

# Developing a Taxonomic Information Model for Biological Data using Ontology

Khin Myo Sett

**Abstract:** In real world applications, the amount of available information and data increases exponentially. Since this growth does not entail as just as large quality increase of the available knowledge, methods and tools are needed to filter and process this mass of information. For this purpose, ontologies which represent data and their interrelations in a computer “interpretable” form get more and more established. Because of the implementation of ontology-based biological information system is presented in this thesis. To observe the ontology-based biological information, “A Manual of Practical Zoology Chordates” written by P.S.VERMA is used as the resource available. The first implementation with Neo4j Community Server installation, configuration, and biology based ontology graph nodes are created in the database server. Taxonomic Biological ontology model is conceptually implemented in biological domain and class orientation of superclass and subclasses hierarchy structure is described. Java implementation of Mynode class, query engine class, search and admin search for any simple query and more complex query is also presented. Secondly, in the java implementation, detail search, group search and admin user for database operations are provided. The most commonly used standard format can be obtained by the use of ontology. Evaluation of taxonomic biological information retrieval using Cypher query and conceptual visualization results are presented. In this thesis, it has been evaluated consistency of the query results with the results of set theory, and the precision and the recall

**Index Terms:** Ontology, Biological Data, Taxonomic Biological Ontology Model

## I. INTRODUCTION

Ontologies and taxonomies provide a structure to the concepts and language used to organize knowledge. Geoffrey P Malafsky and Brian D Newman [26] have described organizing knowledge with ontologies and taxonomies. Without them, the knowledge will inevitably be difficult to find and reuse as people have very different perspectives on how the knowledge is related in the context of their situations. The implementations of ontologies and taxonomies have three options: develop the ontology then develop the supporting taxonomy, develop taxonomy and then develop the over-arching ontology and develop the two in parallel.

Taxonomy is a simple hierarchical arrangement of entities where the data has a parent-child kind of relationship. This is very much similar to the taxonomies in the field of Biology, for example: Panther is a type of Carnivorous animal, which intern is a type of Mammal and all Mammals are types of Animals. Taxonomies are simple arrangements of classes without any restrictions on any properties at any level of the hierarchy.

Ontology is a more complex variation of taxonomy.

Besides having the hierarchical arrangement of the classes that represent entities, each class has intern several restrictions on its relationships to other classes or on the properties a particular class is allowed to possess. An example of this would be that for a hierarchy in the Animalia domain, if Chondrichthyes is a parent of common name of Shark and Dog-Fish etc. and then add another simple statement representing scientific name is another separate entity which is scoliodon and chiloscyllium. It is basically referring to a simple taxonomy. But if the constraints are added on the relationships for these entities, for example such that there exists a relationship named "is\_a\_ClassName" that is a prerogative of only the Chondrichthyes.

In this paper, we proposed an ontology for biological data. To observe the ontology-based biological information, “A Manual of Practical Zoology Chordates” written by P.S.VERMA is used as the resource available. The first implementation with Neo4j Community Server installation, configuration, and biology based ontology graph nodes are created in the database server. Taxonomic Biological ontology model is conceptually implemented in biological domain and class orientation of superclass and subclasses hierarchy structure is described. Java implementation of Mynode class, query engine class, search and admin search for any simple query and more complex query is also presented. Secondly, in the java implementation, detail search, group search and admin user for database operations are provided. The most commonly used standard format can be obtained by the use of ontology. Evaluation of taxonomic biological information retrieval using Cypher query and conceptual visualization results are presented.

This paper is organized as follow. Section 2 describes some related works of this work. Section 3 presents the proposed method. Section 4 analyzes the proposed ontology. Finally, section 5 concludes the paper.

## II. RELATED WORKS

Ontologies have become popular in the fields of intelligent information integration, cooperative information systems, electronic commerce and knowledge management. Recently, ontologies for knowledge representation in the biological domain have also appeared.

