A Hybridized Framework for Ontology Modeling incorporating Latent Semantic Analysis and Content based Filtering

Pushpa C. N.
Department of CSE
University
Visvesvaraya
College of
Engineering
Bangalore
University,
Bangalore-560001

Gerard Deepak
Department of CSE
University
Visvesvaraya
College of
Engineering
Bangalore
University,
Bangalore-560001

Thriveni J.
Department of CSE
University
Visvesvaraya
College of
Engineering
Bangalore University
Bangalore-560001

Venugopal K. R.
Department of CSE
University
Visvesvaraya
College of
Engineering
Bangalore University
Bangalore-560001

ABSTRACT

In the era of Semantic Web, organization of the necessary Semantic Information becomes quite vital for improving overall retrieval efficiency of the Semantic Web contents. Ontologies are one of the most important and yet the most primary entities of the semantic web which is used for representing and modeling knowledge. Authoring of ontologies must be done in a highly systematic and an organized manner in order to validate the correctness of the ontologies authored. Several traditional ontology authoring systems are based on Semantic Wikis which use graphs to store the ontological entities that increase the overall complexity of ontologies which needs to be overcome. A Hash Table based ontology organization strategy is proposed which is further empowered by a Semantic Latent Analysis to compute the ontological relevance. Several agents are incorporated to check the correctness of ontologies. The proposed framework is further enhanced with Content Based Filtering for yielding better results. The proposed methodology yields an accuracy percentage of 88.99.

Keywords
Content Based Filtering, Hash Table, Knowledge Modeling, Ontologies, Semantic Latent Analysis, Semantic Web

1. INTRODUCTION
The World Wide Web is the largest storehouse of Information. It can be regarded as a single entity which stores enormous amounts of data and is expanding exponentially. Organization of information over the World Wide Web is highly essential. With the booming of the trends in Social Networks, a large amount of users' as well as data is being added into the World Wide Web. With increase in the amounts of data, retrieval as well as organization of such data has become cumbersome. Due to the continuous addition of the overall number of users' as well as the data contents to the Web, there is a need for proper structuring of the underlying identifiable data contents.

The urgent need for structuring the Web data and establishing a relationship among the data elements gave birth to an intelligent form of Web called the Semantic Web. Semantic Web mainly deals with the application of several Semantic Techniques to the Web existing web contents to either structure, organize or relate the existing Web data elements.

This further enables the Semantic Web retrieval algorithms to work efficiently on the structured, organized and relationship established data.

Ontologies form the most intrinsic and primary constituent of the Semantic Web structure. Ontologies form the core of the Semantic Web [1]. Ontologies can be related to a set of terms that are related and linked. Ontologies are explicit specification of shared knowledge. Ontologies can be added to the Semantic Web in several means. The most traditional method of including the ontologies into a specific domain is by ontology authoring, where ontology experts like Domain Experts and Knowledge Engineers are involved directly in contributing to building of ontologies. Ontologies can be conceived from the available Web data using expert systems for building ontologies, which to a certain extent automate the process of authoring ontologies. Ontology Modeling involves the overall process of ontology authoring, structuring the ontologies, establishing relationship among the ontological entities and axiomitization of ontological entities.

The ontologies authored or accepted by an ontology authoring system must be validated for the correctness of ontologies and their best fit into a Domain of relevance. Moreover, an ontology processing and modeling system must validate the correctness of the ontologies and their relevance at least into a Domain. Ontology Modeling is never a single step process rather, it is an iterative process which, comprises of several sequences of validation. Most ontology authoring systems mainly concentrate on the methodology of viewing the ontological entities and on the processes adapted to organize the ontologies. The traditional ontology modeling systems always rely on human resources like the Domain Experts and the Knowledge Engineers to validate the acceptance of the ontologies into the domains. Only a very few ontology modeling systems focus on inclusion of a specific paradigm to compute the relevance of ontologies, to be organized into a system.

The quality of ontologies that are accepted into a system matters the most. Only if the ontologies modeled are of a high quality, the algorithms dependent on them would perform well. Irrelevant ontologies themselves would serve as a threat, if they are not placed properly in the Domain of relevance. Since the Domain Experts and Knowledge Engineers are the source of ontologies, their presence is required but their

consent is always not needed for validating the domain relevance of ontologies. The interaction of the ontology authors can be minimized by incorporating agents as well as semantic techniques into the ontology authoring systems.

Ontology authoring became robust and interactive when the collaboration between the Domain Experts and Knowledge Engineers [2] was introduced. This enhanced the quality of the ontologies as the Domain Experts always had a check on the ontologies that were authored. Domain Experts are skillful in constructing narrow ontologies [3]. Domain Experts serve as an excellent source of ontologies due to their expertise in a specific domain but their presence in the system for even computing the correctness of ontologies and their relevance can be minimized by employing agents and efficient semantic strategies. It is the role of semantics which enable the extraction of core [4] ontologies. An ideal ontology authoring system must be able to organize the ontologies with high efficiency and reduce the complexity of authoring and storage. Additionally, it must include a proper logic formulation and axiomitization with a high level of accuracy. Ontology Modeling Systems are an important class of Information Retrieval Systems with the main focus on Knowledge Management.

