International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-,December 2015
ISSN: 2395-3470
www.ijseas.com

Ontology Model for Phylogeny Visualization

M.Sharmila[1],
M.Phil Scholar,
sharmimanohar@gmail.com,
Department of Computer Applications,
Bharathiar University, CBE.

Dr.V.Bhuvaneswari[2],
Assistant Professor,
bhuvanes_v@yahoo.com,
Department of Computer Applications,
Bharathiar University, CBE.

ABSTRACT

Phylogeny is the genealogical study of living and non living organisms. Phylogeny represents the historical pattern of relationship among organisms that has genealogical unity of given hominidae species like, human, gorilla, chimpanzee, and orangutan. The Ontology is a branch of metaphysics concerned with the nature and relations of being. Ontology is a specification of entities and their relationships. The research analyses the ontology for phylogeny visualization. The purpose of the research is easy to understand the phylogeny visualization. This research has performs various ontology terminologies and DL Query. The comparison of the existing phylogeny visualization and proposed phylogeny visualization using ontology is carried out. The proposed ontology based phylogeny model the extraction of related organisms based on gene functionality and gene can be done using DL Query which is not possible in traditional Phylogeny methods.

Keywords: *Phylogeny Visualization, Ontology, DL Query.*

I. Introduction

Bioinformatics is an interdisciplinary field that develops and improves methods of storing and retrieving of biochemical and biological data using mathematics and Computer Science [16]. Phylogenetic tree diagram shows the evolutionary interrelations of a group of organisms that usually originated from shared ancestral form [5]. Phylogenetic analysis is the inferring or estimating evolutionary relationships among organisms. The result of an analysis is drawn in a Cladogram diagram called a cladistics. Cladistics is the classification of organisms based on the branching of descendent lineages from a common ancestor. The organisms in Phylogeny Visualization methods provide are related with other biological entities like gene, protein and involved in other activities. The existing phylogeny visualization method does not relate the organism with any other related biological entities.

Ontology is the philosophical study of the nature of being, becoming, existence or reality and the basic categories of being and their relations. Ontology collects the set of individual instances of the kinds of entities it specifies constitutes a knowledge base. It is a major bioinformatics project to unify the illustration of gene and gene product qualities across all species [12]. The advantages of using ontologies have been argued extensively, but the main reason is that ontologies are attempting to capture the precise meaning of terms. The creation of domain ontology is also fundamental to the definition and use of an enterprise architecture framework.

In this paper ontology model for phylogeny visualization is proposed. The organization of the chapter includes the second section reviews various methods available from the literature for Phylogenetic methods and

visualization tools, Gene Ontology. The third section includes the Ontology Model for Phylogeny Visualization. The fourth section the experimental results of this work are discussed. Various screen shots are discussed in detail. The fifth section includes summaries the research work, contribution to the domain knowledge and direction of future work.

II. Review of Literature

The Ontology model consists of various reviews.

B.Orgun et.al [2] proposed a work for providing interoperability among domain ontologies. They discussed about some key issues are that still need to be addressed if there to move from semi to fully automated approach to provide consensus among heterogeneous ontologies. The issues outlined are addressed in order to establish a generic, domain independent, fully automated approach interoperability across heterogeneous ontologies. Dan He [3] proposed an ontology based feature weighting strategy for text classification. The ontology based likelihood functions of features can be computed with the combination of their corresponding layers in the given ontology and the original feature frequency under bag-of-words representation. Purvesh Khatri [11] presented the several analysis tools. An automatic ontological analysis approach has been proposed for biological analysis of results. The comparisons of several analysis tools have a number of limitations and drawbacks. The result shows that the large number of tools implementing every similar approach. Kaustubh Supekar et.al [8] proposed method to examine the use of metadata using ontology for clinical research database. The ontology based Meta data management system allows a customized integration of heterogeneous clinical databases. Daniel.L.Rubin[4] developed a method using data from the visible human. The authors have demonstrated the usage of ontologies with medical images to support computer reasoning about injury based on images. Go-Ebi et.al [6] reviewed GO project. It provides ontologies to describe attributes of gene products. Peter.D.Karp [10] developed a functional ontology for Ecocyc database. Function based data queries have been demonstrated for how the ontology can be used. Stuart Blair et.al [15] reviewed the gene ontology to solve the computational problems of biology.

