International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Genomic and Proteomic Semantic Annotations Integrating Cross Ontology

Karthik K, S.Rajaprakash, S. Muthuselvan, Fayas alam, Cris Mathew

Abstract: ***Considering the intricate biological phenomena that demand solving the difficult queries on biomedical- based on biomolecular content in sequences, these are sent via various proteomic and genomic semantic annotations that are distributed in many heterogeneous format. With those knowledge and dispersion various biologic scientist's skill of enquiring various problems and solving them continuously which becomes tedious work for them. To put an end to this problem I developed a software based architecture which creates and maintain a GPKB -Genomic and Proteomic Knowledge Base (GPKB), this combines various important gene diseases and its relevant information.***

The main problem of such discrete information. The answer to this problem is simple, since the software uses as modular, flexible and a big multilevel schemed data which is based on socializing the combined features of data and its abstraction. It also sets up a trial method for deleting all the combined data that have its structure, data and its numbers. Such methods will also provide consistency, quality and tracking methodologies for all combined data.

Keywords—Ontology, GPKB.

I. INTRODUCTION

Data and information specifically, Increasingly large amounts of valuable, but heterogeneous and sparse, biomolecular are characterizing life sciences . In particular, controlled terms that describe the biomolecular entity features or functions and semantic controlled annotations of biomolecular entities, i.e. the associations between biomolecular entities, are of great value; they support scientists with several terminologies and ontologies describing structural, functional and phenotypic biological features of such entities (e.g. their sequence polymorphisms, expression in different tissues, or involvement in biological processes, biochemical pathways and genetic disorders). These semantic annotations which can be used to formulate and validate biological hypotheses and possibly discover new biomedical knowledge can effectively support the interpretation of genomics and proteomics test results and the extraction of biomolecular information. A comprehensive approach to such data integration, querying and analysis by answering related complex biomedical questions can help understanding complex biological processes and their pathological alterations. Yet, the scattering of genomic and proteomic annotation data in many complementary but also overlapping sources is an important and not yet completely solved challenge. Specifically, the facets of a very hard data integration problem are data source heterogeneity in data representation and format, their fast evolution in number, data content and structure, the high variety of available data types, and also the great amount of data produced over time.

I developed a software architecture to create and maintain an updated and publicly available integrative data warehouse of genomic and proteomic semantic annotations, taking advantage of our previous experience with the GFINDer system. For integrated data management, It adopts a modular and multilevel global schema that we propose. This data schema supports integration of data sources, possibly overlapping, which are fast evolving in data content, structure and number, and assures provenance tracking of all the integrated data.

The outline of this paper is as follows. Section 2 discusses the related work in data integration, focusing on the biomedical domain. Section 3 describes our integrated data schema. Section 4 illustrates the developed software architecture for data integration, which ensures consistency, quality and provenance tracking of all the integrated data, eases their updating and extension and perform semantic closure of the integrated ontology hierarchical relationships. Section 5 presents the Genomic and Proteomic Knowledge Base (GPKB), which benefits from our integrated data schema and software architecture and provides Web interfaces to easily compose queries, although complex, on the integrated semantic data. Section 6 illustrates a relevant example of GPKB use for discovering common biological aspects in apparently unrelated genetic disorders. Section 7 discusses significant aspects of our work and concludes.

II.RELATED WORKS

In Many methods and schemes have been planned to mixed information from many varied information resources. The final tactic is identified to have upkeep above, mutually keeping the mixed information till date with the unique designated bases and in increasing the storeroom with extra information and kinds of information from latest bases [19]. [11]

Manuscript published on November 30, 2019.
* Correspondence Author
Karthik K* , Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India. karthik@avit.ac.in
S.Rajaprakash, Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India.
R.jaichandran S , Dept. of Computer Science and Technology, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India. csmuthuselvan@gmail.com
Fayas alam Final year CSE, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India
Cris Mathew Final year CSE, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India

