# Ontology Design for Gene Integration and Feature Representation-OGFR

## Dr.B.L.Shivakumar[1], V.Bhuvaneswarir[2]

[1]*(Professor and Head, Department of Computer Applications, SNR Sons College, Coimbatore,Tamilnadu )*
[2]*(Assistant Professor, Department of Computer Applications, Bharathiar University, Coimbatore,Tamilnadu )*

**ABSTRACT:**
Recent few decades many ontology have been developed for various domains which mostly includes educational domain, agricultural domain and biomedical domain. Particularly in the biomedical domain, large set of ontologies were developed to integrate variety of genomic information. Gene related functional information is stored in various databases after the sequencing of human genome in 2003. Searching of the information for any purpose it becomes essential to understand the schema of database to fetch information, which has becomes the time consuming and complex task. .  This paper provides a unified framework that integrate diverse source of genomic data from various repositories such as ontologies and data bases using ontology based approach. The proposed ontology design coverst all possible relations for genomic data such as gene names and its products, gene types, gene identifier, mapping indexes in various repositories for genes, etc. The proposed ontology design is presented in terms of ontological concepts and the ontology design is evaluated in terms of syntax, structure and semantics using various evaluation metrics. The experimental result of our ontology shows that the proposed ontology is designed with well defined classes and conceptual relationships.

**Index Terms: Concepts, Ontology design, Tools, Structural, Semantic similarity.**

## 1. Introduction

Ontologies provide a common representation of terms and relations of a particular domain. They are used as the backbone of Knowledge representation related to a domains for integrating various sources of data or systems.  In recent years Ontologies are becoming more popular in the field of bioinformatics. The development of molecular biology has brought out the explosions of large data after the sequencing of human genome in the year 2003.  Protein and gene information with its various functions generated out of the sequencing are represented in genomic databases by various organizations. The searching and acquiring information from the various data sources becomes a challenging task. The unification of the information   is addressed through ontology representation. The GO Ontology is the famous ontology in bioinformatics, providing various information and functionalities of genes.

Ontologies have been extensively used in data integration systems because they provide an explicit and machine-understandable conceptualization of a domain. A common ideal for ontology is that it should be re-usable [9]. This ambition distinguishes ontology from a database schema, it is intended to satisfy only one application, but ontology could be re-used in many applications even though both are conceptualizations. The paper mainly focuses in integration of Genomic features for gene information spread across various relational databases. The ontology OGFR is created and designed using protégé tool .

The distinct feature of the OGFR ontology design is that it makes Concept map to represent domain information. Concepts map provides an easy visualization approach, which helps to understand the ontology by the users and domain experts. Evaluating ontology framework design requires the supervision of domain experts. The ontology designed has to be evaluated against the quality, cohesion, domain covering and its richness. To validate the resulting ontology, this paper uses the ontology evaluation methodology and the results are obtained in evaluating the corpus generated from databases. The OGFR provides an environment for accessing gene information directly instead of searching through various databases.

The paper is organized as follows:  Section 2 briefly presents the study related to domain Ontologies in biomedical domain and also the study related to evaluation of the ontology and the tools used for creating ontology. Section 3explains the framework proposed by us for integrating gene information using ontology based approach OGFR by exploring the design , creation and mapping the information from datasets. Section 4 explains the evaluation methodology used for the proposed ontology. Section 5 discusses the results of the ontology evaluations and section 6 provides the application of the OGFR ontology for bioinformatics researchers and final section draws the conclusion of the paper.

## 2. Related Study

This section presents related work on ontology design in various domains and biomedical domains. The section describes the work related for integrating data using ontology. The literature also discusses the methods used for evaluating and validating the ontology.

### 2.1 Domain Ontology Creation

In the paper the authors has reviewed  the ontologies Gene Ontology (GO), Mammalian Phenotype (MP) Ontology, and Adult Mouse Anatomical (MA) Dictionary developed at MGI (Mouse Genome Informatics laboratory). The ontologies helps to integrate the genomic and functional information of mouse lying at various disparate sources.

They have also discussed the common vocabularies, tools, browsers used to support ontology driven querying [4].

In this paper the author Yu-Ting Huanga et.al has proposed a frame work for extracting knowledge from database like pubmed using ontology inference and semantic processing techniques. Ontology based knowledge extraction is proposed to infer information after processing using NLP techniques. The authors have also compared their results with KEGG pathway database for Aptopsis domain [23].

