American Association for Science and Technology
AASCIT Communications
Volume 1, Issue 3
October 20, 2014 online

AASCIT American Association for Science and Technology

Semantic Integration of Biological Entities in Phylogeny Visualization: Ontology Approach

Velumani Bhuvaneswari, Manohar Sharmila

Keywords: Phylogeny Visualization, Ontology, DL Query

Semantic integration of information has become necessary to fetch knowledge related in various domains. Bio-informatics is a domain where integration of information of biological entities and annotations from various knowledge bases has become necessary. 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. Current there exists various approaches for phylogeny visualizations which provides the relationships in forms of hierarchy. Understanding the relationship among organism represented using phylogeny tree needs the user to search for other related data of the organisms. The proposed approach overcomes the drawback of the existing phylogeny visualization approaches by integrating the data of organisms using an Ontology approach. The ontology is implemented using Protégé tool provides visualization of phylogeny relationships with integrated data and found to be efficient.

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]. The need for semantic integration of data has become important currently due to huge volume of datasets generated related to biological entities. Ontological approaches are widely used in Semantic Web for integration of data. The ontology modeling is found to be well suited for application domains for integrating highly semantic data. Various ontology integration models are proposed by various authors for bioinformatics [2][11][4][6].

Phylogenetic tree construction is an core area in bioinformatics which involves study of evolutionary relationships among 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 existing phylogeny visualization approaches do not integrate biological entities of organisms visualized in phylogenetic trees. A similar work is carried out for integrating gene information with other biological entities [17][18]. The work presented overcomes the limitation of the existing phylogeny visualization approaches by integrating the relevant data of biological entities like protein, gene, GO annotation, database references in phylogeny visualization using ontology. The organization of the Paper is as follows: the section II reviews various methods available from the literature related to Phylogenetic methods and visualization tools. Section III explores the proposed ontology framework for Phylogeny Visualization. The IV section discusses the experimental results of the work with snapshots followed by conclusion in Section V.

Review of Literature

The related reviews of various ontology approaches are discussed in the below section. 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 et.al [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 et.al. [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. Bhuvaneswari et.al [17] [18][19] has proposed ontology models for integrating gene information with other biological entities.

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.

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 3 phases: Phylogeny Creation, Ontology Design, and Visualization. The phase 1 the phylogeny tree is constructed using sequence information in FASTA format. In the phase 2 of the framework the ontology is constructed by extracting relevant data of biological entities. The phase 3 the phylogeny tree is visualized using ontology constructed.

The dataset used for the proposed work is downloaded from NCBI for hominidae family. 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.

Phylogeney Creation:

The Phylogeny Creation is the first phase which consists of three steps: Identifying Similarity of sequence, Sequence Alignment, and Phylogeny tree construction. The degree of sequence identity between 2 nucleotide sequences is used for identifying the similarity of sequences.. Sequence alignment is classified in to two categories as pair wise sequence alignment and Multiple Sequence alignment (MSA). MSA technique is used to infer the similarity among group of sequences. The similarity matrix is used for aligning sequence and the phylogenetic tree is constructed as given in Figure 2.

