Manifestation of Redundancy Conflict Diminution by Reusing Ontological Description

Aradhana Negi[1], Parminder Kaur[2]
[1,2] *Department of Computer Science, Guru Nanak Dev University, Amritsar, Punjab*
(E-mail: aradhana.csersh@gndu.ac.in)

***Abstract*—**Although the idea of endorsing meaning with web resources gives tremendous success to Semantic Web, but on the other side of coin these semantics leads to some serious interpretation issues e.g. multiple meaning of a term (Ambiguity) and same description in multiple Ontologies (Redundancy). In order to fully exploit the features of Semantic Web these issues must be treated as a high prioritized task. This paper presents Semantic Redundancy issue of Semantic Heterogeneity and describes Ontology Reuse as Semantic redundancy conflict diminution. An analysis of redundancy conflict is presented using Semantic Web Search Engines and it is conclude that a) universal strategies are required to reuse Ontologies in an effective way b) Web Resource dealing with different knowledge domains need to relate different Ontologies competently for annotation.

Keywords*—Semantic Heterogeneity; Semantic Inconsistency; Semantic Redundancy; Ontology; Ontology Reuse*

I. INTRODUCTION

The advancement of Information Technology promotes rapid data generation via web pages, blogs, social media and e-commerce. Every day voluminous big data has been generated by millions and millions of users [1]. As Internet is hub of heterogeneous distributed data, so people use to search, share, and compose information from a wide variety of data sources. Moreover, with advanced tools and techniques, it is very easy to publish data on internet but this ease of publication can leads to data or Semantic conflict. Data Conflict is a clash between two or more merging web resources, where some of the data is not matched whereas Semantic Conflict is a clash between web resources, where resources get merged effectively at syntactical/textual level but behave differently with regards to meaning of events, operations, objects, actions.

Semantic Conflict (also called Semantic Inconsistency) is a problem generated with Semantic Heterogeneity. Semantic Conflict can be classified into two major problems a) Semantic Redundancy and b) Semantic Ambiguity. Semantic Redundancy means, a situation where several semantic descriptions of different Web resources represent same intended meaning either using same term e.g. word 'Emergency' description in three Ontologies or different terms e.g. "Cash and Money" in five Ontologies. Semantic Ambiguity means, one word/object/event having different contexts in different web resources. Both of them are hammering problem for Semantic Web Applications [2]. By keeping only Semantic Redundancy Conflict in mind, this paper elaborates Ontology Reuse as a proactive tact to redundancy problem

Semantic Inconsistency of vast distributive information sharing system on internet gives rise to Semantic Conflict onto the success path of Semantic Web so the first objective of this paper is to understand Semantic Redundancy Conflict of Semantic Web. In literature, there is no universally accepted Ontology Reuse work flow, because this process is still considered as a skill more than an engineering task therefore second objective is to present Ontology Reuse as an easiest way to lessen Semantic Conflict and to simplify the Ontology Reuse by arranging steps in a work flow formation. The third objective of this paper is to analyse the redundant term in two senses 1) using same term/word 2) using different word.

Even though reusing is not the only solution for Semantic Redundancy Conflict but it is based on "Prevention is better than cure" strategy, which is more pragmatic in real world thus with above mentioned strategy the paper writing is divided into VII sections. Section II discusses related work for Semantic Redundancy Conflict and Ontology Reuse. Semantic Redundancy and Semantic Web is described in Section III. Next, Ontology reuse, its objectives and general Ontology Reuse work flow has been talked about in Section IV. In Section V, Semantic redundancy Conflict analysis and Ontology reuse research gaps is presented. Finally, Section VI gives closing remarks and future work.

II. RELATED WORK

Gracia and Mena [2] have explained Semantic Heterogeneity issues on web. They have proposed two solutions: Sense Clustering and Sense Disambiguation for Semantic Redundancy and Semantic Ambiguity respectively. In [3], a comparative analysis of Ontology engineering methodology is presented with a goal, to discover an effective methodology for Semantic Conflict detection Ontologies. The author recommended a methodology called METHONTOLOGY for Semantic Conflict detection.

