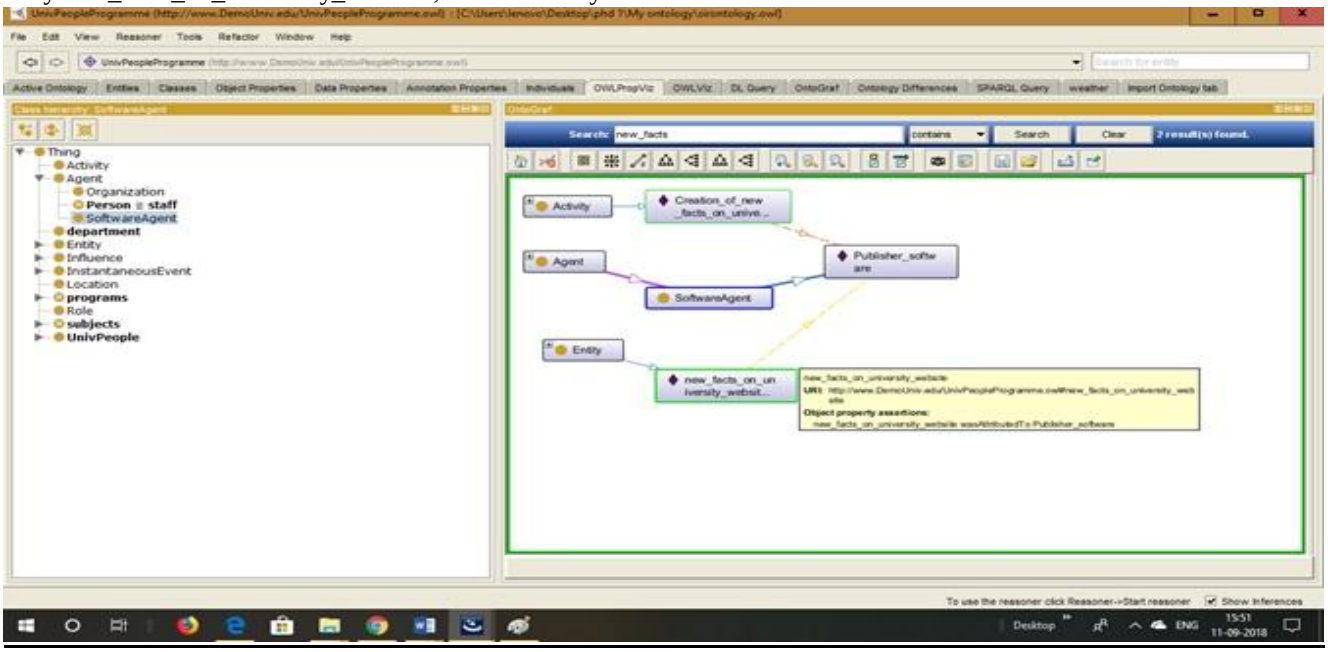

Fig8: Figure showing the relationship of Software agent to Activity and Entity

The code for the above Ontology relationships, in OWL Functional Syntax is shown below[26]-

Declaration(NamedIndividual(:Publisher_software))
ClassAssertion(<http://www.w3.org/ns/prov#SoftwareAgent> :Publisher_software)
Declaration(NamedIndividual(:new_facts_on_university_website))
ClassAssertion(<http://www.w3.org/ns/prov#Entity> :new_facts_on_university_website)
Declaration(NamedIndividual(:Creation_of_new_facts_on_university_website))
ClassAssertion(<http://www.w3.org/ns/prov#Activity> :Creation_of_new_facts_on_university_website)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#wasAssociatedWith> :Creation_of_new_facts_on_university_website :Publisher_software)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#wasAttributedTo> :new_facts_on_university_website :Publisher_software)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Declaration(NamedIndividual(:new_facts_on_university_website))
ClassAssertion(<http://www.w3.org/ns/prov#Entity> :new_facts_on_university_website)
ObjectPropertyAssertion(<http://www.w3.org/ns/prov#wasAttributedTo> :new_facts_on_university_website :Publisher_software)

VII. CONCLUSION

A key use of provenance is to enable the reproducibility of processes, Provenance also provides new information for reasoning. The provenance model provided above provides several processes to reason and determine that a given ontology is trustworthy. The procedures for creating, embedding and reasoning provenance with respect to an Ontology have been provided above and the same thus can be effectively used in other applications as well.