III. Ontology Model for Phylogeny Visualization

The framework of proposed work for visualizing phylogeny relationship using ontology is given in figure 1. The proposed framework consists of Phylogeny Creation, Ontology Design, and Visualization.

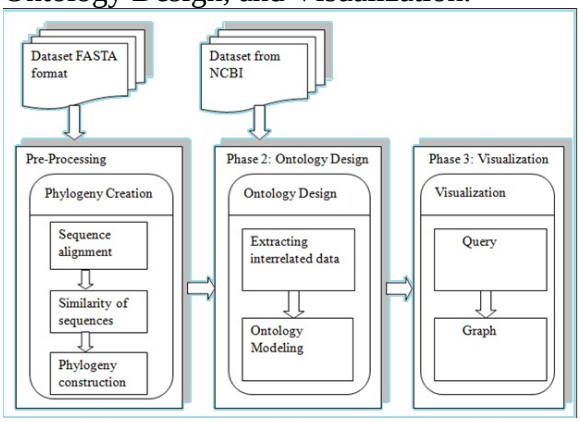

Fig 1: Framework for Ontology Model for Phylogeny Visualization

Dataset

The dataset used for the proposed work is downloaded from NCBI. FASTA Sequence of hominidae family is downloaded for phylogenetic tree construction. Hominidae is the Great apes and humans. The hominidae form a taxonomic family of primates including four

extant genera; 1) Chimpanzee (pan) 2) Gorilla (gorilla) 3) Humans (homo) and 4) Orangutans (pongo). The dataset contains FASTA sequence information of 12 organisms.

Fig 2: FASTA Dataset

Phylogenetic Creation:

The Phylogeny Creation is the first phase which consists of three steps Similarity of sequence, Sequence Alignment, and Phylogeny tree construction.

Similarities of sequences:

The degree of sequence identity between 2 nucleotide sequences. Align more than two sequences it may want to compare many sequence that trying to determine evolutionary relationship among many organisms.

Sequence Alignment:

Sequence alignment is the way of arranging the sequence of DNA, RNA, and protein to identify region of similarity that may be consequence of structural, functional, or evolutionary relationship between the sequences. Sequence alignments are used for non-biological sequences.

Pairwise Distance:

A variety of distance algorithms are available to calculate pairwise distance. Distance metrics compares two aligned sequences at a time, and builds a matrix of all possible sequence pairs. During each comparison, the number of changes(substitutions, insertion and deletion) are counted and presented as a proportion of the overall sequence length. These final estimates of the difference between all possible pairs of sequences are known as pairwise distance.

Multiple Sequence Alignment:

A Multiple Sequence Alignment (MSA**)** is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. The input set of query sequences are assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor. In the result of Multiple Sequence Alignment, Sequence can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins.

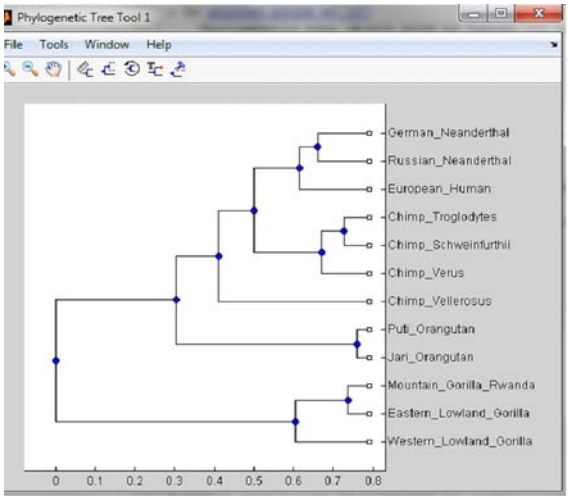

Fig 3: Snapshot of Phylogeny Creation

The Ontology Design is the second phase which consists of two processes Extracting interrelated data and Ontology Modeling.