A feasibility study on Ontology Reuse has been conducted by [4]. The author presents empirical evidence and findings in the form of methods, methodologies, tools and techniques of Ontology Reuse with two case studies on e-health and e-recruitment. The results of this paper argue for the need of context and task sensitive treatment of Ontologies, Optimal Ontology Reuse methodologies and research and development

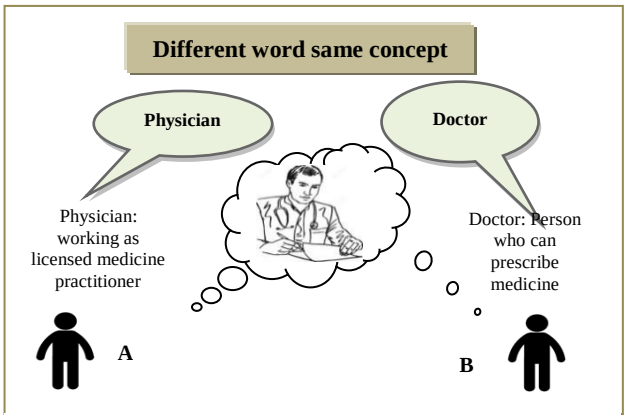

Figure 1. Semantic Redundancy conflict scenario

challenges of Ontology Reuse. A basic question "What is Ontology reuse?" has been addressed in [5], and they stressed on the issue of lack of formal definition and degree of reuse. This paper defined Ontology Reuse and Reusability with respect to various literature contributions and presents a solution for Ontology Reuse by defining seventeen definitions. In the Scheme given in [6], several Semantic inconsistency detection and Ontology merging algorithms have been proposed. The experimental part of these algorithms have been implemented on local heterogeneous Ontologies only and the results shows the time and cost efficiency with respect to existing work. This paper [6] makes human involvement one step down by proposing these algorithms. Recently in 2017, an empirical study of Ontology reuse on BioPortal Ontologies is given in [7] where inclusion of classes and properties within BioPortal Ontologies and from other source Ontologies has been examined.

III. SEMANTIC REDUNDANCY AND SEMANTIC WEB

The world have witnessed the accomplishment of Semantic Web and its applications in the form of web content annotations, software agent, multi-agent systems, knowledge representation, information exchange between various web resources/organizations, seamless information integration, Ontology standardization and Web services. Semantic Web efficiently addressed the Web Interoperability challenge between distributive heterogeneous environments and reduces pricey human interventions with automatic and semi-automatic schemes. Though this achievement of Semantic Web makes it a mature field but still many research gaps exist for its full realization. One step towards this realization was introduction of notion of Ontology for knowledge representation. The intrinsic properties of Ontology make it unambiguous, sharable, adaptable, and reusable [8] in comparison to the other existing ways of classical knowledge base but sometimes the variation in knowledge representation causes Semantic Conflict. These conflicts occur whenever knowledge based systems do not use the identical interpretation [9] of the used information.

Semantic Redundancy Conflicts is assumed one of the crucial steps that should be carried out efficiently so as to pave the way towards establishment of Semantic Interoperability between heterogeneous [3] Web applications. Consider a scenario of redundancy conflict (same description with different words) where a Semantic Web content-publisher **A** wants to annotate its content with some Ontology and being unaware of existing related Ontologies, **A** creates new Ontology **OA** where the notion of *medical practitioner* is described by the term *Physician*. Let's say another publisher **B** uses same concept with term *Doctor* and annotated it with an Ontology **OB** (*existing Ontology*). Now these two Ontologies are unaware of one another from automation view point (see Fig. 1), whereas the basic rules for a good Ontology says [10] that instead of creating new Ontology from starch one should reuse the existing Ontology and add the new features to builds new Ontology. If the applications based on these two Ontologies wants to communicate with each other than due to the reusability feature (of Ontology) these two can understand redundant terms in a seamless manner. Semi-automatic mapping, merging and integration of Ontologies can detect and resolve these redundancy case if Ontologies are related to each other on the account of inheritance and hierarchical domain relationships. Above discussed scenario is also applicable for second case where a term is redundantly defined in different Ontologies. Message exchange, parameters exchange, and integration of some or whole content of Web resources are directly affected by this conflict.