The paper the authors Auxilio Medina and Alberto have presented techniques for constructing, implementing and maintaining Ontologies. They have constructed the ontology semi automatically based on hierarchal clustering of documents. They have provided the methods for constructing and maintaining the ontology. The authors have constructed the ontology using XML language and RDF format from the clustered results and called as records. They have also proposed an idea for maintaining the ontology created by choosing accurate concept map for clusters by continuous revision and rewriting [16].

Anne Macula the authors has constructed and designed ontology for representing amphibian information using protégé tool [7] The ontology include the physical and self connected objects for all amphibians. They have identified 212 semantic concepts and 58 relations to design the ontology. The author Robert Stens has proposed methodology for designing bio ontologies and compared various existing bio Ontologies [18]. The authors have defined a concept for designing ontology related to neuroscience. The ontology design provides the facility to query the required results based on the ontology concepts [20]. In this paper authors has designed ontology for analyzing and integrating protein transport interactions and gene expression for specific cell type, they have also classified the documents based on the ontology constructed [13]. The author Amandeep S. Sidhu have reviewed the traditional data integration methods like keyword based search , integration based on sequence identity , structural similarity and functional identification for Protein. The author has also discussed the complexities that exists in the traditional approach and proposed integration of proteins based on ontology concepts [3]. The Author Barry Smith has discussed the structure of GO ontology model , relationship representation and they also pointed the drawbacks and limitation of the ontology . The authors have also proposed ideas to improve the existing ontology [8]. The authors has reviewed the usage of ontology in Biomedicine and its current trends based on the existing Bio-Ontologies[17].

## 2.2 Validation methods for ontology

The authors have proposed method for integration and discussed the role of Ontologies in semantic Integration for knowledge representation and compared it with existing ontology [5]. The Authors has described similar metrics for comparing two ontological concepts and compared it with other existing approaches. The paper the authors have proposed a Method for measuring semantic similarity

concepts in the same ontology [21]. Measuring of the semantic similarity of concepts based on the factors, Link type, Node depth, Local network Density, Strength of an edge link, Concept attribute, Granularity degree of clusters in Ontology.

The author's have presented a methodology for constructing ontology from text documents. The ontology is created for an educational system, they also have evaluated the ontology created in terms of structure, syntax and semantics and compared their approach with other existing ontology [1][2]. The author has proposed a task based ontology model for integrating information from various domains of AI and software engineering. The author have represented the tasks for process management [19].

The authors have presented a semantic similarity measure based on ontology structure using multiple Ontologies for biomedical concepts. They have proposed three measures: cross path length between concepts, commonality between concepts and local granularity of ontology clusters. They have also validated and evaluated the approach with other existing methods and human expertise for MESH terms and clinical terms [11]. The authors have created semi automated ontology for breeding environments a collaborative network. The first phase of the work the authors have created a common vocabulary for representing common concepts as ontology for collaborative networks using Protégé. The second phase the authors have created a semiautomatic ontology for extracting network specific knowledge related to competencies. They have applied the idea for real time problems for mechanical engineering companies [12].

The authors have developed a framework using ontology for extracting knowledge from biomedical documents for understanding the molecular functions in biological pathways. The ontology is also evaluated using KEGG databases [23].

## 3. Ontology Based Gene Integration

Integration of information from various heterogeneous data sources is becoming very important in recent years, to avoid the problem in searching data. Ontology based approaches are used in data integration as it provides a standard representation of domain in common platform. Many biomedical ontology exist and the uniqueness of each ontology depends on the purpose in which it is designed.