Motivation: Most of the traditional ontology authoring systems is based on the usage of Semantic Wikis. Wikis make use of a graph based structure for storing the concepts and individuals of ontologies which increases the complexity of the system. There is a need for ontology authoring system which does not make use of Wikis and yet facilitate easy interactivity. Most ontology authoring systems need the ontology authors to play a vital role even after they have contributed their part in authoring ontologies, ie, they need to be a part in checking the correctness of ontologies. There is a mandatory need for an ontology authoring system which would validate the ontologies automatically. Moreover, a system which amalgamates the ontologies from ontology authors as well as from web sources is required to increase the density of ontologies in a specific domain. The correctness and the quality of ontologies authored by a system must be maintained very high. The process of ontology approval into a domain needs to be governed by an efficient semantic technique for minimizing the participation of DEs.

Contribution: An ontology authoring and modeling system has been proposed which replaces the Semantic Wikis. Instead of using Semantic Wikis which stores the ontologies as graphs, the ontologies are structured using Hash Tables that are parsed using lexical analysis. A Strategic Semantic Latency computing Methodology is proposed for validating the domain relevance and correctness of the ontologies. A Multiagent approach is followed where several agents with different behaviors are incorporated into the system for achieving quality checks. A content based filtering approach which is driven by semantic latency methodology is proposed for acquiring the ontologies from related Web Documents. A synonym computation strategy is implemented for increasing the domain population of ontologies, thereby increasing the ontology density A logic induced Axiomitization is incorporated for relationship establishment and structuring the ontologies.

Organization: The remaining paper organization is as follows. The Section 2 provides a brief overview of Related Literature of the research conducted. Section 3 presents the Problem Definition. Section 4 depicts the Proposed Architecture. Strategy Incorporated is discussed in Section 5. Section 6 describes the implementation in detail. The Results

and Performance Analysis is discussed in Section 7. Finally the paper is concluded in Section 8.

2. RELATED LITERATURE

Dellschaft et al., [5] has proposed a strategy of Cicero tool which enhances the convenience of ontology authoring. The underlying methodology in this approach is that the facility to provide more documentation while ontology authoring such that the actual meaning of ontology is clearly portrayed in this approach. Moreover, a forum is created such that a discussions based ontology editing is facilitated in this proposed work. Such an approach enhances ease of ontology editing and editing ontologies becomes a simple and yet an important approach. A DILIGENT argumentation model based on Potts and Bruns model wherein issues or potholes in the domain where shallow ontologies are possible is estimated and artifacts are developed for the same initially. Based on these artifacts for resolving issues, ontologies are authored and developed which becomes an innovative methodology for ontology authoring. The drawback for such approach is that a complete baseline work has to be done before ontologies are developed for a domain. An ideology for a domain must be a preconceived entity which definitely is not possible when a dynamic scheme for ontology definition is considered.

Wang et al., [6] have carried out a detailed study of amalgamating semantic ontologies with software engineering paradigms. The major investigations were focused on the reusability of knowledge by using strategic UML for modeling of ontologies. The major domain that was used for study was the Petroleum Remediation domain where the UML modeling was carried out to study the phenomenon of ontology modeling. Several Software Engineering paradigms like abstraction were studied for similar inclusions into process of Knowledge Representation and Modeling. The methodology followed is a detailed investigation of the entities in the petroleum domain. The evident cognitive gap between strategic software engineering models and promising ontology authoring models is bridged.

Kim et al., [7] have proposed a knowledge map framework for knowledge representation and modeling of multi-domain ontologies. The national R&D data is used as the major domain of data and knowledge map service is developed using lightweight ontology modeling strategies. The methodology focuses on representing simple document centric relationships like co-author and co-topic relationships. Several project-output, document-authors relationships are included in this strategy. A triplet generation strategy is incorporated in this approach for eliciting the relationships. Furthermore the data visualization and document level navigation are achieved in the framework proposed.

Fan et al., [8] have proposed a strategic approach for business processing modeling using ontologies. The methodology that is incorporated here is the Process Ontology Based Approach wherein process modeling is done by specifying the ontologies for individual processes in a workflow. The Bunge-Shanks Framework is successfully incorporated for representing and modeling complex business ontologies and overcome semantic ambiguity that exists in the present frameworks. The semantic information for individual process workflow is modeled as business ontologies by information harvesting methodology. This clearly overcomes the contradiction between the business concepts and the final outcome of ontologies is quite satisfactory in this method.

Bao et al., [9] have proposed a strategic model for ontology modeling using the Big Data Framework. In this methodology

ontologies are conceived using a system called Big Data Knowledge Management System (BDKMS) which is fully driven by data. The principle of dynamic updation of ontologies and service oriented publication makes this approach more robust and quite interesting. A high level of intelligence is incorporated in this strategy of ontology modeling. The domain considered for experimentation is the steel manufacturing process domain which gives quite efficient and pleasing results. Ontology Reusability and Knowledge Reusability are the two major criterions which are focused in this methodology. The approach focuses on knowledge level inference driven by semantic principles. The Big Data Oriented Framework enhances the performance efficiency of the proposed framework.