IV. ONTOLOGY REUSE: OBJECTIVE AND WORK FLOW

Ontology is a ceremonial way to describe the concepts of a group of terms linked to a specific domain [11]. Formally it can be defined as "*a process in which available (ontological) knowledge is used as input to generate new Ontologies*" [12]. It is the mainstay of Semantic Web approach and defines the terms used to describe, standardize and represent an area of knowledge [13]. Ontology merging and Shared Ontology can multiply the degree of interoperability between Web resources with single description of overlapping knowledge domains. These have been productively implemented for the possible solutions of problems deriving from the management of shared information, distributed knowledge, and the efficient integration of information across applications [14]. The construction of these Ontologies from starch is a time consuming and mental intensive task but Ontology Reuse can arbitrate this task by utilizing the previously defined concepts from existing Ontologies. This Ontology Reuse provides a manner for resolving semantic heterogeneity and enhancing semantic interoperability among data-sharing systems for the purpose of knowledge sharing and reuses [15].

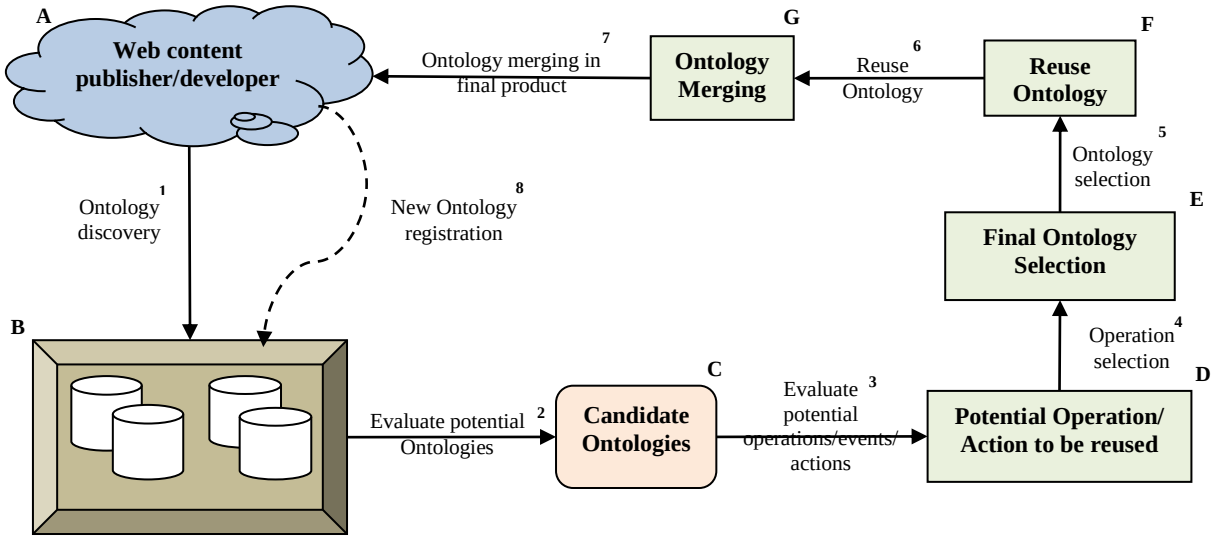

Figure 2. General Ontology Reuse work flow

The huge amount of literature on Ontology Reuse process has attested its importance and utilization. The variation in Ontology Reuse steps leads to mystification, one part of literature support Reverse Engineering, Restructuring and Forward Engineering technique [16][17] as an inseparable part of Ontology Reuse process while other does not emphasise on this strategy [18-20]. The general Ontology Reuse work flow has been diagrammed in Fig. 2. The alphabetic order depicts the general steps and numerical order illustrates the process for corresponding steps. The Ontology Reuse process initiated with the discovery of relevant Ontologies by looking up to the Ontology repository. Once the candidate Ontologies identified, the evaluation process starts to choose the best one for reuse. Requirements matched against Operations/events/action/ parameter of potential Ontologies and after final selection of Ontology/Ontologies the actual Ontology Reuse begins. This reuse can be with Reverse Engineering or with simple code extraction/inheritance of the selected Ontology. Integration and merging of Ontologies subject to the translation of uncommon terms to common one i.e. common description for overlapping terms.

Reusing Ontology avoids redundancy by means of duplicity and supererogatory of terms and it also reduces coding effort when reused via a proper channel. It is a preventive approach where proactively Ontology Engineer put efforts for Reusability and Ambiguity. The efforts arising from Ontology Reuse process could be calculated using a cost estimation model ONTOCOM [21-22] but sometimes it is not feasible to reuse the Ontology because the reusing cost of Ontology exceeded the construction cost of new Ontology, in such rare cases cost-benefit analysis recommend the cheaper one. The tradeoffs between usability and reusability effort and time is still an issue for Ontology Reuse [16], but it is a must practice for coding the knowledge because Redundancy reduction improves information retrieval by narrowing search space.