Extracting interrelated data:

In this phase extracting interrelated data, the inter related details of organisms represented in phylogeny like gene information, Gene functional description in Gene Ontology are extracted from the Gene dataset. The snapshot of extracted and mapped details of organism is given in Fig 3.

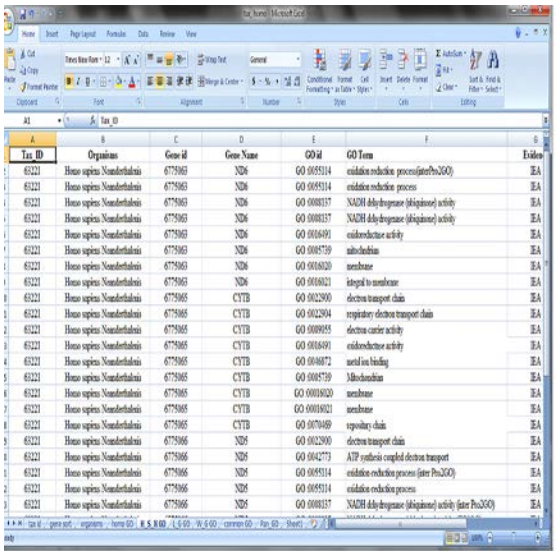

Fig 4: Extracting Interrelated Data

Ontology Modeling

The Ontology modeling is the part of the second phase. Ontology is designed by mapping the gene information for which a schema is designed for mapping gene information with ontology properties of Hominidae species.

Ontology Terminologies

Class:

- The thing is the default main class for protégé tool.
- Organism is the base class which represents the phylogeny hierarchy of organisms for hominidae family.
- Gene is a parent class which holds the gene details.
- Gene functionality class is used to define the functionality of Gene at three levels with sub classes Biological process, cellular component and Molecular Function.
- The class gene type is used to define the category which gene belongs to and has three gene types which forms the subclass of the defined class. The three subclasses are Protein coding, Pseudo, and Unknown.

Members:

The individuals are represented as member. Gene id provides the identification of the gene which is represented as member. Taxon id provides the identification of the organisms which is also represented as member. The individuals are mapped through the gene identifier to the corresponding gene class.

Object Properties:

The following object properties are defined for ontology constructed.

has_gene: The given object property is used to map the gene to the corresponding organisms.

has_go: The given object property is used to map the gene with corresponding GO identification number

has_evidence_as: The given property is used to map the gene to the corresponding genes.

has_genetype_as: The given property is used to map the gene to the corresponding go functionality.

has_organism: The given object property is used to map the organism to the corresponding genes.

belongs_to: The given object property is used to map the gene to the corresponding gene type.

Data Properties:

The following object properties are defined for ontology constructed.

has_gene_id: The given data property is used to assign the gene identification number to the corresponding genes of the organisms.

has_taxon_id: The given data property is used to assign the taxonomy identification number to the corresponding organisms.

Mapping gene information

Mapping of gene information is significant process in phylogenetic visualization ontology. The gene information is mapped to their respective gene id by using the ontology concepts. Organism is defined as the base class for the ontology constructed in the proposed work.

Visualization

The third phase is the Visualization which is used to extract the functionality and relationships among the genes. Visualization consists of two parts; Query and Graph Visualization. Query is used to retrieving the information from the ontology constructed for inferring phylogeny of organism with interrelated data. Graph is used to visualize the ontology.

DL Query:

The information for genes is retrieved using the DL Query tab in the tool. The DL Query tab provides a powerful and easy-to-use feature for searching a classified ontology; it is a standard plug-in for protégé 4. The query language supported by the plug-in is based on the Manchester OWL syntax, user friendly syntax. It is based on collecting all information about a particular class, property, or individual into a single constructed called a frame. The structure of the syntax for the DL Query is gene functionality.