Rathi et al., [10] have proposed a standardized data format incorporating Web Ontologies. A MostoBM benchmark is adapted and rule based induction is followed for modeling of ontologies. An efficient framework is built for data interchange which is in par with the standard benchmark systems. Relations Mapping and usage of metadata is the characteristic feature of this approach. The problems of evolution of ontologies, publishing of open linked data and Vocabulary adaption is achieved in this approach. The scaling of patterns is achieved which enhances the overall performance of the system. This strategy serves as a lone strategy for testing the data exchange systems driven by ontologies.

Sunitha Abburu et al., [11] have proposed a framework for Knowledge Management using ontologies. The proposed framework is Triadic in nature which incorporates Knowledge Acquisition, Knowledge Storage and Knowledge reusability. The knowledge acquisition is from knowledge sources that are categorized into Learning and Non-Learning objects. The acquired knowledge is stored in a local repository which is the Oracle DB. A triplet RDF is used and querying is done using a SPARQL query on the acquired knowledge. The framework for Knowledge Management is proposed to depict that knowledge can be efficiently reused using ontologies. Domain Level ontologies used and the proposed framework addresses to the traditional problems of knowledge management systems.

Rivero et al., [12] have proposed a methodology for mapping RDFs in Knowledge Bases using certain samples. The concept of schema mapping is proposed where the relationships are explicitly expressed. The proposed framework automatically generates schema mappings and associations for RDF Knowledge Bases in the Web of Data. The framework is quite appealing as the schema mapping samples are generated automatically by analyzing the RDF data which is of triplet nature. The associations between the data in the Web of Data are derived on analysis. A single input exchange format is incorporated and the standards are well evaluated to support this framework.

Park et al., [13] have proposed a technique for classification of personal media automatically using an ontological approach. The proposed approach incorporates rule based inference for processing of events. Media Analysis is integrated into the proposed approach where recognition technology is imposed for recognition of useful objects present. In the proposed methodology, automatic video classification is achieved using logic based reasoning and rule based inference. The objects that are considered are video

tags, titles of videos and the script details. A description logic is imbibed which not only enhances the performance of the proposed approach but also increases the level of confidence yielding results up to the mark.

Christopher et al., [14] have proposed a methodology for the inference of ontological structures for using a k-BOOM which is a specialized Bayesian Approach. The strategy for ontologies building includes domains like anatomy and diseases to interleave existing knowledge. Since mappings are loosely coupled, precise logical relationships must be elicited for translation of mappings into precise and logical relationships. A non-trivial translation is proposed which uses hypothetical axioms to aggregate semantically loose mappings together and knit them. A deductive reasoning and probabilistic reasoning methodology called BOOM is proposed where probabilistic ontologies are divided into k sub modules using a lexical technique.

Carvalho et al., [15] have proposed a methodology for uncertainty management in Semantic Web Systems. The uncertainty is represented using Probabilistic Ontologies in a Semantic System. The construction of probabilistic ontologies is not given prime importance. To overcome this proper construct method which constructs probabilistic ontologies is proposed. A PR-OWL probabilistic ontology language is proposed for uncertainty modeling for a specific domain. This methodology is applied to several scenarios in the domain and gives a clear depiction for understanding the representation of uncertainties.

3. PROBLEM DEFINITION

Modeling of ontologies is definitely not a single step process but it requires a high degree of organization, skill and resources. Modeling of ontologies is the first and yet the most important step of Ontology and Knowledge Management. The visible problems can be described as the lack of a properly organized methodology for ontology modeling. The modeling of ontologies needs a lot of involvement of the ontology authors which makes the process more users' dependent which definitely needs to be minimized. Another evident problem prevalent in ontology modeling is the increase in the complexity of the storage of ontologies when Semantic Wikis are employed for authoring of ontologies.

The modeling of ontologies must be independent of semantic wikis and must be able to organize the ontological hierarchies quite effectively with lesser complexity. An intrinsic and automated conflict resolution platform is typically absent in the existing ontology authoring tools. The conflicts with reference to the correctness of domain relevance of ontologies must be resolved. There is a need for a concrete semantic strategy for processing and validating the correctness of the ontologies. The correctness of ontologies mainly influences the quality aspect of authored ontologies. The existing cognitive gap between the manual and automated ontology authoring systems must be reduced by automating most of the stages in ontology modeling process.
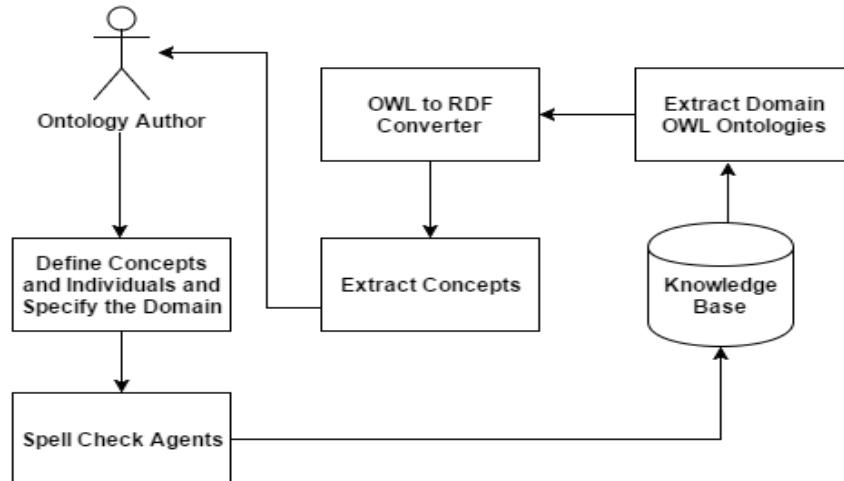
4. PROPOSED ARCHITECTURE