V. AN ANALYSIS OF SEMANTIC REDUNDANCY CONFLICT

Out of many Semantic Web Search Engines, the Swoogle (http://swoogle.umbc.edu) and Watson (http://watson.kmi.open.ac.uk/WatsonWUI/) are two most popular engines where both uses crawler-oriented indexing scheme for the discovery of Semantic Web Content. The other Semantic Web Search engines like Semantic Web Search, Sindice, and Semantic Web Search Engine are not accessible and testable so we have analyzed Semantic redundancy conflict using two mentioned popular engines. The online access for Semantic description of 'Doctor' and 'Physician' is presented in Table 1. Both the selected engines have discovered many Semantic documents containing given words like Watson comes with 60 whereas Swoogle gives 578 Semantic documents for the word 'Physician'. Here, a noticeable thing is Ontological descriptions for word 'Doctor' is 162, this value is much bigger than the possible description of this keyword. The presence of redundant Ontological description can complicate the discovery or reuse of Ontology as in Table 2, where the term 'Doctor' is redundantly present in four Ontologies. One more well defined and fine-grained terminological description of term 'Doctor' is presented in Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) (http://bioportal.bioontology.org/ontologies/SNOMEDCT/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FSNOMEDCT%2F112247003https://bioportal.bioontology.org/ontologies/BDO) as given in Fig. 3. Although the simple reading of these descriptions proves that all these descriptions either directly or indirectly indicate the same thing but here the problem is automatic selection of one. The presented redundancies negatively affect the basic notion of Semantic Web Technology that was invented to enhance machine readability and understandability [19] by means of Interoperable systems.

Table 1. Semantic Redundancy through Swoogle and Watson

Keyword	Swoogle		Watson
	Semantic Document	*Ontology*	*Semantic Document*
Doctor	1062	167	Doctor
Physician	578	92	Physician

Gracia and Mena [2] have suggested a solution to reduce the Semantic Redundancy problem using Sense Clustering, even though this is a good solution for redundancy reduction but here this paper focuses on reusing Ontology or Ontological terms which will never demand for such solutions. Reusing Ontological based description from more general Ontologies can simplify its practice, convenience, and effectiveness. The domain-oriented OWL Ontology reuse i.e. Bone Dysplasia Ontology (BDO) (https://bioportal.bioontology.org/ontologies/HUPSON) and Human Physiology Simulation Ontology (HUPSON) (https://bioportal.bioontology.org/ontologies/FMA) reuses terms of Foundational Model of Anatomy (FMA) Ontology (https://bioportal.bioontology.org) of BioPortal [23] (BioPortal, an on-line storehouse of 656 bio medical Ontologies) is an effective reuse but this sort of reusing is not universal like 'Account', an ambiguous term have redundancy conflict within ambiguity conflict and no domain-oriented reuse like BioPortal is available for such terms.