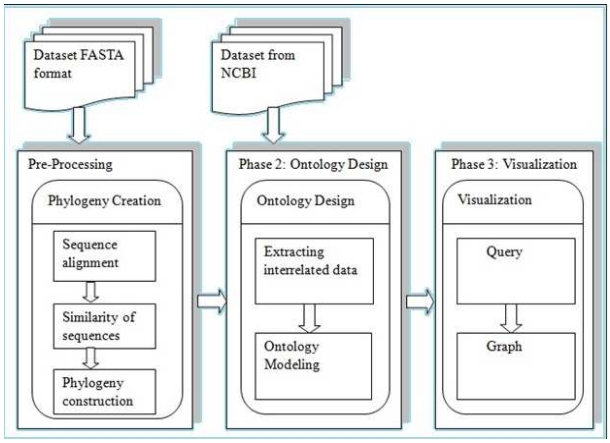

Figure 1. *Framework for Ontology Model for Phylogeny Visualization*

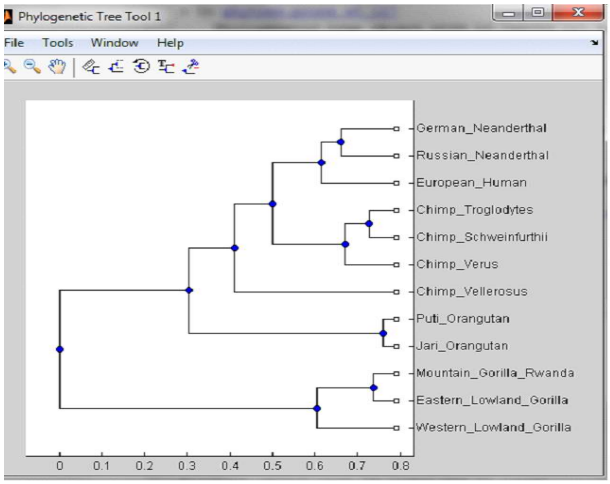

Figure 2. *Snapshot of Phylogeny Creation*

Ontology Model

The Ontology Design is the second phase which consists of two processes Extracting interrelated data and Ontology Modeling. The phylogeny information is given as input for extracting related data for ontology construction.

The inter related details for organisms represented in phylogeny like gene information, Gene functional description in Gene Ontology, Database references are extracted from the Gene dataset downloaded from NCBI.

The schema model for the proposed ontology representation with ontology terminologies is presented in the below sections.

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 object properties are defined for mapping entities with other related entities is given in Table 1.

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 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. Class name and Object property name value “”- The syntax used to query and retrieve information based on class and object property is given in Table 1.

Table 1. *Snapshot of DL Query with various properties*

Class Name	Object Property vale	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_genetye_as	Organism and has_genetype_as some Protein_Coding
Organism	has_evidence_as	Organism and has_evidence_as ome 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_neandertheal 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

Result and Discussion

This section presents with discussion based on the experimental analysis. The proposed approach the framework Ontology Model for Phylogeny Visualization (OMPV) is used to visualize the phylogenetic tree of organisms with interrelated data. The various snapshot of experimental results with visualization approach is presented in the below sections.

The Figure 3 provides a snapshot of the information retrieved using the query with class name and object property using DL Query. The Figure 4 provides a graphical visualization of the retrieving information for Class name and Object Property using Ontology Visualization engine of Protégé tool.