Figure 1: Domain Ontology Modeling

The architecture of the proposed system bridges the cognitive gap between manual and automatic ontology authoring and modeling systems. The organization of the proposed architecture is divided into three different phases that are integrated together. The first phase depicted in Figure 1 is the Domain Ontology Modeling which allows the system to interpret and view the OWL ontologies that are stored in the Knowledge Base. Likewise it allows the ontology authors to define concepts and individuals which are validated using the spell check agent. In the first Phase, the system allows the ontology authors to extract the Domain ontologies from the Knowledge Bases that are generally stored in .owl format.

The OWL ontologies are parsed using Lexical Analysis. Parsing allows the system to understand the organization of the ontologies in the .owl file. Further the domain ontologies are subjected to OWL to RDF transformation from which the concepts and individuals are extracted. The extracted concepts and individuals are furnished to the user. The user can be any ontology author who can either be a Domain Expert or a Knowledge Engineer. The ontology authors need not have any technical knowledge to define the ontologies but rather be equipped only to contribute the ontological entities and be able to relate to them.

Phase 1 allows the ontology authors to define the concepts and individuals and also specify the domain where they should belong to. As the concepts and individuals are defined by the ontology authors, they are subjected to a spell check agent which corrects the spellings of the concepts or individuals in case there are any. This is the first step of validating the correctness of the ontologies. Phase 1 ends with storing the defined ontologies which have no spelling errors into the Knowledge Base.
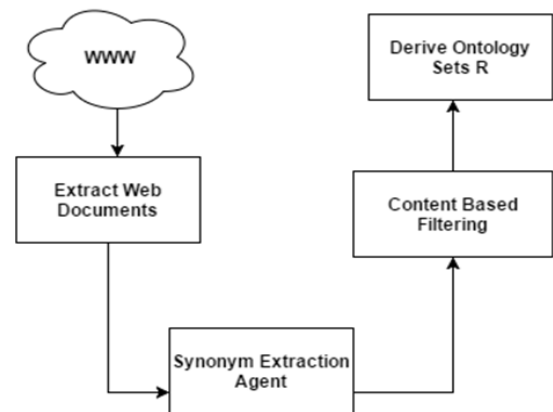

Figure 2: Content Based Filtering on Web Documents

The second phase depicted in Figure 2 of the system mainly focuses on the content based filtering on Web documents to formulate the ontologies that combine synonym extraction using an agent. In Phase 2, the Web Documents are extracted from the World Wide Web and is further subjected to a Synonym Extraction Agent. Furthermore, the ontology set undergoes expansion by extraction of synonyms from the ontological entities or linked data in the World Wide Web. The system furthermore has a provision for automatically including the domain relevant ontologies from the documents over the World Wide Web. A content based sematic filtering strategy is proposed by using the contents in the intermediate ontology set. The ontologies added from several web sources and the ontologies defined by the ontology authors are stored as an ontological set which is again made visible to the ontology authors for logic input. At the end of second Phase, the Ontology Sets with synonymous entities are available.
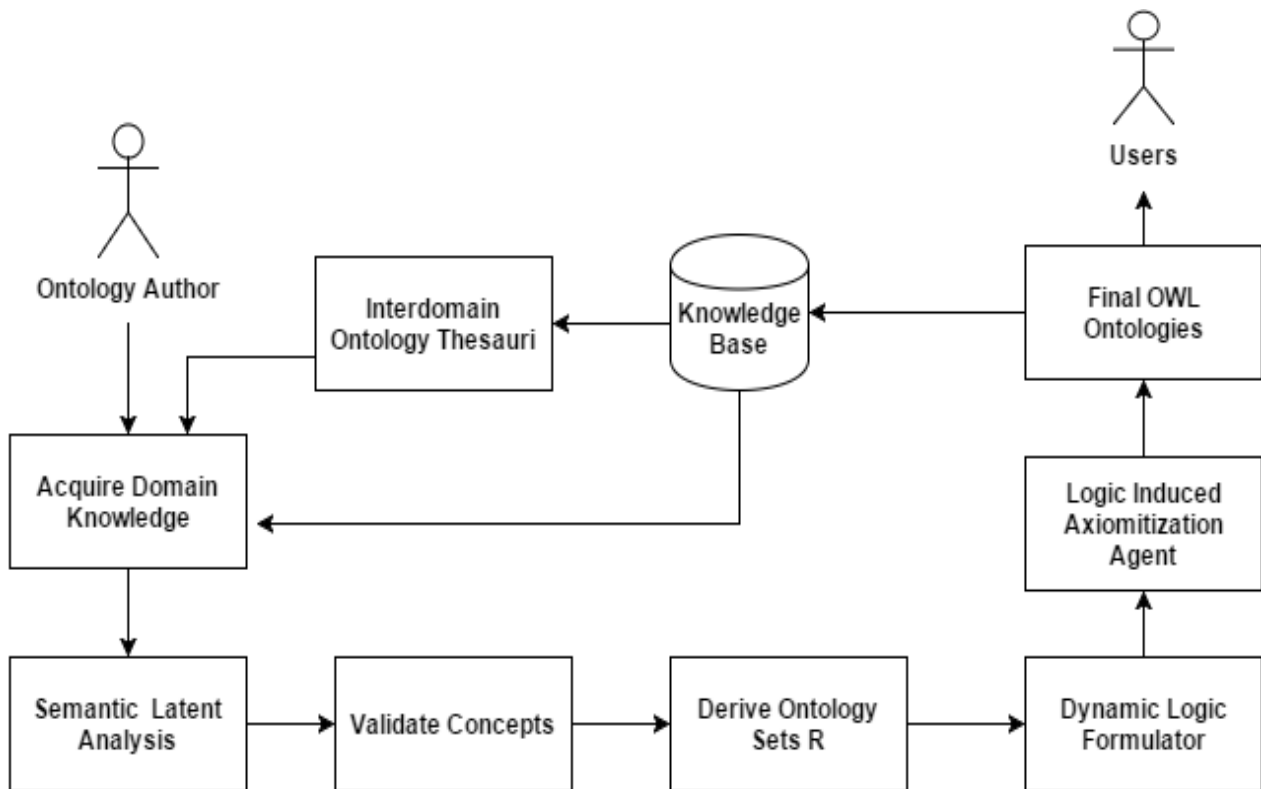
Figure 3: Semantic Latent Analysis on Domain Knowledge