Knowledge Extraction: Ontology can represent knowledge in two forms; declarative and procedural. Declarative knowledge is used for representing domain Ontologies. Difficulties in representing gene information as concepts and other functional attributes are : The gene information of all organisms are available as separate data bases with different information and relations. Recovering all information related to a particular gene is difficult as domain knowledge is necessary to identify the attribute to relate the information. Difficulties exist in identifying the attributes from for constructing ontology. Various biomedical Ontologies are developed for representing gene information, the famous GO ontology is used as a model in framing the concepts.
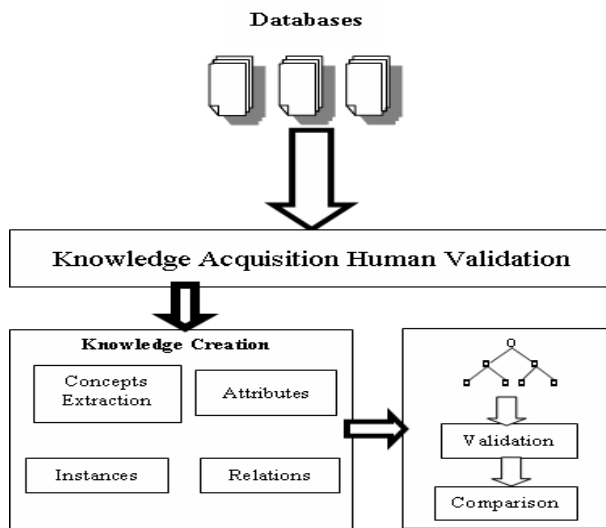
**Figure 1. Gene Integration using Ontology- Framework**

**T**he domain ontology creation of the proposed work consists of various steps as shown in Fig 1.

- The first step in ontology creation is to understand the schema definitions of the various datasets for integrating gene information.
- The main task in information acquisition is to extract the attributes, identify and mine the domain terms for concepts mapping to represent the ontology hierarchy.
- The second step is to convert the extracted terms for ontology representation by detecting classes, attributes, instances, and relations and save them in OWL/XML format using Protege.
- The ontology is validated by human experts to verify the representation of the ontology created. The ontology is validated by using the evaluation metric as proposed in the paper . Finally the ontology is compared with the existing ontology in the domain which is discussed in Section 4.

### 3.1 Information Acquisition Process

Gene is the basic functional unit of any biological living process. The information of genes are maintained in different databases with attributes like type of gene coding , the chromosome, map location, name, identifier of other database references. GO Ontology classifies the gene based on three functional processes namely biological, cellular and molecular. The gene mapped with GO ontology classification is also represented in attributes by assigning GO numbers. The challenging task for retrieving the information from the databases depends on the knowledge of schema information to understand the relations within various datasets. The proposed ontology provides a unique gene information retrieval system through a single ontology.

**Table 1 Gene Dataset**

| Datasets | Fields | Description |
|---|---|---|
| Homo_sapiens.gene_info | Tax_id, Gene_id, Symbol, Locustag, Synonyms, DBxrefs, Chromosome, Maplocation, Description, Genetype, Symbol_nomen_authority, Fullname_nomen_authority, Nomen_status, Modified date. | It defines the core details of genes based on the Tax_id, Gene_id of the Homo sapiens information |
| Gene2ensembal | Tax_id, GeneID, Ensembl_gene_identifier, RNA_nucleotide_accession version, Ensembl_rna_identifier, protein_accession version, Ensembl_protien_identifier. | It defines the ensembal information's of genes based on the Tax_id, Gene_id of the Homo Sapiens information. |
| Gene_group | Tax_id, GeneID, Relationship, Other_tax_id, Other_GeneID. | It defines the relationship among the genes based on Tax_id, Gene_id of the Homo sapiens information. |

The gene information dataset is downloaded from National Center for Biotechnology information Entrez(NCBI Entrez) website and used for the knowledge acquisition process. The ontology created integrates gene information for human taxon from six datasets like Gene-Info, Gene2Ensemble, and Gene-group, Genedetails . The gene information (gene_info) dataset contains genes details like Tax_id, GeneID, Synonyms, Chromosomes, Maplocation, Fullname nomen authority etc. The database includes a variety of genomic and biological information. The Gene2ensembl file contains Ensemble annotation to the respective Gene Identifier and the dataset Gene_group contains the information related to a particular gene with other Gene identifiers.The Gene dataset is shown in Table 1.

### 3.2 Term Extraction

Term extraction process is the discovery of terms for representing concepts in ontology. The prime attributes of the dataset which are unique are mined after a detailed study and selected for representing the core concepts. The dataset of gene information contains 50 attributes. Based on the attribute domain values the other terms are extracted for relating them with concepts through properties.