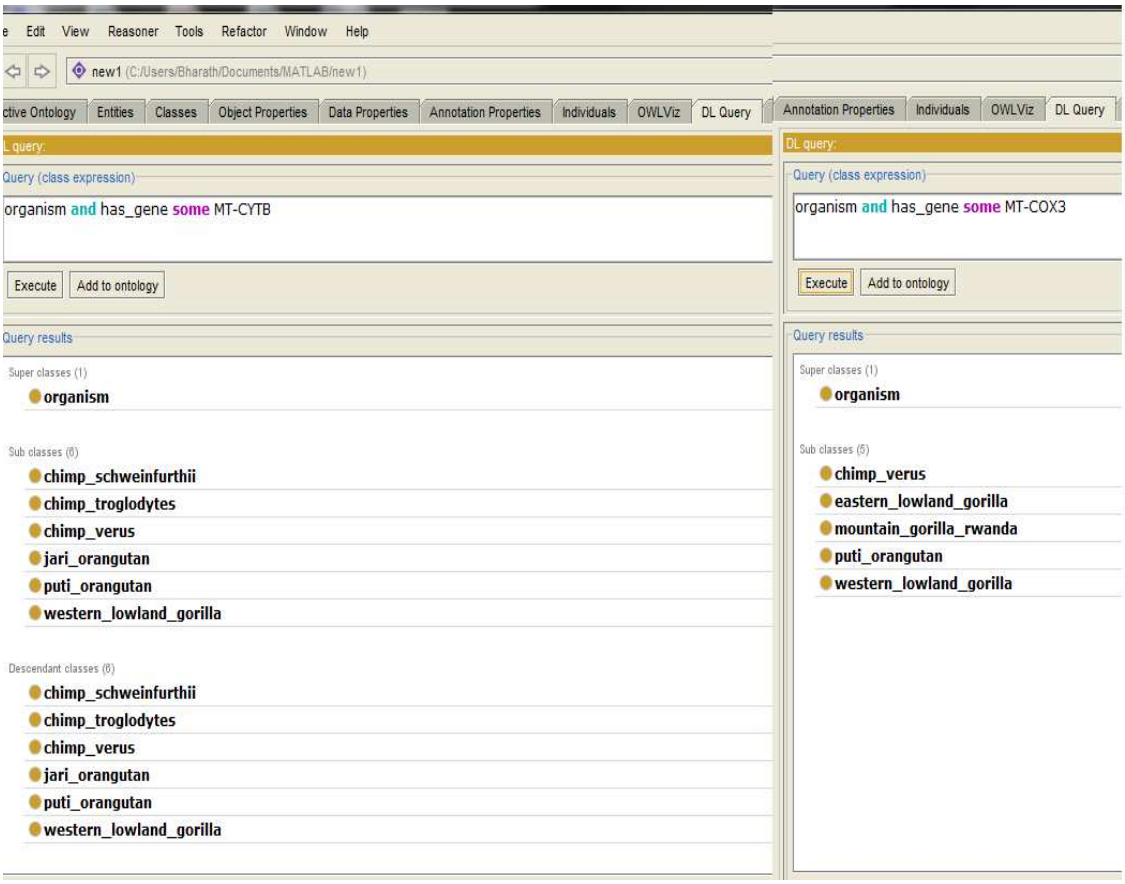

Figure 3. *Retrieving information for Class name and Object Property.*

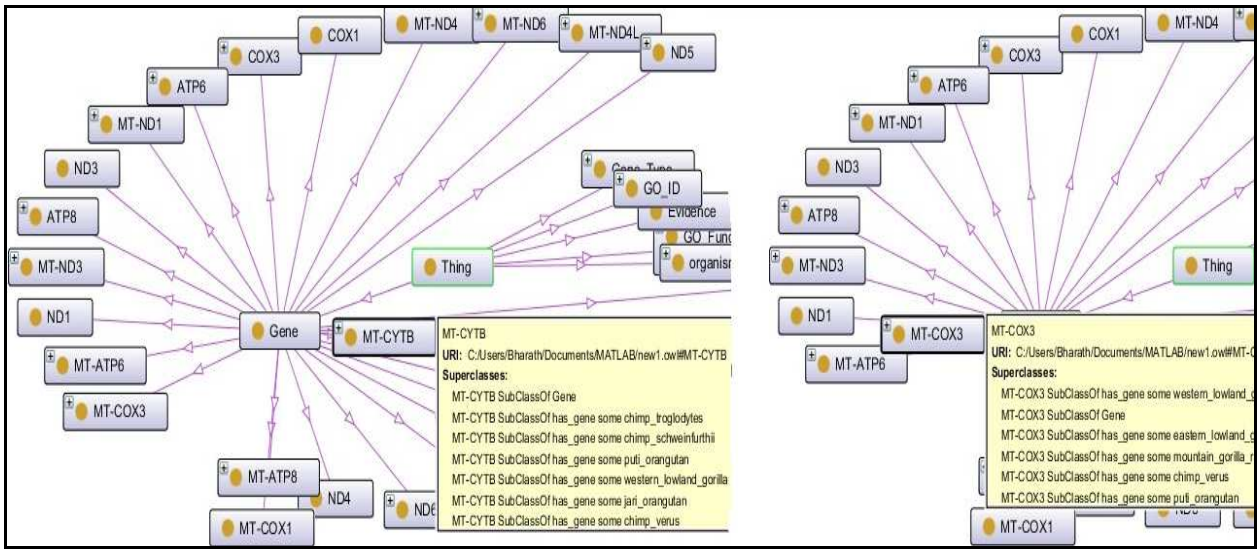

Figure 4. *Graphical Visualization of retrieving information for Class name and Object Property.*

Graph Visualization

The gene information is also viewed using graph visualize by using ontgraf which is a plug-in in protégé tool. The Figure 5 provides a snapshot of the Graph Visualization of Ontology Model for Phylogentic Visualization.