The third phase illustrated in Figure 3 focuses on the Latent Semantic Analysis on the Acquired Domain Knowledge. A very important aspect of acquiring Domain Knowledge is the primary functionality in Phase 3. Acquiring domain knowledge refers to dynamically loading the principal ontologies namely the first two and the last two ontologies from the Hash Table structure of a specific domain. Once the concepts are defined and the required domain knowledge is acquired, Semantic Latency Analysis is applied in order to compute the Semantic Heterogeneity. The Semantic Deviation is computed between the defined concepts and the acquired concepts in the Domain Knowledge. The Semantic Deviation value will decide if the defined concepts are best fit into the Domain or not.

The concepts within the threshold value of semantic deviation are validated and included into the domain. The concepts or individuals are formulated into an ontology set before they undergo Axiomitization. A dynamic logic formulation technique is followed based on the already available inter domain ontology thesauri or the domain knowledge. Furthermore, based on the logics formulated or accepted from the ontology authors, logic induced Axiomitization is done where axioms definition agents structure the ontological entities. The Axiomitization of ontologies is done by inducing the rules between the associations of ontologies stored in hash table. This definitely overcomes the traditional approach of using Semantic Wikis which use a graph based structure and thereby reduces the overall complexity in organizing the ontologies. The final Axiomitized ontological entities are saved in Knowledge Bases, Inter-domain ontological Thesauri and are also sent to the user.

5. STRATEGY INCORPORATED
Latent Semantic Analysis
The strategy incorporated is the Latent Semantic Analysis or the Semantic Latency Approach to compute the Semantic

Similarity between a set of terms. In this case, the terms used are ontological entities namely the concepts or individuals which either contributed by the Ontology Authors or acquired from the Knowledge Bases. The semantic Latency measurement is a methodology of Distributional Semantics where a Term Count Matrix is formulated. In this work, the ontology occurrences are counted within the existing Relevant Knowledge stored in Knowledge bases and is represented in the form of a Matrix. Furthermore, Singular Value Decomposition technique is applied to the Term Count Matrix to minimize the number of rows in the Original Matrix. To the resulting matrix, the cosine of the resultant angle between a pair of vectors is computed to finally deduce the semantic similarity. The step by step procedu.re to compute the Latent Semantic Heterogeneity is illustrated in Table 1.

Table 1: Computation of SLA using Semantic Latent Analysis

Step 1: The Ontological entity O input is used to count the term frequency in the underlying data set, ie, the Documents in the Knowledge Base.
Step 2: Formulate, the Term Frequency Matrix T where the element (x, y) depicts the occurrence of Ontology Term O_x in the existing relevant Knowledge Entity y.
Step 3: Perform Singular Value Decomposition (SVD) to T and reduce the total number of rows preserving the column structure.
Step 4: Find the Cosine Similarity of the angle between the two resulting vectors to deduce the Semantic Similarity.

6. IMPLEMENTATION
The implementation of the proposed system was successfully completed in JAVA using Net Beans as an IDE. A modularity approach was followed where individual functionalities were

developed as separate functional Java classes were further integrated. A Model View Controller design pattern was followed. The centralized Knowledge Base served as a large repository for OWL Domain ontologies where there were .owl files of several domains. 18 distinct and independent domains level .owl files were used for experimentation. The details of .owl Domains used for experimentation comprised of several concepts. Each of the concepts were associated with a specific number of individuals. The details of the number of concepts and individuals in the .owl domain ontologies are depicted in Table 2.