The term geneid is extracted and the corresponding gene names are mined and represented as concepts. The genes are classified as Protein coding genes, pseudo coding genes and unknown. The concepts gene is related with other concept gene type through properties. The functional process of genes is identified through GO values. The ontology constructed contains the GO identifiers as concepts, which is related to Gene. The other terms like Gene fullname, alternate name, database references, chromosome map location, number, taxon are mapped to the core domain concepts.

The section explains the integration of gene details using Protégé tool. The details of the genes information maintained in databases are converted into ontological terminologies like classes, properties, members and annotations for creating ontology using protégé tool. In the ontology creation process the following entities and attributes are modeled as given in Table 2 using ontology terminologies.

**Table 2: OGFR Concepts Representation for integrating Gene information**

| Class | Description | Dataset attributes |
|---|---|---|
| Gene | Gene class that contains all the details related to particular gene. | Gene Symbol |
| Gene type | Class which has gene function coding for protein, pseudo or unknown. | Type of gene |
| GO | The GO identifiers used for mapping to gene functions at Molecular, Biological, Cellular | GO identifier |
| Go Type | The class three structures molecular, biological and cellular to map with Go Identifiers | Functional Level |
| **Object Properties** | | |
| has_synonym | Official and unofficial symbols for the gene given as members are mapped to gene_id to gene identifier through this property. | Synonym |
| has_taxon | Gene symbols are mapped through this property, which can be used to link the ontology for other organisms using taxon id. | Taxon id |
| has_relationship | The property maps all the details of gene maintained in other databases. | All attributes of gene_function dataset. |
| Is_curatedfrom | This property has the reference of gene symbols from curated go references. | GO-Ref:0000002 is curated from go reference. |
| belongs to | This property is used to map Go identifiers at Molecular, Biological are Cellular level . | |
| Has_GO_Ref | This property is used to map gene symbols refereed from GO identifiers. | Descriptions from Gene Dataset |
| has_gene | The property is used to relate Gene with GO. | |
| has gene-type | This object property is used to relate gene to gene function . | |
| has_go | This property is used to map all GO identifiers related to gene at different functional levels.. | |
| **Data Properties** | | |
| is in_chromosome | The given data property is used to assign the chromosome number, which ranges from 1-22, defined in gene type class. | Chromosome |
| is in_maplocation | The given data property is used to assign the map location in chromosomes where gene exist and to map with corresponding Gene symbol. | Chromosome maplocation |
| is_curatedfrom | This property is used to map genes referred in in pubmed literature. | |
| has_dbxrefs | The given data property is used to map the other database references of genes HGNC,MIM,ENSG and HPRD identifiers with their unique database identifier to map with Gene symbol. | Database references from Gene_Ensemble dataset. |
| has_ensembleprotein | This database gene property is used to refer gene present in ensemble for RNA, nucleotide and protein. | |
| gene_description | This property is used to provide the description of gene symbol represented in class. | Descriptions from Gene Dataset |
| has_fullname | This property contains original description of genes represented as classes. | |
| has_alternatename | This database property contains alternate descriptive names for genes mapped to gene class. | |
| has_gene-id | This property is used to map the genes with unique gene identifier. | |
| **Members(Individuals)** | All alternate names, gene id are defined as individuals to map using object or data property. | Gene symbol, Alternate name, Full name from Gene dataset. |
| **Annotation** | Description of all classes and genes given as annotations. | |
| **Reasoner** | FACT++,HERMIT used to classify ontology for querying. | |
| **DL Query** | DLQuery is used for querying the ontology. | |
| **Visualization** | The ontology is visualized in various view using ontgraph and ontvisualizer. | . |

**3.3 Knowledge Creation Process**

In knowledge creation process the domain Ontologies are mapped to concepts extracted and the Ontologies terminologies like classes, properties, attributes and instances are determined.

**3.3.1 Classes**

The ontology OGFR constructed the domain terms are considered as primitive classes and some terms as defined classes. The domain terms Gene names, Gene types, Go identifiers are defined as primitive classes. The gene symbols equivalents are defined as equivalent classes as identified from the database. The associations between classes are represented through object properties. The gene

symbol is associated with gene coding types through object properties to form associations. Currently the ontology consists of classes belonging to human taxonomy. The same genes present in other organisms may be linked through the has_taxon property , where taxon is the unique identifier used for classifying different organisms.