L. Elazary and L. Itti, "Interesting objects are visually salient," J. Vis., vol. 8, no. 3, pp. 1–15, 2008.Still, the information warehouse method is greater in enhancing applications that require off ‐ line information processing in order to professionally organize the combined information and widely use them for facts finding, which is my box. [1]
T. Etzold, A. Ulyanov, and P. Argos, "SRS: Information Re ‐ trieval System for molecular biology data banks", Methods En ‐ zymol., vol. 266, pp. 114 ‐ 128, 1996. Also, it lets completely checking information class and uniformity, within a sole or diagonally many information resources, in order to combine and use only superior excellence constant information. [24] M. Masseroli, O. Galati, and F. Pinciroli, "GFINDer: Genetic disease and phenotype location statistical analysis and mining of dynamically annotated gene lists", Nucleic Acids Res., vol.33, pp. W717-W723, 2005.It also simply let reunion non-mixed information, e.g. from separate information bases with diverse informing periods, [10] C. Privitera and L. Stark, "Algorithms for defining visual regionsof-interest :comparison with eye fixations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 9, pp. 970–982, Sep. 2000.
by winning benefit of accessible ancient growth information during storeroom building. For all these details, I accepted the information store housing method.
To come out of these problems,E. Cadag, B. Louie, P.J. Myler, and P. Tarczy-Hornoch, "Biome-diator data integration and inference for functional annotation of anonymous sequences", Pac. Symp. Biocomput., pp. 343-354, 2007. I established and accepted a linked and many level feature based combined information plan, which is defined in part 3. [3] R. Stevens, P. Baker, S. Bechhofer, G. Ng, A. Jacoby, N.W. Pa ‐ ton, C.A. Goble, and A. Brass, "TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources", Bioinformat ‐ ics, vol. 16, pp. 184 ‐ 185, 2000. It not only comforts information store housing updates and extensions, but also confirms origin chasing of all the combined information
Freshly, a proportion of importance has been located on the usage of related information for organic information [23]; [7] A. Freier, R. Hofestädt, M. Lange, U. Scholz, and A. Stephanik, "BioDataServer: a SQL ‐ based service for the online integration of life science data", In Silico Biol., vol. 2, no. 2, pp. 37 ‐ 57, 2002.Still, linked information provides only dual networks between couple of source items and enquiring them yet remains problematic due to the absence of consistency in the picture of connected information sets [24].[21] W. Sujansky, "Heterogeneous database integration in biomedi-cine", J. Biomed. Inform., vol. 34, 4, pp. 285-298, 2001.
Also, instinctive borders for enquiring organic connected information and by means of mined outcomes are yet very partial. [10] C. Privitera and L. Stark, "Algorithms for defining visualregionsof-interest :comparison with eye fixations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 9, pp. 970–982, Sep. 2000.
Equally, attentive on some designated bases, my method mixes and facilitates information items mined from many bases, with general larger information class and continuous enquiring and outcome usage support.

III. INTEGRATED DATA SCHEME

In this Section, I exemplify and consider the universal data schema that defines the combined various, diverse, controlled annotation data, i.e. data regarding unlike character or topics represent through various controlled vocabularies and ontologies, as well as their relations