For segregating the concepts and individuals from the OWL domains, a combination of simple parsing and lexical analysis is done. The lexical analysis is done using JAVA Scanner classes including a HashMap, an ArrayList and an Iterator for repeating the process until all the contents are parsed and lexically analyzed. The OWL to RDF is further realized in JAVA by instantiating the predefined java class *"RDFXMLOntologyFormat"* that belongs to the inbuilt java package *"org.semanticweb.owlapi.io"*. The ontological classes and individuals are yielded to the ontology authors, so that they will be able to view the ontological contents. The spell checker agent is designed using *Jazzy Open Source Java API* for checking the spellings. The Java code for Semantic Latent Analysis is written by implementing the procedures depicted in Table 1 systematically and is realized as a separate reusable class. The HashSet collection framework of Java is used for storing the formulated ontology set. The UI interface is designed using JAVA Swings as well as HTML.

A content based filtering strategy is incorporated for downloading text documents containing the ontological terms in the intermediate ontology formulated set. Further, the context text of the content term is extracted. The context text refers to the sentence embedded between white spaces or full stop notation where the term content is present. The entire text is further parsed and processed for semantically similar terms using Latent Semantic Analysis method. The semantically similar terms are added to the ontology set that was formulated.. A Synonym extraction agent which was designed in Java was incorporated in order to enhance the number of similar ontologies in the ontology set before Axiomitization.

Table 2: Details of Domain Ontologies

OWL Ontologies	No. of Concepts	No. of Individuals
Cars Ontology	36	81
Radio diagnosis Ontology	18	56
Ice Cream Ontology	12	42
Furniture Ontology	18	62
Finance Ontology	42	94
Biomedical Ontology	20	34
Dog Breeds Ontology	14	31
Shopping Ontology	15	63
Mathematical Ontologies	9	46
Pharmacology Ontology	10	44
Soil Ontology	24	65
Comedy Movies Ontology	8	26
Geology Ontology	12	63
Computer Science Ontology	5	18
Environment Ontology	7	24
Microorganisms Ontology	3	36
Weather Ontology	5	62
University Courses Ontology	4	24

The Axiomitization is induced on the ontological concepts as well as the individuals. Initially each individual is linked to its specific concept and their relationship is expressed. Once the individuals are associated with their respective concepts, then the logics are enforced on the concepts. The logics are either requested from the Domain Experts where the Domain Experts have a choice to enter logic or review the logics entered by other ontology authors. The system is capable of identifying and inferring the basic logics by analysis of logics in the inter domain ontology thesauri. If a logic is not enforced or input by a Domain Expert, then the system is capable of Dynamic Logic Formulation where the logics of similar terms in the inter domain ontology thesauri is learnt and a similar logic is induced. The Axiomitization is implemented by already stored basic logic rules in the database. The Axiomitization is done using a logic enforcement agent which is designed in Java. The state of Axiomitization agent is to identify two concepts that are related and the behavior is to enforce axioms and rules between the concepts.

Table 2: Proposed Algorithm for Ontology Modeling

Input:	Initial user specified concepts C_d or individuals I_d. existing concepts C_e and individuals I_d in OWL Domain ontologies.
Output:	Axiomitized and Organized Final Domain Level OWL ontologies.

Begin

Step 1: Specify the Domain of Choice D_l for viewing or authoring the ontologies.

Step 2: The .owl Domain Level Ontologies of D_l is loaded and parsed.

Step 3: The Domain Level OWL ontologies are subjected to parsing and Lexical Analysis for identifying

the concepts C_e and individuals I_e

Step 4: The OWL files are converted into its RDF format and the concepts C_e and Individuals I_e are stored in a hash table.

Step 5: Ontology Authors define the concepts C_d and Individuals I_d explicitly whose spell check is conducted and validated using Jazzy Agent.

Step 6: The Domain Knowledge of D_l is acquired where the principal ontologies P_o of D_l is loaded.

Step 7: The Semantic Latency Analysis SLA is computed between C_d and concepts of P_o or between I_d and Individuals of $P_o.$

Step 8: If SLA < 0.35, then add C_d or $I_{d.}$ into the Arbitrary Set R.

Step 9: Expand R using Synonym Extraction Agent such that SLA value of a term T in R and its Synonym T' <0.15

Step10: Incorporate Content based filtering using term T in Extract the context of a T and tokenize t_k in it. Compute the SLA for $t_k.$ If SLA < 0.35, then add t_k into R.

Step 11: Using the logic specified by the ontology authors or by dynamic logic inference, the elements of R are Axiomitized.

End

The Algorithm depicted in Table 2 is a chronological algorithm which describes the steps in a sequential manner for authoring and modeling the domain specific ontologies. The algorithm is an intelligent amalgamation of techniques for extracting the already existing ontology entities, defining and validating new ontological concepts and individuals and conceiving ontologies automatically through context oriented content based filtering. The threshold value of 0.35 or 35% semantic similarity is considered. The reason for considering 0.35 is that this range allows quite a good number of ontologies and it's a good threshold value for semantic similarity where the quality of ontologies can be regulated. The threshold of 0.15 or 15 % similarity is considered for synonymous terms as the number of synonyms should be more and the similarity in the terms must be preserved. The algorithm specified in Table 2 is implemented as a framework in Java.