### 3.3.2 Relationships
The relationships between pair of classes are related through object and data property assertions as shown below. The object property is defined to relate between a pair of classes. The gene codes for a particular function are mapped via GO numbers. GO numbers are unique identifiers classified based on the functionality of gene which participates at three levels namely Biological, Cellular and Molecular. The object property is defined for the gene classes to map with corresponding Go functionality where gene will be the subclass for the corresponding functionality.

### 3.3.3 Subclasses
The classes which share similar characteristics between concepts are assigned as subclasses. In the ontology all the genes codes for a particular functionality at biological, cellular and molecular level. The gene type class classifies

### 3.4.1 Ontology Evaluation
The domain ontology created is validated and evaluated. This section explores the evaluation techniques of our ontology created to check the performance based on various measures. The ontology designed is validated to check the correctness and characteristics. There exists various ontology evaluating and validating methods which is an ongoing research. The ontology is evaluated on structure, syntactic and its semantic representation. The structural evaluation checks the ontology in terms of concepts representation. The ontology is structurally validated using various metrics. The syntactic evaluation checks the ontology for consistency.

### Syntactic Evaluation
The syntax of the OWL Ontologies is evaluated to verify the consistency of the ontology using Reasoner. The unified ontology frame work is found to be consistent after the evaluation using Reasoner. The inconsistent concept is verified by human domain experts.

### Structural Evaluation
The structural evaluation is done by using the metrics defined by [1]. The ontology is evaluated structurally based on four metrics Class Match Measure, the Density Measure, Betweennes measure and Semantic Similarity Measure. A total score is then computed from all these measures. This score can be used to rank the ontology with respect to the given search terms. The values of all metrics and final scores are set between 0 and 1.

genes coding for proteins, pseudo genes and unknown functionality. The class gene type is defined as primitive class and all the other genes are mapped as subclass to gene type classification. There exists more than one gene symbol name for each gene called as synonyms which does not include the type of function it codes for, which are defined as instances for the gene class. The attributes are mapped via data property and object property. The data property are designed to relate the other characteristics of gene functional property like chromosome no, map location, database references.

### 3.4 OGFR Ontology Concepts
The database attributes referred from various dataset for integrating gene information are represented using ontology concepts. The modeling of gene information is done using Protégé tools. The attributes are identified after consultation with domain experts for modeling it using ontology concepts. The various concepts identified are represented as classes. The relationships between classes are mapped through various object properties and data properties. The consistency of the ontology is evaluated using Reasoner. The ontology can also be queried and visualized using DLquery and Ontgraph tools in Protege.

### Class Match Measure –CMM
The CMM evaluates the ontology for the given key terms. The terms given are searched in ontology to determine the terms as exact ontological classes or labels in the ontology as partial match. The weights are assigned to evaluate the metrics based on threshold. A common weight of 0.25 is used to calculate the score for all the three corpus. The threshold is varied based on the domain expertise. To evaluate the class match measure both exact and partial matches are considered. The exact matches the value increases than partial matches.

### Density Measure - DES
The Density Measure expresses the degree of richness of the attributes of a given concept. Density measure includes subclasses, inner-attributes, siblings and the relationship maintained with other classes. The metric are evaluated for varying thresholds. The DEM measure increases when number of concepts is increased.

### Betweennes Measure - BEM
The betweennes measure calculates the betweennes of the terns in the generated Ontologies. It measures the class centrality in the Ontologies. A value with high betweennes measure shows the centrality of the class. The betweennes measure in the proposed work we have used as the count of the class it is related with other concepts instead of using count of shortest path between other concepts. In ontology created all the concepts have only path related with other concepts, since each concept is unique to itself.

### Overall Scores:
Finally based on these four metrics, an overall score is computed. The weights assigned to each of the metrics can

be either the same or different. The overall score is explained in the comparative analysis.
Let M ={M1,M2,M3;M4 } represent class measure , density , betweennes measure and semantic similarity measures respectively. wi be a weight factor, and O be the set of ontologies to rank. The score is computed as follows [2]

$$Score(o \in O) = \sum_{i=1}^{4} w_i \frac{M[i]}{\max_{1 \leq j \leq |O|} M[j]}.$$

**Semantic Similarity Measure – SSM**
The last measure computes the proximity of the classes that terms match in the ontology. Alain and Brewster stated that the terms are used as representatives of the domain concepts, the ontology should link them through relationships using object and data properties. The SSM is based on the count of links that maps the concepts.