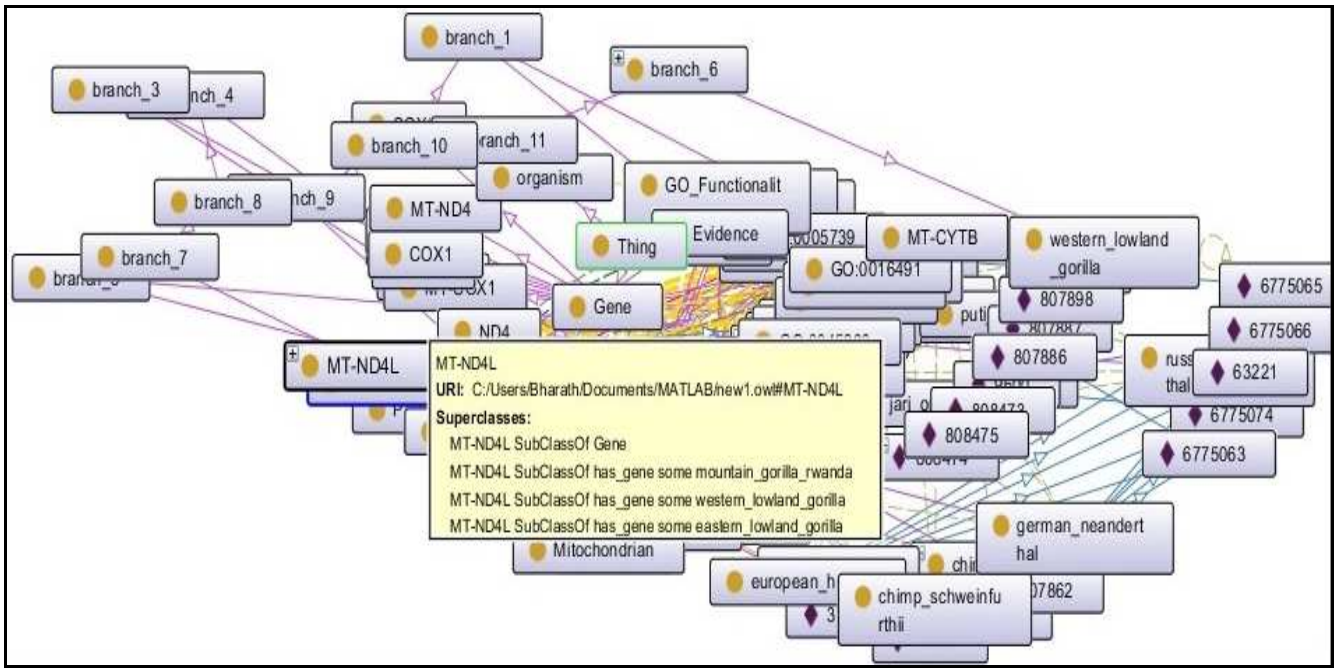

Figure 5. *Graph Visualization of Ontology Model for Phylogentic Visualization*

The visualization approaches is available in form query and graph representations. The query interface provides mechanism to query related data dynamically represented in ontology modeling. The formation of query is also so simple which makes combination of class and other related object and data properties. The graph visualization can also be viewed through drag of ontology classes dynamically. Hence the proposed approach of integrating related biological data in phylogeny visualization is found to be efficient compared to other existing approaches.

On implementation the limitation of the work wer found to be integration of data generated in phase 1 and phase 2 with respect to programming languages. In this work we have considered only the core biological entities like Gene, Gene functionality and its databases for integration. The other classifications like disease gene, protein structures, motifs which are not considered can be upgrade as future work.

Conclusion

This paper an Ontology model for Phylogeny Visualization is proposed using a framework. The framework consists of three phases which includes Phylogeny Creation, Ontology Model, and Visualization. Phylogeny is constructed in phase 1 based on multiple sequence alignment. The second phase the relevant data is extracted and semantic mapping is carried out and represented in ontological terminologies. The third phase of the framework the ontology model designed is explored with various visualization approaches using query and graph. The proposed approach has overcome the existing limitations and found to be an efficient approach for integration related data into phylogeny for analyzing evolutionary relationships among organisms.

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.■

V. Bhuvaneswari
Department of Computer Applications, Bharathiar University, Coimbatore, Tamilnadu, India
bhuvanes_v@yahoo.com

M. Sharmila
Department of Computer Applications, Bharathiar University, Coimbatore, Tamilnadu, India
mcasharmi8@gmail.com

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: 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/bti551.

[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.

[17] B.L.Shivakumar, V.Bhuvaneswari,"Ontology Design for Gene Integration and Feature Representation", Published in IOSR Journal of Engineering, Vol. 2, Issue 5, pp: 11, May. 2012.

[18] V Bhuvaneswari , R Priyadarshini , "An Ontology based framework to integrate gene information for homosapiens taxon" Publised in Lingaya's journal of professional Studies(LJPS), Volume-4(2), Jan-June issue ,2011.

[19] V.Bhuvaneswari, R.Priyadarshini, "An Ontology Based Approach for Classifying Gene Information and Comparing with Traditional Methods", Published in Proceedings of International Conference Emerging Trends in Computing, Coimbatore , ICETC, March 2011.