REFERENCES

1. Tim Berners-Lee, Jim Hendler, and Ora Lassila. (2001 ,May 17). "The semantic web. Scientific American." pp. 35-43. [Online]. Available:http://www.sciam.com/article. cfm?id=the-semantic-web.
2. S.K. Rajapaksha ; N.Kodagoda ,"Internal Structure and Semantic Web Link Structure Based Ontology Ranking",In Proc.IEEE , 4th International Conference on Information and Automation for Sustainability,2008,pp 86-89.
3. Tim Berners-Lee, Jim Hendler, and Ora Lassila. (2001 ,May 17). The semantic web. Scientific American. pp. 35-43. [Online]. Available: http://www.sciam.com/article. cfm?id=the-semantic-web
4. E.Cutrell, D.C. Robins, S.T.Dumais, R.Sarin "Fast, Flexible Filtering with Phlat—Personal Search and Organization Made Easy", In Proc. CHI 2006,pp 261-270..
5. Dumais, S, Cutrell, E, Cadiz, J, Jancke, G, Sarin, R, Robbins, DC (2003) "Stuff I've seen: a system for personal information retrieval and re-use".In Proc. 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, SIGIR 03,2003,pp.72-29.
6. E. Freeman, D. Gelernter, "Lifestreams: a storage model for personal data." *SIGMOD Record*, Vol. 25, No. 1,pp. 80-86, March 1996 .
7. S.Stumpf, E.Fitzhenry, T.G.Dietterich ,"The Use of Provenance in Information Retrieval",In Proc. Workshop on Principles of Provenance (PROPR), 19 - 20 November 2007..
8. N.J.I. Mars and A. Schreiber, "Direct Access to Knowledge in Bibliographic Databases," Proc. ARTINT Workshop Artificial Intelligence and Information Retrieval ,Sept. 1985, pp. 83–86, 1985
9. N.J.I. Mars, "The Role of Ontologies in Structuring Large Knowledge Bases*," Knowledge Building and Knowledge Sharing*, K. Fuchiand T. Yokoi, eds., pp. 240–248, Tokyo: Ohmsha, 199.
10. P.E. van der Vet ,N.J.I. Mars ,"Bottom-Up Construction of Ontologies",IEEE Transactions on Knowledge and Data Engineering, Volume: 10 , Issue: 4 , Jul/Aug 1998.
11. T. Blanc-Brude, D.L..Scapin,, "What do people re-call about their documents?: implications for desktop search tools", In Proc. IUI,pp.102-107.
12. S.Stumpf, E.Fitzhenry, T.G.Dietterich," The Use of Provenance in Information Retrieval",In.Proc. Workshop on Principles of Provenance (PROPR), 19 - 20 November 2007, pp.1-7.
13. L.Viljanen,"Towards an Ontology of Trust", In Proc "TrustBus'05 InProc. of the Second international conference on Trust, Privacy, and Security in Digital Business",22 August,2005pp.175-184.
14. D. Gambetta," Can we trust trust? Trust: Making and Breaking Cooperative Relations", *213–237 (2000) Electronic edition*
15. T. Grandison, M. Sloman,"Specifying and analysing trust for internet applications". In: Proceedings of the 2nd IFIP Conference on e-Commerce, e-Business, e-Government I3e 2002", October 2002, pp. 145-146.
16. P. Buneman and W.-C. Tan, "Provenance in databases," in SIGMOD'07: In Proc of the 2007 ACM SIGMOD international conference on Management of data.,2007, pp. 1171–1173.
17. J. Widom, "Trio: A system for integrated management of data, accuracy,and lineage," In Proc. CIDR, 2005, pp. 262–276.
18. W. C. Tan, "Provenance in databases: Past, current, and future," *IEEE Data Eng. Bull*., vol. 30, no. 4, pp. 3–12, 2007.
19. R. Bose and J. Frew, "Lineage retrieval for scientific data processing: a survey", *ACM Compute. Survey*., vol. 37, no. 1, pp. 1–28, 2005.
20. M. Blount et. al., "Century: Automated aspects of patient care," In Proc. of the 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications,2007, pp. 504–509.
21. M. Jayapandian, A. Chapman, V. G. Tarcea, C. Yu, A. Elkiss, A. Ianni,B. Liu, A. Nandi, C. Santos, P. Andrews, B. Athey, D. J. States, and H. V.Jagadish, "Michigan molecular interactions (mimi): putting the jigsawpuzzle together," *Nucleic Acids Research*, vol. 35, no. Database-Issue,pp. 566–571, 2007
22. A.Kementsietsidis, M. Wang, "On the Efficiency of Provenance Queries",In Proc."25th International Conference on Data Engineering",2009,pp.1223-1226.
23. Y.Theoharis, I.Fundulaki, G.Karvounarakis, and V.Christophides,"OnProvenance of Queries on Semantic Web Data", *IEE Internet Computing*, vol. 15, pp. 31-39.
24. B.Ö.Czerkawski,"The semantic web in teacher education", *The Turkish Online Journal of Educational Technology*, October 2014vol 13,pp. 144-146.
25. V.Devedzic,"The semantic web in education", *International Journal of Artificial Intelligence in Education* ,2004,pp. 39-65.
26. Protége[online] Available: https://protege.stanford.edu/
27. World wide Web Consortium: "PROV-DM:The Prov Data Model,"W3C recommendations,2013.[Online].Available: http://www.w3.org/TR/2013/REC-prov-dm-20130430/