Class name and Object property name value “”- The syntax used to query and retrieve information based on class and object property with value. e.g. ‘western_lowland_gorilla’and has_gene some MT-CYTB and belongs_to only Molecular_Function and has_evidence_as some IEA has_genetype_as some Protein_Coding.

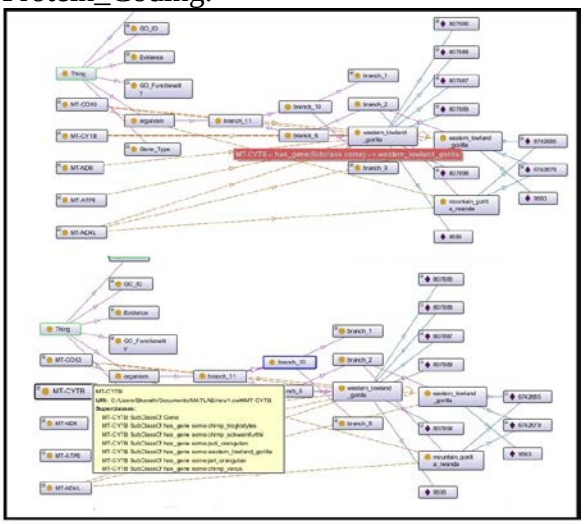

Fig 5: Sample Visualization of class_name and object_property

The information of gene is retrieved using DL Query. DL Query tab provides a powerful and easy to use for identifying the organisms.

Class Name	Object Property value	Query of DL
Organism	has_gene	Organism and has_gene some MT-COX3
Organism	has_go	Organism and has_go some GO:0005739
Organism	belongs_to	Organism and belongs_to only Biological_Process
Organism	has_genetype_as	Organism and has_genetype_as some Protein_Coding
Organism	has_evidence_as	Organism and has_evidence_as some IEA
ATP6	has_gene_id	ATP6 and has_gene_id value 6775074
MT-COX3	has_taxon_id	MT-COX3 and has_taxon_id value 9600
Organism	has_gene	Organism and has_gene some MT-CYTB and has_genetype_as some Protein_Coding
Organism	has_gene, belongs_to, has_genetype_as	german_neanderthaal and has_gene some ATP6 and belongs_to only Biological_Process and has_genetype_as some Protein_Coding
Organism	has_gene, belongs_to, has_evidence_as	western_lowland_gorilla and has_gene some MT-CYTB and belongs_to only Molecular_Function and has_evidence_as some IEA
Organism	has_gene, belongs_to, has_evidence_as ,has_genetype_as	chimp_troglodytes and has_gene some MT-ATP6 and belongs_to only Molecular_Function and has_evidence_as some IEA and has_genetype_as some Unknown
Organism	has_gene, has_gene_id, has_taxon_id	eastern_lowland_gorilla and has_gene some MT-ND4L and has_gene_id value 6742685 and has_taxon_id value 9593

Table 1: Snapshot of DL Query with various properties

International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-,December 2015
ISSN: 2395-3470
www.ijseas.com

In Fig 6, represents the DL Query visualization. In that Sample Query, organism and has gene some MT-COX3', this gene is present in two organisms like chimp_verus, eastern_lowland_gorilla, mountain_gorilla_rwanda, jari_orangutan and western_lowland_gorilla.

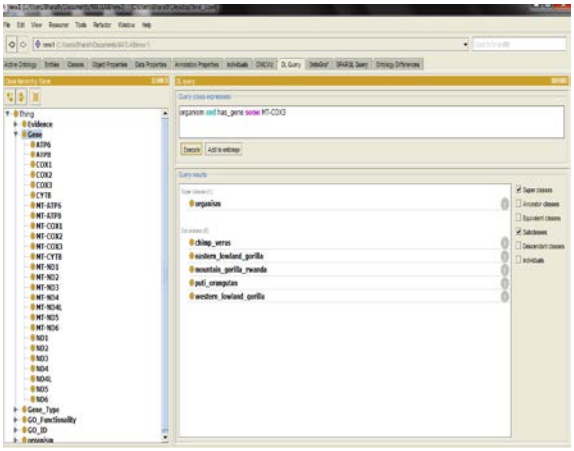

Fig 6: DL Query snap shot for Class name and Object Property.