**Semantic Evaluation**
The semantic evaluation in ontology is domain dependent validated by experts. The ontology can be compared with existing same ontology concepts to verify the domain ontology created. We in this work have evaluated and validated the approach by the methodology proposed in [1][2]. The same metrics validation method is chosen to evaluate the ontology created, we have done a minor modification in the two metrics used by them to suit the gene integration for which the ontology is created.

## 4. Experimental Results
The dataset used for gene integration is downloaded from NCBI for human taxon. The downloaded dataset is represented with various ontological concepts to form the ontology hierarchy. To validate the ontology the domain terms represented as classes are used as the main key terms and relations that exist between other terms are also considered. The following keyterms as shown in Table is used to form the corpus and validate the ontology. The various terms given in table are represented as concepts in ontology. Most of the terms exist as classes with detailed description and linked with other concepts through relations.

**Table 4 Domain Terms**

| Terms |
| --- |
| Gene_id , Taxon Identifier,GO-id |
| Symbol |
| Synonyms of Gene Symbols |
| Database References |
| RNA,Protein,MIM,Ensemble |
|  |
| Chromosome , Maplocaion |
| Genetype |
| Full name nomen |
| Authority, Alternate name |
| Gene related with other Taxon |

**Table 5 : Experimental Dataset Corpus**

| Corpus | Concepts | Gene Integration Ontology Concept |
| --- | --- | --- |
| K-C1 | One Primitive Concept | Gene , GeneType Go Functionality ,Go Number |
| KA-C | Two Concepts related with two or more Data Property Gene Type , Gene present in chromosome and map Location | Gene type with Data Property<br><br>Gene Functionalty and Go Term\<br><br>Gene type , Go Term with Data Property |
| K-C2 | Link between two concepts | Gene type and Go Term |

The dataset is extracted from the ontology as shown in Table for concept representations to verify the ontology . The corpus contains maximum of combination of four primitive concepts with different attributes and relationships to verify the ontology. The corpus is evaluated based on the concept I. The three values are {I=1, 2, 3}.Each corpus is the superset of the previous one, when the value of I increase the relations get filtered.
The first dataset evaluates the ontology with concepts which are modeled as primitive classes like gene name with different attributes and relations included as data and object properties. The gene names which are unique is selected and tested verify the consistency of the ontology is consistent. The corpus K-C1 is extracted from the ontology hierarchy with single primitive concepts to evaluate and find its relation with other concepts and links. All the primitive concepts are represented as direct classes.The corpus KA-C consists of ontology hierarchy related within a concepts and its mapping with other functionalities. The corpus K-C2 is two evaluate the ontology links between two or more direct concepts.
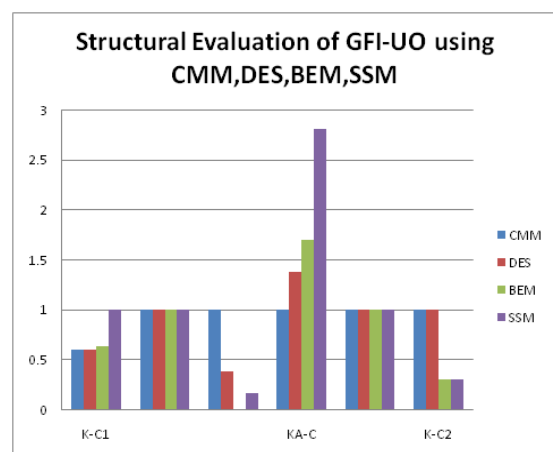


The Fig.shows the results of the structural metrics evaluated on the dataset for three different ontology concepts. The class match measure is very high when there were exact matches and comparatively less when partial matches are considered. The Density measure is found to be increased when more relations exist between concepts in KA-C. The betweennes measure is found to be high when the concepts

is related with more relations in KA-C and minimum value is also reached when the concepts does not have relations as in K-C1 of second dataset. The SSM measure we have found that it has nearer values for all ontology concepts in all datasets. The overall score of the dataset is discussed in the section below.