AUTHORS PROFILE

Ms. Mrinal Pandey is a Research scholar, from Amity Institute of Information technology, Lucknow, Uttar Pradesh. She is MCA. The author is doing Ph.D. in Information Technology from Amity Institute of Information Technology, Amity University, Lucknow Campus, Uttar Pradesh. Her research areas include Semantic web, Artificial Intelligence, Provenance. Her papers are published in IEEE International conferences and are all Indexed in Scopus.

Dr. Rajiv Pandey, Senior Member IEEE is a Faculty at Amity Institute of Information Technology, Amity University, Uttar Pradesh, Lucknow Campus, India. He possesses a diverse back ground experience of around 30 years to include 15 years of Industry and 15 years of academic. His research interests include the contemporary technologies as Semantic Web Provenance, Cloud computing, Big-Data, and Data Analytics. He has been ontechnical Committees of Various Governmentand Private Universities. He is intellectuallyinvolved in supervising Doctorate Research Scholars and Post graduateStudents. He is also an active contributor in professional bodies like IEEE, IET and LMA. He is a member of Machine Intelligence Labs.

Dr. Manuj Darbari, At present working as Associate, Professor in Department of Computer Science, BBD University, Lucknow, UP, India. Job includes teaching and guiding students intheir Project/Thesis work and upgrading the syllabus as per the industry needs.Experienced Associate Professor with a demonstrated history of working in the telecommunications industry. Skilled in Mathematical Modeling, Analytical Skills, Computer Science, Research Design, and Entrepreneurship. Strong education professional with a PhD focused in Business Administration from University of Lucknow. He research areas include-Semantic web,Software Modeling Using Model Driven Architecture, Soft Computing,Business Analytics,Workflow Modeling using Petrinets,Urban Traffic Systems,Mobile Computing.