IV. Result and Discussion

The result of the proposed work for extracting interrelated data and mapping gene information by using ontology concepts. The proposed approaches using the framework Ontology Model for Phylogeny Visualization (OMPV) is used to visualize the phylogenetic tree of organisms are discussed in detail in the following section.

DL Query:

The DL Query has the facility to retrieve gene information using class name, object properties, data properties from hominidae organisms' ontology. The Fig. 7 provides a snapshot of the information retrieved using the query with class name and object property. The Figure 8 provides a graphical visualization of the retrieving information for Class name and Object Property exists in Fig 7.

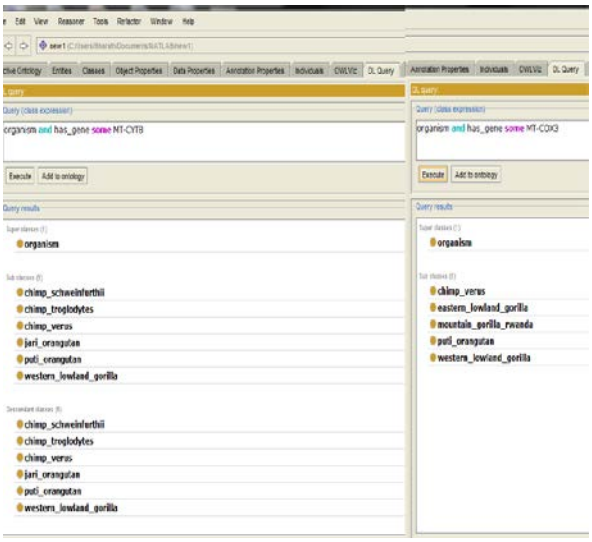

Fig 7: Retrieving information for Class name and Object Property.

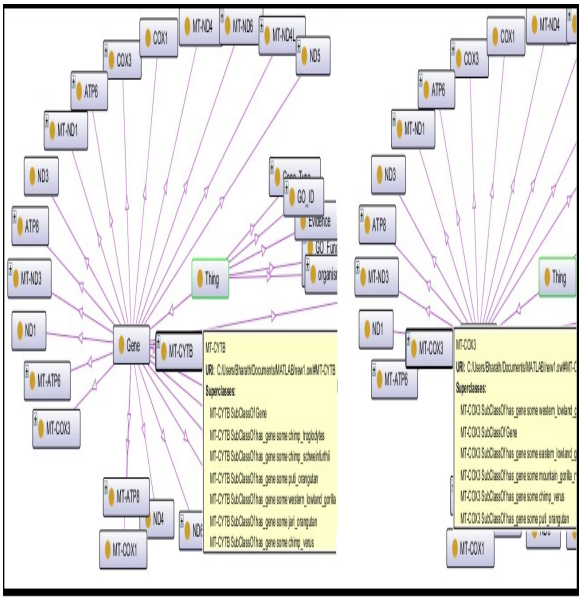

Fig 8: Graphical Visualization of retrieving information for Class name and Object Property.

The Fig.9 provides a snapshot of gene ATP6 with interrelated data of organisms like german_neanderthal, european_human, russian_neanderthal retrieved using the

International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-,December 2015
ISSN: 2395-3470
www.ijseas.com

query with different class name and object property. Eg. german_neanderthal and russian_neanderthal and european_human and has_go some GO:0004129 and has_gene some ATP6 and belongs_to only Molecular_Function.