Thus, to realize Semantic Redundancy Diminution there must be some universal ways to describe terms once and reuse as many times as possible regardless of their Ontological language. These universal ways not only increase Semantic Interoperability but also lift the efficiency of data discovery systems.To develop cost-effective and high-quality Ontologies the obvious considered methodologies and guidelines is Ontology Reuse. Unfortunately, this reuse still poses a considerable amount of challenges for the Ontology developer community [5]. The observed challenges are:

- Non-technical challenge- It is the willingness and level of experience of Ontology Engineer for sharing their knowledge with others and accepts the knowledge from outside Ontologies. It is assumed as an art more than a science [4] where recycling and extension of existing knowledge is considered.
- Identification of existing Ontology usability [6] [30] by understanding the importance of reusable content. It involves the decision making about cost-effectiveness of Ontology to be reused. Moreover, lack of efficient cost estimation and cost-benefit analysis techniques for Ontology Reuse make this challenge more complicated. The current paper observes that little work has been done in literature for cost estimation analysis for successful Ontology Reuse methodologies.
- The exponential growth of web open Interoperability challenge for Ontology Reuse and this growth of web is putting intense complex computational load on the information systems, which are especially meant for discovery, translation, mapping, merging, optimization, and integration. The variation of knowledge representation among various information sharing systems like sharing between Intelligent Systems/Expert Systems, Web Services and Web Applications raises Intercommunication challenges for Ontology Reuse.
- The Ontology dependencies hinder integration in Ontology Reuse process. These dependencies cause interpretation issues. Some research work emphasized Generic Ontologies but the lack of sufficient Generic Ontologies faded this notion. So here is the high need for generic Ontologies and their proper advertisement among Ontology Engineers.
- The state-of-the-art Ontology Merging, Mapping and Integration are semi-automatic this means human intervention is still required for final Ontology selection. Insufficiency of Ontology reusing tools and techniques make detection and resolution of semantic inconsistency [6] problematic. Some other challenges are standardization of reusability definition, degree of reusability [5], and requirement of pragmatic methods for utilization of source Ontologies to a maximal extent [12].

Table 2. Ontology Redundancy

Sr. No	Doctor	
	Description	*Ontology*
1	..person who can prescribe medication	http://micra.com/COSMO/OlderVersions/COSMO.owl
2	The profession of being a medical doctor, i.e. having attended medical school and being licensed to practice medicine.	OWLS-TC4_PDDL/OWLS-TC4_PDDL/htdocs/ontology/Mid-level-ontology.owl
3	A doctor is an individual who practices medicine, which is concerned with promoting, maintaining or storing human health through the study, diagnosis, and treatment of disease, injury, and other physical and mental impairments.	http://semanticscience.org/ontology/sio.owl
4	A doctor is a general or specialized physician under medical practitioner subcategory of healthcare profession category in BioPortal.	https://bioportal.bioontology.org

VI. CONCLUSION AND FUTURE SCOPE

Ontology is the backbone for sharable knowledge representation, without utilizing its reusable property it is just a classical knowledge base. It is one of the possible proactive and preventive solutions of Semantic Inconsistency or conflicts among heterogeneous information system. In this paper Semantic Redundancy has been discusses from Ontology Reuse perspective and analysed that available Semantic documents are not practicing stated standards of Semantic Web Technology. Although reusing Ontology reduces redundancy, but the limitation of Ontology Reuse approach does not make it