Vaida Jakoniene and Patrick Lambrix [12] have proposed ontology-based integration for bioinformatics. This system presents the large variety of bio-ontologies. They differ in the type of biological knowledge they describe, their intended use, the adopted level of abstraction and the knowledge representation language. For instance, via OBO (Open Biomedical Ontologies) they have accessed a number of ontologies having different

Revised Version Manuscript Received on 12 July 2018.

Dr. Khin Myo Sett, Department of Computer Studies, Mandalay University, Mandalay, Myanmar, E-mail: [khinmyosett.mu@gmail.com](mailto:khinmyosett.mu@gmail.com)

## Developing a Taxonomic Information Model for Biological Data using Ontology

Biological focus and that are developed for different purposes. GO (Gene Ontology) ontologies describe biological process, molecular function and cellular components of genes and proteins in all organisms. The goal is to produce structured, precisely defined, common and dynamic ontologies that can be used for annotating gene products.

Erick Antezana, Martin Kuiper and Vladimir Mironov [1] have described biological knowledge management. This system presents bio-ontologies and most of the bio-ontologies are hosted under the umbrella of the OBO foundry. The OBO foundry is a collaborative effort not only for gathering bio-ontologies but also to provide a set of foundational principles to structure the further co-ordinate development of bio-ontologies. The OBO foundry now constitutes a set of 53 domain-specific candidate ontologies, which are becoming widely accepted as a reference by the life science community.

Natalia Sklyar [14] has surveyed the existing bio-ontologies that are intended to provide a brief state-of-art introduction into ontology-based biological systems. This investigation is aimed to the role of ontologies and the possibilities opened by the use of ontologies in bioinformatics. For every ontology, the structure, scope, area of application, characteristics, representation language etc. are considered. A comparison of ontologies and also future possibilities for the use of bio-ontologies are discussed. Some selected bio-ontologies are presented. They are Transparent Access to Multiple Bioinformatics Information Sources (TAMBIS), Gene Ontology (GO), and Ontology for Molecular Biology (MBO), EcoCyc Ontology and Cell Signaling Ontology. The bio-ontologies described above are not the only ones that deal with the bio-molecular data. Existing ontologies, in general, have much more differences than similarities. They differ in their intention, structure, their coverage, and detail level.

Nico M. Franz and David Thau [13] have explained that the prospects of integrating biological taxonomies into an ontological reasoning framework are reviewed. Traditionally ontological representations of taxonomy have adopted the model of a single and static hierarchy. This model is contrasted with a more realistic situation involving dynamic revisions of particular groups and alignments among alternative taxonomic perspectives. Taxonomic practice is bound by a range of epistemological constraints and linguistic conventions that run orthogonal to the logical background from which ontological entities and relationships originate, resulting in severe challenges for ontological representation and reasoning. In particular, the purported existence of a single hierarchy in nature forces taxonomists to gradually approximate this hierarchy and make frequent rearrangements in light of new evidence. The taxonomy/ontology interface should focus either on strictly nomenclatural entities and relationships or on ontology-driven strategies for aligning multiple taxonomies, but not on building static networks for large portions of the tree of life. The prospects of using ontology-based services in taxonomy will largely depend on the ability of the taxonomic expert community to present its products in ways that are more compatible with ontological principles than concurrent practice.

Lars Vogt [15] has presented that role of bio-ontologies for developing a general data standard in biology. Due to lack of common data standards, the communicability and comparability of biological data across various levels of

organization and taxonomic groups is continuously decreasing. However, the interdependence between molecular and higher levels of organization is of growing interest and calls for co-operations between biologists from different methodological and theoretical backgrounds. A general data standard in biology would greatly facilitate such co-operations. This article examines the role that defined and formalized vocabularies (i.e. ontologies) could have in developing such a data standard. They suggested basic criteria for developing data standards on grounds of distinguishing content, concept, nomenclatural, and format standards and discuss the role of data bases and their use of bio-ontologies in current activities for data standardization in biology. General principles of ontology development are introduced, including foundational ontology properties (e.g. class, subclass, parthood), and how concepts are defined. After addressing problems that are specific to morphological data, the notion of a general structure concept for morphology is introduced and why it is required for developing a morphological ontology. The necessity for a general morphological ontology to be taxon-independent and free of homology assumptions is discussed and how it can solve the problems of morphology. The article concludes with an outlook on how the use of ontologies will likely establish some sort of general data standard in biology and why the development of a set of commonly used foundational ontology properties and the use of globally unique identifiers for all classes defined in ontologies is crucial for its success.