### Table 6 Overall Score

| Corpus | Overall Score | Rank |
|--------|--------------|------|
| K-C1 | 0.73 | 3 |
| | 1 | 2 |
| | 0.36 | 6 |
| KA-C | 0.6 | 5 |
| | 1 | 1 |
| K-C2 | 0.65 | 4 |
| GO | 0.65 | 4 |

The overall score for the dataset with different ontology concepts is provided in the table 6. The table ra nking is based on the same weight for all metrics. The corpus with single concept K-C1 found to have better scores as the CMM value was high. The Corpus K-C2, KA-C has an average score and can be considered as good ontology model based on all the metrics. We also have noticed there were identical scores in the ontology K-C1 and KA-C. Based on the scores we found that CMM, SSM , DES are important to evaluate an ontology.

### 4.1 Discussion

Results of this experiment we can infer that, the proposed Ontology, obtain high score than Gene ontology. The ontology is evaluated by providing same weights to all the metrics, varying the weights for the ontology results in different scores. The class measure CMM obtains higher values when the classes are exact matches. Partial match of the classes we found that the value is minimum. The Ontologies KA-C and K-C2 was found to be compact, in terms of representing ontology elements like classes and other conceptual relationships related to the classes. The domain terms sought are found for the above said classes. The proposed ontology was found to have better conceptual relationships for representing attributes, which was found lacking in GO ontology. The keywords when given to fetch information using domain terms, it was found that proposed ontology returns results with its relations. The ontology have sufficient quantity of parents, and has defined concepts which does not exist in GO.

The DEM measure returned high values, when there exists richly described concepts with relationships. The ontology KA-C was found to have high score with well defined concepts and relationships. The ontology is evaluated based on syntax using reasoner to find the consistency, structurally evaluated using the metrics as proposed in the paper. The ontology is also evaluated semantically by experts.

## 5. Conclusion

The ontology based approaches proposed in bioinformatics domain gives better classification of data. The ontologies are constructed automatically, manually or semiautomatic approaches. The proposed ontology based OGFR framework is implemented as semi automatic model for integrating gene information using OWL language representation. The ontology concepts for the gene details are clearly specified. The proposed ontology helps the user to extract all possible information related to human gene with good visualization. The proposed ontology is evaluated in terms of syntax and structural evaluation metrics and also evaluated semantically by domain experts.

The ontology is designed with well defined classes and attributes which are related through rich conceptual relationships. Further, it is decided to improve the proposed ontology by converting its creation methodology from semi automatic to automatic and add more new conceptual relationships among ontology concepts to have high degree of association links.

## Acknowledgement

## References

**Journals**
[1] Amal Zouaq, Roger Nkambou," Building Domain Ontologies From Text For Educational Purposes". IEEE Transactions On Learning Technologies, Vol. 1, No. 1, January-March 2008.

[2] Amal Zouaq ,Roger Nkambou," Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project", IEEE Transactions On Knowledge and Data Engineering, Vol. 21, No. 11, November 2009.

[3] Amandeep S. Sidhu,"Protein Data Integration through Ontologies", Published in Proceedings of Nature,2009.

[4] Anna v. Anagnostopoulos, Judith A. Blake, "Using Bio-ontologies as Data Annotation, Integration & Analytical Tools at the Mouse Genome Informatics Resource."

[5] Andre Valente , "Types and Roles of Legal Ontologies", Lecture Notes in Law and Semantic Web Springer-Verlag Berlin Heidelberg 2005.

[5] Antonio, María "A Biomedical Ontology for Nutrigenomics Research.", Published in Proceedings of International Conference on BioMedical Engineering and Informatics,2008.

[9] Huiru Zheng et.al.,"Integration Of Genomic Data For Inferring Protein Complexes From Global Protein–Protein Interaction Networks", IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 38, No. 1, February 2008.

[10] Hisham Al-Mubaid, Hoa A. Nguyen, "Measuring Semantic Similarity Between Biomedical Concepts Within Multiple Ontologies", IEEE Transactions On

Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 39, No. 4, July 2009.

[11] Jo¨El Plisson, Peter LjubiˇC, Igor MozetiˇC, And Nada LavraˇC IEEE "An Ontology For Virtual Organization Breeding Environments " ,Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 37, No. 6, November 2007 1327.

[12] Lawrence Hunter, "OpenDMAP: An open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression", BMC Bioinformatics 2008.