3.1 Characteristics Module

This combined data schema is constitutes of numerous interlinked modules; every module signifies a single characteristic, whose data are provided by one or more of the combined data sources, and contains derivation information for every single characteristic instance entry (Figure 1). As I aim at annotation data controlled biomedical-molecular, a characteristic can be a bio-molecular entity (i.e. DNA sequence, gene, transcript, protein) or a biomedical characteristic (e.g. pathway, genetic disorder, etc.), a characteristic instance can be, for example, protein, a specific gene pathway or genetic disorder, and a characteristic entry is a particular representation of a characteristic instance (e.g. the data of a specific gene in a particular data source such as the Entrez Gene database). Each characteristic entry is recognized by the importance of its Source ID and Source name features (since each characteristic instance can have numerous IDs from different sources), and contains the Reference attribute, representing the basis that provides the data, i.e. their provenance (which can be unlike from their ID source); for instance, Table 1 shows two Gene characteristic entries, from different sources, that signify the Gene characteristic instance of the human BRCA1 gene Each characteristic module of the common schema can also include History and/or corresponding data. The earlier ones represent outdated discontinued source IDs and, if they have been circulated, the current ID that restores each of them. The later ones illustrate similarity of different characteristic entries (from the similar or different sources) through their ID pairings, which connects the different characteristic entries that the IDs identify (e.g. gene characteristic entries identified by different IDs from distinct sources, but representing the similar gene; see Table 1 for an example). These similarity data can be imported from one or more sources, or placed in by specialist curators, or automatically inferred by a computational process (e.g. based on Processing of textual descriptions of Natural Language existing within some data attributes, or based on semantic analysis of ontological data). Both history and corresponding data are paramount to merge unsynchronized data and recognize multiple characteristic instance entries, from single / multiple data sources, as representing the identical characteristic instance (e.g. data regarding different gene IDs which actually represent the similar single gene; for instance, in the case of history data, this can happen when the gene ID modifies and the gene data are supplied by different sources, among the combined ones, with different revising time, in some of the gene ID that has not yet been updated/revised). Thus, history and similar data directly enables the precise mapping and combination of different characteristic data. 3.2 Characteristic Module Associations Data characteristic modules are pair wise related (through associating annotation data); and these connotations are categorized in an import and an aggregation tier. In the later one, the co notate data, that are included in the import tier as pairs of characteristic entry IDs, are repeatedly translated into pairs of exceptional OIDs and are matched to the characteristic entry OIDs of the two associated characteristics. By taking benefit of existing ID history and similarity data, this translation also allows reconciling discontinued IDs to their current ones, and recognizing as such different IDs that represent the similar characteristic instance.nIn the end, a third, higher and more common combined tier (not shown in Figure 1 and Figure 2)

completes the integrated data schema by signifying all the exceptional characteristic instances, or theories, (e.g. all distinct genes, proteins, pathways, genetic disorders, etc.) and these associations illustrated by the integrated data, despite of the source(s) that supply(s) them (e.g. all the integrated different genes and their annotated characteristics, despite of the multiple IDs of each of them and their providing source).

IV. SOFTWARE ARCHITECTURE FOR SEMANTIC INFORMATION INTEGERATION

Profiting from idea, modularity and configurability of our mixed information scheme, in Java programming language I created a generalized and parametric software architecture, which helps the modified mechanical formation of an information warehouse accepting our information schema, and keeps apprising the information warehouse and spreading it with new information bases relaxed. My approach for the combined dispersed multi-source varied information is separated in two major ladders, performed according to the defined configuration meta information:

1. Importing information from their varied bases in the source-import tier of our combined information scheme
2. Mixing them in the illustration-combination and idea-combination tiers of the information schema.

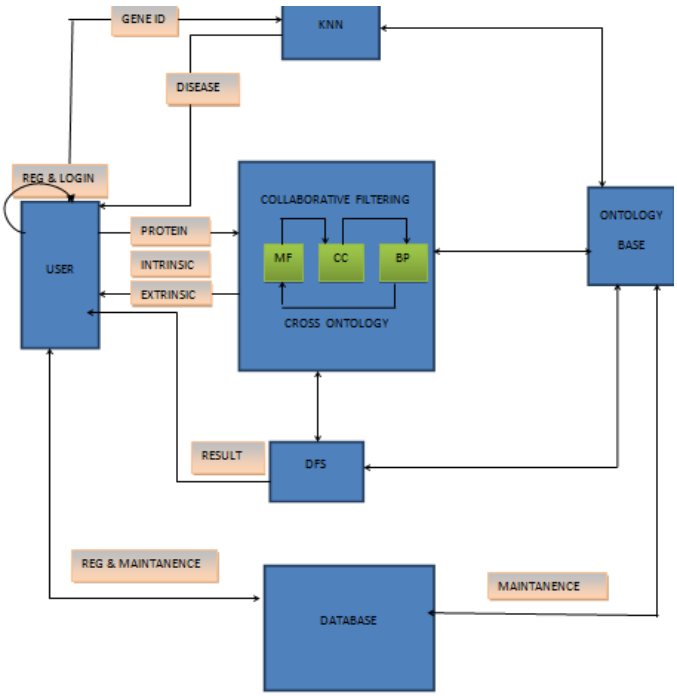

Fig 1: Architecture Diagram

4.1 INFORMATION IMMORT:

The information import process is directed by an import manager that instantiates, configures and executes an importer for each considered information source. Each source specific importer organizes a set of loaders (a loader for each information file, group of homogeneous information files, or information access API provided by the source) and a set of parsers (a parser for each information format). Each parser excerpts the information from its related input file(s) or API(s) and generates information tokens usable by a loader. Each loader is answerable for connecting a semantic sense to the tokens shaped by the related parser and interleaving them into the warehouse.