Stefan Schulz, Holger Stenzhorn and Martin Boeker [16] have described that the ontology of biological taxa in which different approaches on how to represent biological taxa using existing standards for biomedical ontologies such as the description logic OWL DL and the Open Biomedical Ontology Relation Ontology. The researchers have demonstrated that how hidden ambiguities of the species concept can be dealt with and existing controversies can be overcome. A novel approach is to envisage taxon information as qualities that inhere in biological organisms, organism parts and populations. The presented methodology has been implemented in the domain top-level ontology BioTop. BioTop may help to improve the logical and ontological rigor of biomedical ontologies and further provides a clear architectural principle to deal with biological taxa information. For their practical purpose of biomedical ontologies the formalization of species or more generally of biological taxa that they proposed, is intended to be neutral to the different and conflicting species conceptualizations. It departs from the principle that biological taxa are something that regardless of its existence in nature or its (fiat) attribution by biologists has a highly ranked importance in biology and therefore requires to be accounted for in biomedical ontologies.

### III. PROPOSED METHOD

#### A. Ontologies

Ontologies specify the primary concepts and the relationships among the concepts in a particular domain.



The term means several things depending on the field in which it is used. In philosophy, ontology is concerned with the metaphysical nature and relationships of being.

In contrast, computer science uses ontologies to describe specific conceptual terms and relationships in a standardized machine readable format. Any knowledge management effort must grapple with the challenge that there are several viable and valid perspectives on any given topic or business domain. To make the knowledge useful and an effective enabler of organizational success, the knowledge management manager must create a single shared understanding among people of what the knowledge means to the organization within the context of its business domain and how it is intended to be used. Ontology provides this unifying map of concepts and relationships.

The ontology can be represented either graphically or in a structured text format. The former is usually used when the primary goal is to forge a shared understanding of the domain and provide guidance to the members of the group. The latter approach is most often used for computer applications that perform language analysis and concept matching, such as the goal of greater automated semantic capabilities on the Internet (i.e. the Semantic Web). Machine readable ontologies require a computer language to define the concepts and associated relationships. One standard language is the OWL Web Ontology Language developed by the World Wide Web Consortium (W3C). [18]

### B. Taxonomy

Taxonomies are the classification scheme used to categorize a set of information items using a hierarchical structure, which is a tree-like structure, with the most general category as the root of the tree. Each node is an information entity that represents some object in the world that is being modeled. Each link between two nodes in the taxonomy represents a sub classification of relationship. The biological classification of the structure is the most commonly known of life itself. This is described in terms of phylum, family, genus, and species and so on. "Taxonomies are things of interest arranged in a hierarchical structure, typically in a supertype/subtype relationship."

They represent an agreed vocabulary of topics arranged around a particular theme. Although they can have either a hierarchical or non-hierarchical structure, or hierarchical taxonomies such as in libraries, biology, or military organizations are typically encountered. This type has a tree-like structure with nodes branching into sub-nodes where each node represents a topic with a few descriptive words.

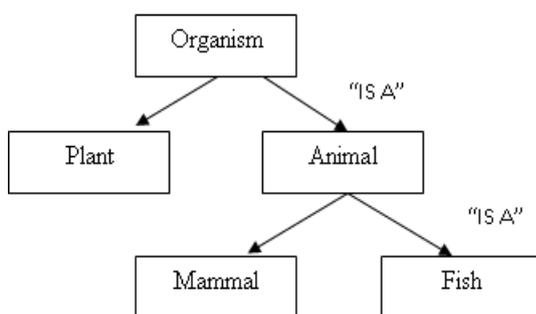


Figure. 1 Example of an ontology (in a Hierarchical Form)

Aarti Krishnan [17] has explained that ontology is a formal description of concepts and relationships that can exist for a particular domain of knowledge. Like a dictionary, ontologies collect and organize related entities. It describes accurately the structure of a main domain and how its entities are inter-related. The backbone of an ontology is most often a taxonomy (classification of entities in a hierarchical form, Figure 1), but is however not limited to taxonomical hierarchies alone. The structure is usually in the form of a tree (or graph), where the entities are specified as the nodes and the edges specify the relationship between them. The graph is directed and acyclic, where the terms follow an "IS A" relation with one another and their definition gets more specialized as we move down the graph.