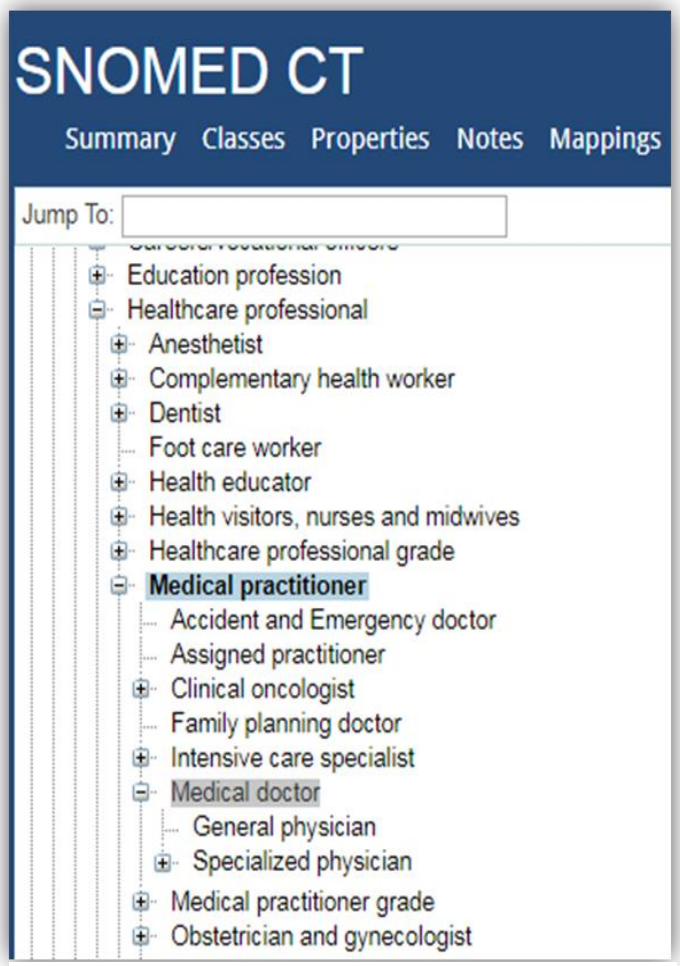

Figure 3. Terminological description of 'Doctor' in SNOMED CT [25]

best solution where reuse cost is higher and much complicated then creating new one. In near future, the authors are intending to increase the scope of Ontology reuse in Semantic Web services publication and discovery.

REFERENCES

[1] I. Akram, "Growth and Capital," 17 MARCH 2015. [Online]. Available: https://growthandcapital.com/2015/03/17/beyond-big-data-to-data-driven-decisions/. [Accessed 5 May 2016].

[2] J. Gracia and E. Mena, "Semantic Heterogeneity Issues on the Web," *IEEE INTERNET COMPUTING*, pp. 60-68, 2012.

[3] A. A. A. G. Ibrahim Ahmed Al-Baltah, W. N. W. Ab and R. a. R. Atan, "A Comparative Study on Ontology Development Methodologies towards Building Semantic Conflicts Detection Ontology for Heterogeneous Web Services," *Research Journal of Applied Sciences, Engineering and Technology*, pp. 2674-2679, 2014.

[4] E. Simperl, "Reusing ontologies on the Semantic Web: A feasibility study," *Data & Knowledge Engineering*, pp. 905-925, 2009.

[5] M. Katsumi and M. Gruninger, "What s Ontology Reuse?," *Formal Ontology in Information Systems, ISO Press*, pp. 9-22, 2016.

[6] M. Fahad, N. Moalla and A. Bouras, "Detection and resolution of semantic inconsistency and redundancy in an automatic ontology merging system," *Journal of Intelligent Information System*, vol. 39, p. 535–557, 2012.