7. RESULTS AND PERFORMANCE ANALYSIS

The performance of the proposed system is evaluated considering Recall, Precision and Accuracy as Metrics. The metrics of Recall and Precision originally belonged to a retrieval system but this does not mean that they may be restricted to a retrieval system alone. The Precision and Recall can be associated to any system (preferably an information system) which is driven by an algorithm or modeled on a framework but it must involve a change in state of that system once the algorithm is applied or even on interaction with the system. Precision in common terms can be interpreted as the correctness of the system and Recall in general can be inferred as the relevance of instances.

Since the proposed Knowledge Modeling System is a class of Information Systems, the parameters chosen for metrics are justifiable. The Precision P for the proposed approach is defined as the measure of the ratio of the count of relevant and accepted ontologies into the ontology authoring system to the total number of ontologies accepted by the system. The Recall R is defined as the ratio of the relevant and accepted ontologies into the system to the total number of relevant ontologies allowed into the system. The Accuracy is the average of Precision and Recall of the system. The Precision, Recall and Accuracy are depicted in equations (1), (2) and (3) respectively.

$$P = \frac{\text{No. of Relevant and Accepted Ontologies into the system}}{\text{Total No of Ontologies Accepted by the system}} \quad (1)$$

$$R = \frac{\text{No. of Relevant and Accepted Ontologies into the system}}{\text{Total No of Relevant Ontologies Authored}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{Precision} + \text{Recall}}{2} \quad (3)$$

Table 3: Performance Evaluation of the Proposed Methodology incorporating Semantic Latent Analysis

Domain Ontologies	Precision %	Recall %	Accuracy %
Cars Ontology	87.88	91.58	89.73
Radiodiagnosis Ontology	87.33	91.18	89.26
Ice Cream Ontology	87.80	90	88.9
Furniture Ontology	86.21	90.36	88.29
Finance Ontology	89.29	91.91	90.6
Biomedical Ontology	86.67	89.66	88.17
Dog Breeds Ontology	86.36	90.48	88.42
Shopping Ontology	87.5	89.74	88.62
Average	**87.38**	**90.61**	**88.99**

Although the experimentation was done by considering 18 distinct domains, only the first 8 domains are considered for evaluation and tabulation of the performance. The reason behind this is to keep the evaluation simple and even if the performance of all 18 domains is considered and tabulated, the overall average performance just differs in 2 decimal points. This is the reason for considering the first 8 domains. Table 3 depicts the Precision, Recall and Accuracy for the proposed system. Moreover, the usage of multiple agents for

authoring of ontologies increases the correctness of ontologies. The spell check agent ensures the correctness of the suggested ontology terms. The Axiomitization agent induces the logics inferred and enhances the process of ontology authoring. The synonym extraction increases the number of related and similar concepts thereby increasing the density of semantically related ontologies.

The performance is evaluated based on the metrics chosen and is depicted in Table 3. The difference between the precision, recall and accuracy between individuals is not very high. This proves that the proposed algorithm is efficient and performs without any major deviation. This justifies the reason for choosing only the first 8 domains for performance evaluation. The proposed methodology yields an average precision percentage of 87.38. The overall recall and accuracy of the proposed methodology are 90.61% and 88.99 % respectively. From the overall values of precision, recall and accuracy, it can easily be inferred that the correctness of the strategy is quite high. This proves that the inclusion of the semantic latency strategy for analyzing the acquired domain knowledge increases the relevance of the ontologies modeled by the system.

Table 4: Performance Evaluation of the approach without Semantic Latent Analysis

Domain Ontologies	Precision %	Recall %	Accuracy %
Cars Ontology	82.35	84.85	83.6
Radio diagnosis Ontology	81.38	83.33	82.36
Ice Cream Ontology	84.21	88.89	86.55
Furniture Ontology	82.35	87.5	84.93
Finance Ontology	82.71	85.27	83.99
Biomedical Ontology	80.77	84	82.39
Dog Breeds Ontology	84.61	86.84	85.73
Shopping Ontology	81.82	85.14	83.48
Average	**82.53**	**85.73**	**84.13**

In order to show that the proposed methodology which incorporates Semantic Latency Analysis is better than a methodology which does not use a semantic strategy, the experimentation is again conducted for the same domain ontologies in the absence of Semantic strategy. Table 4 depicts the Performance of the framework without Semantic Latency Analysis. The percentage of precision, recall and accuracy is 82.53, 85.73 and 84.13 respectively for the approach which does not incorporate Semantic Latency Analysis. Figure 5 depicts the comparison of the proposed ontology modeling strategy that incorporates Semantic Latent Analysis with the same strategy which does not incorporate Semantic Latent Analysis. This indicates that the inclusion of a Semantic Strategy like Semantic Latency Analysis increases the overall performance of modeling the ontologies. Moreover, the overall quality of ontologies that are modeled by this system are increased due to the inclusion of a Semantic Methodology.
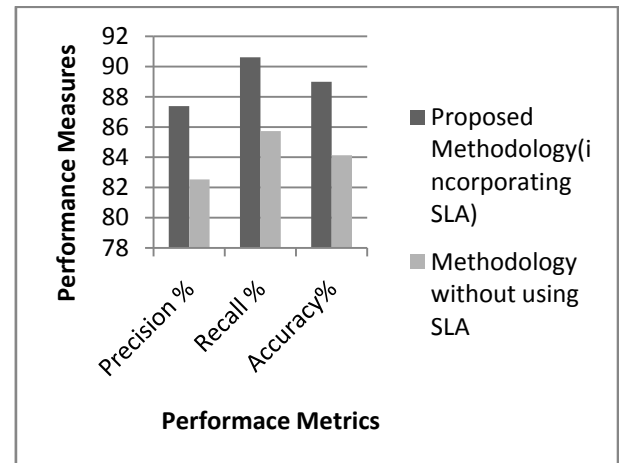
Figure 5: Comparison of the Performances