### C. Ontology based Information System

Information retrieval (IR) system, the key functionality of the emerging "semantic Web", is one of the main challenges over the real world applications. Ontologies now appear to be a de facto standard of semantic IR systems. By defining key concepts of a domain, they introduce a common vocabulary that facilitates interaction between the user and the software. Meanwhile, by specifying relationships between concepts, they allow semantic inference and enrich the semantic expressiveness for both indexing and querying document corpus. It is natural that works relating to ontology integration in IRS are growing. Abdelkrim Bouramoul, Mohamed-Khireddine Kholadi and Bich-Lien Doan [5] have described the solution is to build ontology from the corpus on which IR tasks will be performed and the reuse of existing resources. Ontologies as a support for the modeling of IRS have been discussed in the literature and the contribution of ontologies in an IRS can be understood at three levels:

- In the document indexing process: by combining it with the techniques of natural language processing, the documents in the database will be summarized and linked to the ontology concepts. If this step has been properly done, the search would be easier in the future.
- At the queries reformulation level in order to improve the initial user queries. This aspect was also used as a complement to search this type of such queries.
- In the information filtering process, this aspect will be the subject of the contribution that we present in this paper. The idea is to use ontology to add the semantic dimension to the evaluation process. This can be done by extracting the query terms and their semantic projection using the WordNet ontology on the set of returned documents. The result of this projection is used to extract concepts related to each term.

### D. Biological Ontology

Ontology-based biological information system is proposed in this thesis. Among the ontology domains, biological ontology has been constructed in this research area but ontology-based biological information system is constructed in graph database is the first one approach of this IR system. This system facilitates query in different styles such as single query, group query, and cypher query in graph view. Though most IR systems rely on ontologies, they often use one of the two following extreme approaches:



## Developing a Taxonomic Information Model for Biological Data using Ontology

Either they use most of the semantic expressiveness of the ontology and hence require complex query languages that are not really appropriate for non-specialists; or they provide very simple query language that almost reduces the ontology to a dictionary of synonyms used in Boolean retrieval models.

Another drawback of most IR systems is the lack of expressiveness of their results. In most cases, results are simply proposed as a set of documents with no further explanations concerning the match between the documents and the query.

Biological knowledge is inherently complex and so cannot readily be integrated into existing databases of molecular data. Ontology is a formal way of representing knowledge in which concepts are described both by their meaning and their relationship to each other. Unique identifiers that are associated with each concept in biological ontologies (bio-ontologies) can be used for linking to and querying molecular databases.

### E. Model of Ontology-Based Biological Information System

Ontology-based biological information system is described in this thesis which provides the efficient and semantically related information retrieval features. Queries can be processed single search for detailed description, group search, and graph view with cypher query with colorful graph nodes. The design of this system is discussed in this section in detail. The model of ontology construction is the main component in this system and conceptualization in biology concept should be clear and which has completed description, many properties, and various relationships. Query engine component facilitates the detailed search and group search for any simple query and more complex query. Graph engine component provides graph search for cypher query.

### F. Biological Ontology Construction

The key important factor for IR system based on ontology is construction ontology. The ontology construction model captures domain specific knowledge for both humans and computer applications. A set of vocabulary definitions that embedded a community's knowledge of a domain is considered in this ontology construction model component. Ontology takes a variety of forms, but necessarily it includes a vocabulary of terms, and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms. Biological research involves synergy of data, models, and knowledge. Each of these areas has a more abstract representation, which a person thinks and reasons about, and a more computational representation, which is what the computer uses. A challenge for the biologist is to bridge the cognitive gap between the abstract representation and the computational representation.