The importing outline allocates to each imported "information record" an OID, which is exclusive across the information warehouse. It is used as the main identification of the information entries, subsequently there is no assurance that the IDs given by many sources do not struggle with one another. In order to confirm perfection of imported information, a set of systematic terminologies has been defined to check and find IDs [22]. They are used by the ID matcher, an extra module of my construction that acts as a moderator between the loaders and the information warehouse. The key role of this mediator is to verify ID syntactic correct-ness and classify the semantic type of each ID, in order to inject the right information in the appropriate information warehouse tables. During this process, each inserted tuple is also improved with derivation information to track its source. Right ID identification is supreme since information from many sources are then connected together cheers to association information provided by the combined sources as couple of IDs in different information sources.

4.2 DATA INTEGERATION

The data combination or integration step includes two automated tasks: combination and aggregation. In the first task, the data imported from the former import data step and the data from other sources, are collected & regulated into one illustration in the instance-aggregation tier of our schema of global data. In the second assignment, data is prearranged into revealing groups in the concept-combination tier of the schema of the combined data.

Then, connotations within pairs of article entries are formed by performing OID rendition of the imported connotation data articulated through the article entry IDs. By this, relationship data are joined together with the associated feature records. The featured records or entries that aren't introduced in the data warehouses, getting referred by the associated data will depend on the data sources that are imported and their common organization. In such cases, missing combined feature records are blended and highlighted as incidental through fusion from association data. When an omitted entry has an outdated ID and its most recent ID can be gained through unfolded chronological ID data, the connotation is relocated to the most recent ID and marked as incidental through chronological data. This connotation rendition procedure conserves, post combination, all the connotations communicated by the imported connotation data from different sources. Thus, it permits consequently using such connotations for biomedical information detection.

Whilst in the final combination phase, with the help of a "resemblance scrutiny", it is verified if the particular article entry from diverse sources signifies the same feature theory (e.g. various gene records recognized by various IDs of the same gene from divergent sources). In such case, they are associated with an original distinct theory OID (e.g. a gene theory OID). Additionally, new records can be incidental from the associated data [30]. The incidental characteristic in the combined tables is used to keep track of the corollary method active, if any, to develop a record

V PROTEOMIC AND GENOMIC INFORMATION BASE

To demonstrate the relevance and usefulness of the designated broad software architecture and the combined data schema, which was executed in a PostgreSQL RDBMS, we used them to form, sustain reorganized and increasingly encompass a multi-organism combined Proteomic and Genomic and Data Warehouse (GPDW). It constitutes a high worth and constant combination of various biomolecular interaction and semantic annotation data labelling several biomedical-molecular structures of many biomolecular records, specifically proteins and genes. Such data are brought in from various scattered data sources, cautiously picked up for their transformed significance, which include UniProt, IntAct, Entrez Gene, GOA, GO, KEGG, Expasy Enzyme, Reactome, B ioCyc, and OMIM. Initially, the GPDW comprises in excess of 2.36 billion data tuples; which aggregate to a total of about 737 GB of disk space in the multi-level data schema (together with their indexing). They also contain about 17,535,404 genes of 14,995 various organisms, 19,544,576 proteins of 23,368 species and a total of 16,772,399 gene annotations and 30,440,619 protein annotations articulated though 10 biomedical well-ordered ontologies or terminologies.

The concluding ones include 41,829 Gene Ontology terms and their 42,057,775 semantic annotations, 359,511 biochemical pathways from, KEGG (469), Reactome (37,210) or BioCyc (321,832) and 1,733,389 pathway annotations, as well as 7,936 human genetic disorders from OMIM and their 12,473 gene annotations, together with 34,177 phenotypes (signs and symptoms). We dug up phenotypes from the OMIM clinical synopsis semi-structured descriptions as illustrated in [5]; initially, to the best of our information, they are not counted in organized, easily questionable form in any additional integrative database which is publicly available. Additionally, the GPDW incorporates 626,516 appreciated interaction data from IntAct between various biomolecular records, together with 609,864 protein interactions. In addition, the GPDW contains 3,616,108 semantic annotations of 988,899 genes which were identified by transitive correlation from the integrated annotations of the proteins that these genes encode; the semantic handling that we used is thoroughly showed and deliberated in [30].