8. CONCLUSIONS

The traditional problem of storing the ontological entities in a hierarchical manner is overcome by replacing the graph driven Semantic Wikis with a Hash Table. The process of lexical analysis that is incorporated facilitates easy parsing of ontologies. The cognitive gap between the manual ontology authoring and automated conception of ontologies is reduced in the proposed methodology. The incorporation of Semantic Latent Analysis computes the semantic heterogeneity between the existing and authored ontologies, thereby increasing the relevance of ontologies accepted by the system. The usage of agents accelerates the process of ontology authoring in the proposed methodology and ensures the validity and correctness of ontologies. The Domain Knowledge Analysis and logic induced axiomitization are major highlights in the approach as they ensure the further credibility of ontologies. The proposed approach yields an average precision percentage of 87.38, recall percentage of 90.61and an accuracy of 88.99% which is the best performance achieved by such knowledge modeling systems.

9. REFERENCES

[1] Ding, Li, Pranam Kolari, Zhongli Ding, and Sasikanth Avancha "Using Ontologies in the Semantic Web: A Survey" in Ontologies, pp. 79-113. Springer US, 2007.

[2] Chiara Di Francescodimarino, Chiara Ghidini, Marco Rospocher, "Evaluating Wiki Collaborative features in Ontology Authoring," in IEEE Transactions on Knowledge and Data Engineering, vol.26 issue 12, pp. 2997 – 3011, 2014.

[3] Zhitomirsky-Geffet, M., Erez, E.S. and Judit, B.I., "Toward Multiviewpoint Ontology Construction by Collaboration of Non-Experts and Crowdsourcing: The Case of the Effect of Diet on Health" in Journal of the

Association for Information Science and Technology, 2016

[4] M. Krotzsch, D. Vrandecic, M. Volkel, H. Haller, and R. Studer, "Semantic Wikipedia," Journal of Web Semantics, vol. 5, pp. 251– 261, 2007.

[5] Dellschaft, K., Engelbrecht, H., Barreto, J.M., Rutenbeck, S. and Staab, S., "Cicero: Tracking Design Rationale in Collaborative Ontology Engineering", Springer Berlin Heidelberg, pp. 782-786, 2008.

[6] Wang, X. and Chan, C.W., "Ontology Modeling using UML" in OOIS Springer London, pp. 59-68, 2001.

[7] Kim ST, Lee WG "Knowledge Map Service based on Ontology of Nation R&D Information" in the Journal of Digital Convergence vol.14, issue 3, pp. 251-60, 2016.

[8] Fan, S., Hua, Z., Storey, V.C. and Zhao, J.L., "A Process Ontology Based Approach to Easing Semantic Ambiguity in Business Process Modeling" in Data & Knowledge Engineering, pp.57-77, 2016.

[9] Bao Q, Wang J, Cheng J. "Research on Ontology Modeling of Steel Manufacturing Process Based on Big Data Analysis" In MATEC Web of Conferences, vol. 45 EDP Sciences, 2016.

[10] Rathi, Snehal, Jyoti Shinde, Akshata Sonwane, Mohini Suryawanshi, and Mital Shah "Standardizing Data Exchange using a Benchmark in the Field of Semantic Web Ontologies" International Journal of Research in Advent Technology and the proceedings of National Conference NCPCI-2016, pp.200-204 19, 2016.

[11] Sunitha Abburu,, and G. Suresh Babu "A Framework for Ontology Based Knowledge Management" in the International Journal of Soft Computing vol. 3, no. 3, pp.21-25,2013.

[12] Rivero, C.R., Hernández, I., Ruiz, D. and Corchuelo, "Mapping RDF Knowledge Bases Using Exchange Samples" in Knowledge-Based Systems, 93, pp.47-66, 2016.

[13] Park, Hyun-Kyu, Chi-Seung So, and Young-Tack Park "Ontology Modeling and Rule-based Reasoning for Automatic Classification of Personal Media" in the Journal of KIISE, Vol.43, Issue 3 pp. 370-379, 2016.

[14] Mungall, Christopher J., Sebastian Koehler, Peter Robinson, Ian Holmes, and Melissa Haendel "k-BOOM: A Bayesian Approach to Ontology Structure Inference, With Applications in Disease Ontology Construction." in bioRxiv, 2016.

[15] Carvalho, Rommel N., Kathryn B. Laskey, and Paulo CG Da Costa. "Uncertainty Modeling Process for Semantic Technology" in PeerJ Preprints, 2016.