[14] Madhusudan Therani," Ontology Development For Designing And Managing Dynamic Business Process Networks", IEEE Transactions On Industrial Informatics, Vol. 3, No. 2, May 2007.

[16] Olivier Corby et.al., "Searching The Semantic Web: Approximate Query Processing Based On Ontologies", Published in IEEE Computer Society,2006.

[18] Robert Stevens et.al., "Ontology-based knowledge representation for bioinformatics", Published in Briefings in Bioinformatics, 2000.

[19] Stefano Bianchi, Anna Burla, Costanza Conti, " Biomedical Data Integration – Capturing Similarities while Preserving Disparities.", Published in Proceedings of 31st Annual International Conference of the IEEE .

[20] William J. Bug et.al.,"The NIFSTD and BIRNLex Vocabularies: Building Comprehensive Ontologies for Neuroscience", Published in Journal Neuroinform ,2008.

[23] Yu-Ting Huanga Et.al.," Automatic Extraction Of Information About The Molecular Interactions In Biological Pathways From Texts Based On Ontology and Semantic Processing", Published in IEEE International Conference On Systems, Man, and CyberneticsOctober 8-11, 2006, Taipei, Taiwan

.**Proceedings**

[3] Amandeep S. Sidhu,"Protein Data Integration through Ontologies", Published in Proceedings of Nature,2009.

[6] Anne M. Maglia et.al.,"An Anatomical Ontology For Amphibians", Published in Proceedings of Pacific Symposium on Biocomputing,2007.

[7] Barry Smith, PhD et.al. , "The Ontology of the Gene Ontology", Published in Proceedings of AMIA Symposium 2003.

[8] Gruber T.R., "The role of common ontology in achieving sharable, reusable knowledge bases" In: Proceedings of KR'1991: Principles of Knowledge Representation and Reasoning. pp. 601–2, 1991.

[13] Lingxian Yang, Jinguang Gu and Heping Chen. "Clustering Algorithm Based on Semantic Distance for XML Documents.", Published in Proceedings of First International Workshop on Database Technology and Applications,2009.

[15] Ma. Auxilio Medina, Alberto Ch´avez-Arag´on, R. ,"Construction, Implementation and Maintenance of Ontologies of Records " Published in Proceedings of the Fourth Latin American Web Congress (LA-WEB'06).

[21] Xiang-hua Xu, Jia-lai Huang," A Method for Measuring Semantic Similarity of Concepts in the Same Ontology." Published in Proceedings of International Multi-symposiums on Computer and Computational Sciencesm,2008.

[22] Xang Feng, Wen Hui, Cheng Wei," OWL-based Emergency Preplan Ontology Creation and Application in Knowledge Sharing", Published in Proceedings of IEEE.International Conference on Information and Automation,June 22 -25, 2009.

## Authors

B.L.Shivakumar received Ph.D. in Computer Science from Bharathiar University, Coimbatore. M.Phil. in Computer Science from Manonmaniam Sundaranar University, in 2003 and M.Sc. in Computer Science from Bharathidasan University, in 1996. He also received Post Graduate Diploma in Business Administration (PGDBA), Co-operative Management (PGDCM) and Bachelor of Library and Information Science from Annamalai University. In 1997, he joined SNR Sons College as a Lecturer in the department of Computer Science, and currently is the Professor and Head of the department of Computer Applications. He has authored or co-authored over 15 Research Papers in journal and conferences. He is recipient of Bharat Jyoti award conferred by The India International Friendship Society, New Delhi and Best Programme Officer award by Bharathiar University. His interest includes Computer Forensic Science, Digital Image Processing and Cloud computing.

**Ms. V. Bhuvaneswari** received M.Phil in Computer Science at Bahrathiar University in 2003, Masters Degree (MCA) in Computer Applications in IGNOU in 2002 and the Bachelors Degree (B.Sc.) in Computer Technology in 1997 from PSG Tech,. She has qualified UGC-JRF in the year 2003. She is pursuing her doctoral research at Bharathiar University in the area of Data mining. Currently she is working as Assistant Professor in department of Computer Applicaion, Bharathair University, Coimbatore,India. Her research interest includes Data Mining, Computational Biology, Bioinformatics and Evolutionary Computing. She has authored more than 20 papers in Journals and Conferences.