Query outcomes are presented in a tabular view, whose columns can be freely composed by the user for greatest and informal exploration of the outcomes (Figure 6). Furthermore, the user can choose all or a subset of the query outcome (and of their characteristics) and increase the initial completed query in order to improve or supplement them, basis the liquid query principle [31]; this augments very valuable explorative quests of the abundant and diverse data combined in the GPDW and eases biomedical information mining, specifically when the search query full description and data filtering values are not known in the beginning but can be found by observing limited query results. For example, a GPKB user can initiate query for the genes identified to be involved in a given pathway (e.g. Apoptosis, or RNA transport). Then, the user can refine the initial query outcomes by opting for only few of the established genes (e.g. the ones the user is more concerned in). This semantic development permits recovering also all such indirect, less precise ontological annotations of a biomolecular record in the GPDW, e.g. of a gene, easing the appreciative of all the biomolecular entity features and the use also of such annotations, e.g., for gene enhancement analysis [6], [32] or semantic annotation prediction

VI EXAMPLE USE CASES

In order to display embryonic and congruity of the GPKB, in this Segment we epitomize and examine a complete illustration of its purpose in order to acknowledge indicative biological inquests about multiple anarchy and disclose common biological conditions in ostensibly unrelated genetic anarchies. Multiple disease components can be argued by molecular pathways, involving subgroups of genes and their synergy, where phenotypic intricacy results from interactions amidst genomic irregularity in various disposition. In precise, polygenic attributes are due to the collective activity of multiple genes, which consecutively can be pleiotropic genes, i.e. intricate in more pathological phenotypes. Thus, the multiple pertinence among gene variants and diseases can be well inferred by indicating the genes and metabolic pathways involved in pathological process. extensive search of integrated biomedical information and principles from complex sources can assist formulating which genes and pathways (processes among genes) may be the seminal candidates of a disease. Literally, few genes may be pivots associating distinct complicated disease modules (e.g. different cancer types) and so they may have a vital part in the disease progression (e.g. in carcinogenesis). This advent can also be the basis for few new analysis, or partial description, of extensive phenomena.

Analogous discoveries might support exposing natural microscopic sources of the diseases or disorders, against their phenotypic instability. for instance, the relevant autologous recombination and cellular divisional pathways precisely point out that probable disease causing changes take place during cell division of the pathogen stripe, where meiosis is at the base of volatility and primogeniture through genetic recombination. moreover, using the GPKB we have analysed that differences of one of the genes endowed intricate in both breast and prostate cancers, the cadherin 1, E-cadherin (epithelial) (CDH1), type 1, human gene, are recognised to be convoluted also in other cancer forms, i.e. Gastric cancer, Ovarian cancer, Colorectal cancer and Endometrial cancer . this proves that this gene might play a pivotal aspect in standardizing various biological mechanisms, and so divergent are cancer disease modules.

Test case no	Description	Pre-conditions	Pass/Fail	Expected results
GO_001	Validate user Registration	New user only allowed	Pass	Registered Successfully
GO_002	Validate user login	Registered user only allowed	Fail	Login successfully
GO_002a	Validate user login	Registered user only allowed	Pass	Login successfully
GO _003	User get input from pharmacist or doctor or scientist	Registered user only allowed	Pass	Get input successfully
GO_004	Show all details for that id	Registered user only allowed	Pass	View gene disease successfully
GO_005	Get protein value for that id	User file are allowed	Pass	Get value successfully
GO_006	Using protein value to identify intrinsic or extrinsic	Valid user only allowed	Pass	Get function of Input ID successfully
GO_007	View gene disease by user	Registered user only allowed	Pass	View disease successfully
GO_008	View gene symptoms by user	Valid user are allowed	Pass	View symptoms successfully
GO_009	View curing possibilities by user	Registered user only allowed	Pass	View cure successfully
GO_010	Generate tree structure	Registered user only allowed	Pass	Tree generate successfully

Fig 2: Test Case

VII DISCUSSION AND CONCLUSION

Semantic and bimolecular data are huge in number and quality in addition to being partially connected and dispersed, because of the relevant progresses in biotechnology and system. Understanding complex biological phenomena, normal or pathological, and ultimately of enhancing diagnosis, prognosis and treatment involve the usage of Integration and mining of these distributed and evolving data and information have the high potential of discovering hidden biomedical knowledge; but such integration poses huge challenges. Maintaining, updating and extending an integration of many evolving and heterogeneous data sources and defining them have attained a generalized and novel way because our work has tackled them in an effective way.