[7] C. Ochs, Y. Perl, J. Geller, S. Arabandi, T. Tudorache, MA. Musen, "An Empirical Analysis of Ontology Reuse in BioPortal," *Journal of Biomedical Informatics*, pp. 165-177, 2017.

[8] J. Ye, L. Coyle, S. Dobson and P. Nixon, "Ontology-based models in pervasive Ontology-based models in pervasive," *The Knowledge Engineering Review, Cambridge University Press*, vol. 22, no. 4, pp. 315-347, 2007.

[9] L. Stoimenov, "Mediation and Ontology Based Framework for Interoperability," in *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends*, IGI-Global Press, 2009, pp. 491-507.

[10] A. Duque-Ramos, M. Boeker, L. Jansen, S. Schulz, M. Iniesta and J. T. Fernández-Breis, "Evaluating the Good Ontology Design Guideline (GoodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE)," 22 August 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/25148262. [Accessed 30 January 2017].

[11] A. Negi, M. Singh and P. Kaur, "Semantic Web: An Integration approach towards Intellectual Integration of Distributed Information," in *INDIACom-2017*.

[12] E. P. Bontas, M. Mochol and R. Tolksdorf, "Case Studies on Ontology Reuse," in *Proceedings of the 5th International Conference on Knowledge Management (IKNOW05)*, 2005.

[13] J. Heflin, "OWL Web Ontology Language: Use Cases and Requirements," W3C, ERKIM, Keio, Japan, 2004.

[14] J. Davies, D. Fensel and F. v. Harmelen, Towards the Semantic Web:Ontology-driven Knowledge Management, UK: John Wiley and Sons LTD, 2002.

[15] D. Park, H.-C. Kwon and S. Park, "Ontology for Semantic Representation of Marine Metadata," in *ICUIMC(IMCOM)'13*, Kota Kinabalu, Malaysia, 2013.

[16] A. Gómez-Pérez and M. D. Rojas-Amaya, "Ontological reengineering for reuse," in *International Conference on Knowledge Engineering and Knowledge Management*, 1999.

[17] D. N. Peralta, H. S. Pinto and N. J. Mamede, "Reusing a time Ontology," *Enterprise Information Sysytem*, pp. 241-248, 2004.

[18] T. Russ, A. Valente, R. MacGregor and W. Swartout, "Practical experiences in trading off ontology usability and reusability," in *12th Workshop on Knowledge Acquisition, Modeling and Management KAW*, 1999.

[19] M. Uschold, M. Healy, K. Williamson, P. Clark and S. woods, "Ontology reuse and application," in *First International Conference on Formal Ontology and Information Systems – FOIS*, 1999.

[20] A. A. A. B. J. C. T. B. W. I. Laresgoiti, "Ontologies as vehicles for reuse: a mini-experiment," in *10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop KAW*, 1996 .

[21] E. P. Bontas and M. Mochol, "A Cost Model for Ontology Engineering," Free University of Berlin, Berlin, 2005.

[22] P. Doran, V. Tamma and L. Iannone, "Ontology Module Extraction for Ontology Reuse: An Ontology Engineering Perspective," in *16th Conference on information and knowledge management, ACM*, Lisbon, Portugal , 2007.

[23] E. P. B. Simperl and M. Mochol, "Cost Estimation for Ontology Development," in *Business Information Systems BIS2006*, Poland , 2006.

Aradhana Negi is a Ph.D. scholar in the Department of Computer Science, Guru Nanak Dev University, India. She received her Master of Computer Applications from Guru Nanak Dev Engineering College, India. Her research interest includes Semantic Web, Automatic Semantic Web Service Discovery and Composition, and Cognitive Psychology in computing.

Parminder kaur is working as Assistant Professor in the Department of Computer Science, Guru Nanak Dev University, India. She has received her Ph.D. degree in 2011 from Department of Computer Science, Guru Nanak Dev University, India. Her research interests are Component-Based Software Engineering, Semantic Web Services, Version Controlling mechanism in Open Source Development System.