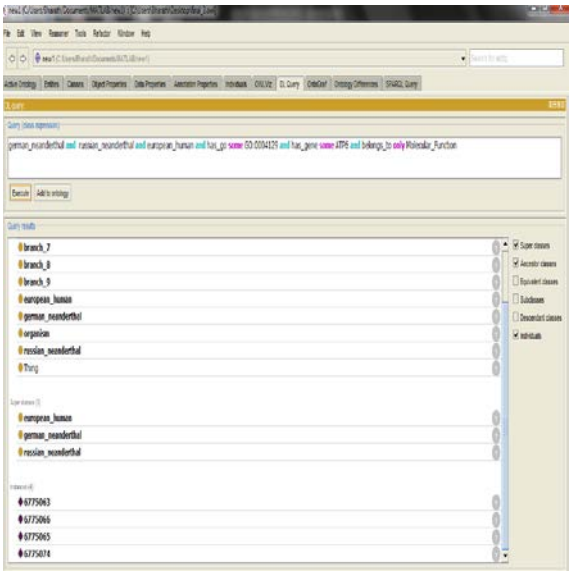

Fig 9: Snapshot of retrieving information of gene

The Fig. 8 provides a graphical visualization of the retrieving information for Class name and Object Property exists in Fig 7.

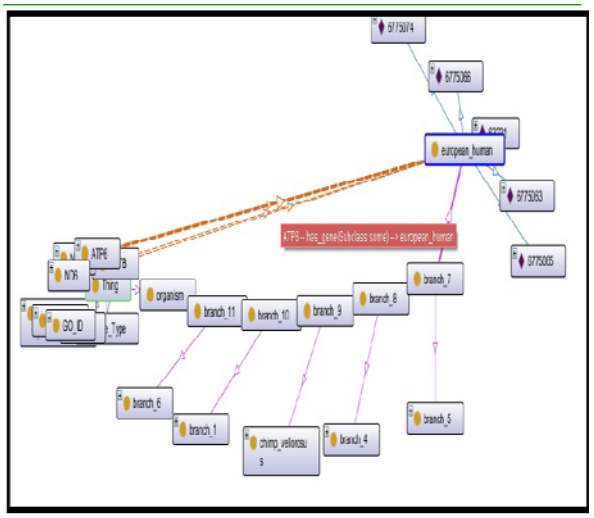

Fig 10: Graphical Visualization of retrieving information for Class name and Object Property.

Graph Visualization:

The gene information is also viewed using graph visualize by using ontgraf which is a plug-in in protégé tool. The Figure 8 provides a snapshot of the Graph Visualization of Ontology Model for Phylogentic Visualization.

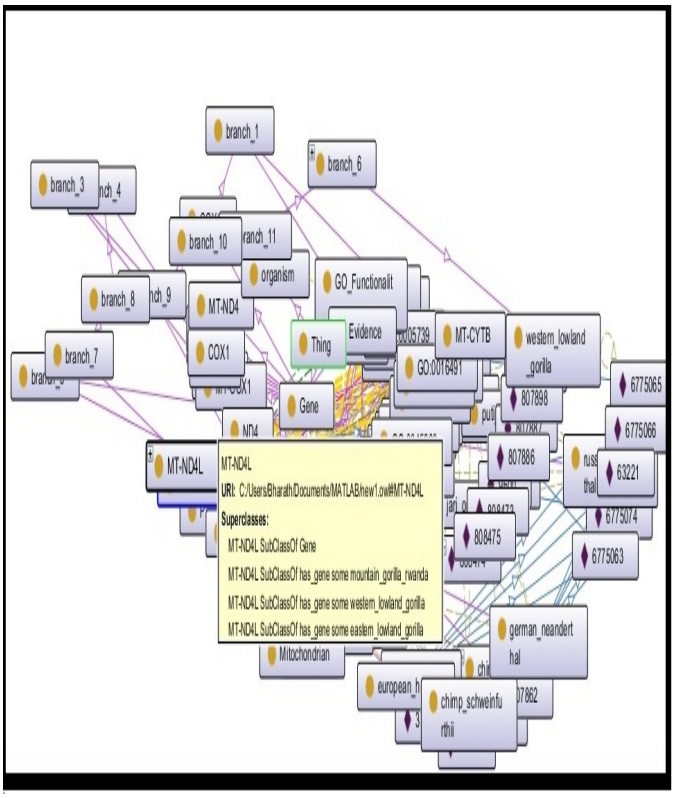

Fig 11: Graph Visualization of Ontology Model for Phylogentic Visualization

V. Conclusion

The Ontology model for Phylogeny Visualization is proposed. The three different phases of Ontology Model is:

Phylogeny Creation, Ontology Model, and Visualization. The result analysis of proposed ontology is compared with the clustered phylogeny of organism. The existing approach does not provide any interrelated data of the organisms. In the Ontology model for Phylogeny Visualization the phylogenetic tree is linked with interrelated data of organism like relative gene functionalities. The querying interface of ontology model helps to query the details of organisms with interrelated data using DL Query, which is not possible in traditional phylogenetic methods.

References

[1]. Arun K Pujari "Data Mining Techniques", University Press (India) Private Limited 2001. ISBN 978 81 7371 380 4.

[2]. B.Orgun, M.Dras, et.al. "Approaches for Semantic Interoperability between Domain Ontologies", Published in Proceedings of International Conference in Research and Practices in Information Technology (CRPIT) in IEEE explore, 2006, Vol. 72. Australian Ontology Workshop (AOW) Hobart Australia, 2006.

[3]. Dan He "Ontology –based Feature Weighting for Biomedical Literature. Department of Computer Science, University of Vermont, Burlington VT05405, USA.

[4]. Daniel L.Rubin, Oliver Damerson et.al "Using Ontologies linked with geometric models to reason about penetrating injuries", Published in Journal of Elseiver doi: 10 1016/, 2006.

[5]. Glenn Blanchette et.al "Inference of Phylogenetic tree:Hierarchical Clustering Versus Genetic Algorithm", Published in Journal of LNCS 7691, pp.300-312, 2012.

[6]. Go-Ebi, Embl-Ebi, et.al., "The Gene Ontology (GO) Database And Informatics Resource", Published in Journal of Nucleic Acid Research, 2003,Vol 32, pp 258-261, doi: 10.1093/nar/gkh036.

[7]. Iván Cantador, Martin Szomzor, et.al., "Enriching Ontological User Profiles With Tagging History For Multi-Domain Recommendations" Cited at http//www.pnas.org 2008.

[8]. Kaustubh Supekar and Yugyung Lee "Ontology based Metadata Management in Medical domains" Published in Journal of Research and Practice in Information Technology Vol. 35, 2, 2003.

[9]. Mike Uschold and Robert Jasper "A Framework for understanding and Classifying Ontology Applications" Published in Proceeding of the IJCAI, Workshop on Ontologies and Problem Solving Methods, 1999.

[10]. Peter.D.Karp "An Ontology For Biological Function Based On Molecular Interactions", Published in Journal of Bioinformatics Ontology, 2000, Vol 16, Issue 3, pp 269-285.

[11]. Purvesh Khatri and Sorin Dru ghici "Ontological Analysis of Gene Expression data: Current tools, limitations and open problems", 2005.

[12]. Robert Stevens et.al "Ontology based Knowledge Representation for Bioinformatics", Published in Journal of Briefings in Bioinformatics Vol.1, 4, pp 398-414, 2000.

[13]. Sahar Hassan, Franck Hetroy et.al "Ontology Guided Mesh Segmantation", Published in "Focus K3D Conference on Semantic 3D Media and Content",2010.

[14]. Steven Maere et.al (2005) "BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks" ,Published in Journal of System Biology, 2005, Vol 21, Issue 16,pp 3448-3449 doi:10.1093/bioinformatics/bt i551.

[15]. Stuart Blair et.al "A review of the Gene Ontology: past developments, present roles, and future possibilities" 2010.

[16]. Yi Ping Phoebe Chen(Ed) "Bioinformatics Technologies", Published in Springer International Edition, 2005.