Multiple ontologies and originally available separately in many different sources pave a way for our proposed data schema and architecture enable us to easily create, keep updated and progressively extend the GPKB which is a publicly available collection of numerous, semantic bimolecular annotation data expressed. Running on big data our system supports comprehensive semantic queries with adequate performance and by recording data provenance and modeling associations among the integrated data, GPKB supports comprehensive reliable data analyses and mining. This helps answering complex multi-topic biomedical questions, which cannot be equally managed by other available systems such as Bio- Mart. As our example use case shown, the GPKB is mostly useful to unveil hidden bimolecular and biomedical associations. The developed GPKB Web interface well enables users to easily compose queries, although complex, on multiple features/topics and to extract valuable biological insights, including hidden associations, which may answer significant biomedical questions.

Opened on the web about 30 months ago, relevance of the GPKB is also proved by the about 47,000 accesses received by more than 1,280 visitors and frameworks for detection and prediction of semantic bimolecular annotations and the Bio-SeCo system in support of distributed bio=data explorative search to complex biomedical questions are some of the usage of GPKB being made available on web in addition to a few other projects.

Accessing GPDW via web services generally used as a collection of web services for a programmatic access, that are being developed for the purpose of keeping it open to the public. Making the service interface open for public availability will enable the access to the GPDW data by other computational systems, providing inclusion in scientific workflows and new foreseen applications, e.g. in drug repurposing; we are carrying on a project on this topic in collaboration with the US National Library of Medicine.

REFERENCES

1. **T. Etzold, A. Ulyanov, and P. Argos, "SRS: Information Re-trieval** System for molecular biology data banks", Methods En-zymol., vol. 266, pp. 114-128, 1996.
2. T.A. Tatusova, I. Karsch-Mizrachi, and J.A. Ostell, "Complete genomes in WWW Entrez: data representation and analysis", Bioinformatics, vol. 15, pp. 536-543, 1999.
3. R. Stevens, P. Baker, S. Bechhofer, G. Ng, A. Jacoby, N.W. Pa-ton, C.A. Goble, and A. Brass, "TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources", Bioinformat-ics, vol. 16, pp. 184-185, 2000.
4. Z. Ben Miled, N. Li, G.M. Kellett, B. Sipes, and O. Bukhres, "Complex life science multidatabase queries", Proc. IEEE, vol. 90, pp. 1754-1763, 2002.
5. S.B. Davidson, C. Overton, V. Tanen, and L. Wong, "BioKleisli: a digital library for biomedical researchers", Int. J. Digit. Libr. , vol. 1, pp. 36-53, 1997.
6. D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, A. and Kasprzyk, "BioMart - Biological queries made easy", BMC Genomics, vol. 10, 22, pp- 1-12, 2009.
7. A. Freier, R. Hofestädt, M. Lange, U. Scholz, and A. Stephanik, "BioDataServer: a SQL-based service for the online integration of life science data", In Silico Biol., vol. 2, no. 2, pp. 37-57, 2002.
8. E. Cadag, B. Louie, P.J. Myler, and P. Tarczy-Hornoch, "Biome-diator data integration and inference for functional annotation of anonymous sequences", Pac. Symp. Biocomput., pp. 343-354, 2007.
9. K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, et al., ″The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud″, Nucleic Acids Res., vol. 41, no. Web Server issue, pp. W557-W561, 2013. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
10. C. Privitera and L. Stark, "Algorithms for defining visual regionsof-interest :comparison with eye fixations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 9, pp. 970–982, Sep. 2000.
11. L. Elazary and L. Itti, "Interesting objects are visually salient," J. Vis., vol. 8, no. 3, pp. 1–15, 2008.
12. S. Hwang and K. Grauman, "Learning the relative importance of objects from tagged images for retrieval and cross-modal search," Int. J. Cmput.Vis., vol. 100, pp. 134–153, 2011.
13. M. Spain and P. Perona, "Measuring and predicting object importance," Int. J. Comput. Vis., vol. 91, no. 1, pp. 59–76, 2011.
14. Displacement prediction model of landslide based on multi-gene genetic programming Jiejie Chen; Zhigang Zeng; Ping Jiang 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)
15. Identification of gene network motifs for cancer disease diagnosis Rohit Gupta; S. M. Fayaz; Sanjay Singh 2016 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics

16. An effective approach to identify gene-gene interactions for complex quantitative traits using generalized fuzzy accuracy Xiangdong Zhou; Keith C. C. Chan 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)
17. Gene selection using interaction information for microarray-based cancer classification Songyot Nakariyakul 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)
18. Simultaneous execution method of gene clustering and network inference Shuhei Kimura; Masato Tokuhisa; Mariko Okada- atakeyama 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)
19. A Bayesian approach to construct Context-Specific Gene Ontology: Application to protein function prediction Hasna Njah; Salma Jamoussi; Walid Mahdi; Mohamed Elati 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)
20. M.Y. Galperin, D.J. Rigden, and X.M. Fernández-Suárez. "The 2015 Nucleic Acids Research Database Issue and Molecular Bi-ology Database Collection", Nucleic Acids Res., vol. 43, D1, pp.D1-D5, 2015.
21. W. Sujansky, "Heterogeneous database integration in biomedi-cine", J. Biomed. Inform., vol. 34, 4, pp. 285-298, 2001.
22. B. Louie, P. Mork, F. Martin-Sanchez, A. Halevy, and P. Tarczy-Hornoch, "Data integration and genomic medicine", J. Biomed. Inform., vol. 40, 1. pp. 5-16, 2007.
23. C. Goble, and R. Stevens, "State of the nation in data integration for bioinformatics", J. Biomed. Inform., vol. 41, pp. 687-693,2008.
24. M. Masseroli, O. Galati, and F. Pinciroli, "GFINDer: Genetic disease and phenotype location statistical analysis and mining of dynamically annotated gene lists", Nucleic Acids Res., vol.33, pp. W717-W723, 2005.
25. M. Masseroli, "Management and analysis of genomic functional and phenotypic controlled annotations to support biomedical investigation and practice", IEEE Trans. Inf.echnol. Biomed., vol. 11, 4, pp. 376-385, 2007.

AUTHORS PROFILE

Mr. K.Karthik ME (Ph.D) currently working as Assistant professor Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India published more than 7 national and international journal and conference and organizing committee for 4 international conference,2 national conference and 15 years of teaching experience with 4 years of research experience. He is a member in following professional societies: CSI and ISTE.

Dr.S.Rajaprakash M.sc, M.Phil M.E Ph.D. currently working as Associate professor of CSE in Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Missions Research Foundation (Deemed to be University), Tamil Nadu, India. He has 18 years of experience in academics, research, and development activities. Published 17 research papers in referred Journals and Conferences. His area of Interest Artificial Intelligence, Computational Intelligence, Discrete Mathematics and Automata theory. .Received grants from Tamil Nadu State Council for Science and Technology .He has peer Reviewed Manuscripts in reputed international Journals and Conferences. He is a member in following professional societies: CSI and ISTE and Ramanujam Mathematical Society.

Mr. S. Muthuselvan, M.E., (Ph.D) currently is working as Assistant Professor Gr. II, Aarupadai Veedu Institute of Technology an ambit institution of Vinayaka Mission's Research Foundation (Deemed to be University), Tamil Nadu, India. Published more than 17 national and international journal and organizing committee for four international conference, two national conference and Five years of industry experience, 11 years of teaching experience with 6 years of research experience. He has peer Reviewed Manuscripts in reputed international Journals and Conferences. He is a member in following professional societies: CSI and MISTE. Area of the interests is DBMS, Data Mining and Data Analytics.

Fayas alam Final year CSE, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India

Cris Mathew Final year CSE, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation Chennai, India