The biological ontology can construct the different types of data structures available, their characteristics, the contexts in which they are useful, the well-known problems in computer science, and the algorithms corresponding to them. Sandhya Revuri, Sujatha R Upadhyaya and P Sreenivasa Kumar [15] have presented for using domain ontologies for efficient information retrieval and for building ontology, all the possible concepts are listed in the domain. And then the properties of each of these concepts are identified. Once the

properties are identified the domain and range are defined. The next step would be identifying the characteristics of the properties like Transitive, Symmetric, Functional and Inverse Functional. Value or cardinality constraints on properties can be posed to make the concept description more specific. Value constraints are of the form all Values From, some Values From and has Value. Cardinality constraints are min Cardinality, max Cardinality and cardinality.

Biological ontologies construction considers animalia as the domain of this thesis. There are six main classes according to the class name and they are chondrichthyes, osthychthyes, amphibia, reptilian, aves and mamalia. Each class has many animals which describe common name, scientific name, feeding habit, breeding habit, habitat, locomotion, body cover, etc.

The following steps are used in construction model of ontology based biology.

- Captures knowledge for biological domain
- Creates a shared understanding – between humans and for machine
- Makes knowledge machine process
- Makes meaning explicitly by definition and context
- Create main categories or classes
- Create properties for each class
- Create relationship between each class
- Create instances under each class

Start the connection with Neo4j database which is a graph database is presented. Input the first node data which is defined as root node and process the root node creation process. Repeatedly a new node and relationship up to user stop condition. Every created node is stored in biological ontology graph database. After all nodes are added, disconnect the database connection and output shows the database creation successful message.

### G. Query Engine

One of the important portions of this research is data visualization. A graph database stores data in a graph, the most generic of data structures, capable of elegantly representing any kind of data in a highly accessible way.

Neo4j is a graph database, which means that it does not use tables and rows. Instead, it uses nodes and relationships. Both nodes and relationships can have a number of properties. While relationships must have one direction and one type, nodes can have a number of labels. For example, the following Figure 2 shows eight nodes and their relationships, where every node has a label while relationships have a type (is\_a\_ClassName, is a, has and lives in). The properties used in the graph are name, label, and habitat, etc as shown in the Figure 2.

Query engine component provides for querying the detailed search any animal and group search of the same features such as feeding habit, breeding habit, etc. For example the query for Shark with common name produces Common name is Shark, Scientific name is Scoliodon, feeding habit type is Carnivore, breeding habit type is Viviparous, habitat is Marine Water, locomotion has Fin, body cover has Scale and is illustrated in Figure 3.



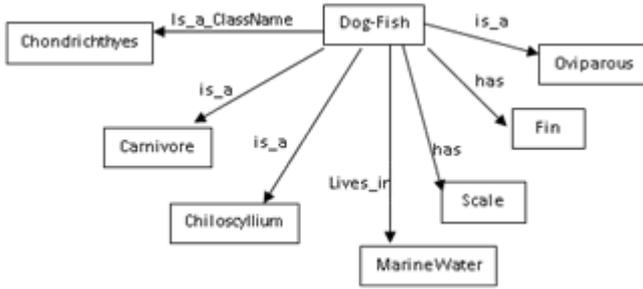


Figure 2 Sample Graph Database

|                         |                  |
|-------------------------|------------------|
| <b>Common Name:</b>     | <b>Shark</b>     |
| <b>Scientific Name:</b> | <i>Scoliodon</i> |
| <b>Feeding Habit:</b>   | Carnivore        |
| <b>Breeding Habit:</b>  | Viviparous       |
| <b>Habitat:</b>         | Marine Water     |
| <b>Locomotion:</b>      | Fin              |
| <b>Body cover:</b>      | Scale            |

Figure 3 Sample of Simple Query by Common Name (Shark)

In detailed search using common name, the query engine component searches for six main classes and continue for retrieving all common names with the same relationship according to the user query common name search and produces the detailed info as an output.

In detailed search using scientific name, the query engine component searches for six main classes and continue for retrieving all scientific names with the same relationship according to the user query scientific name search and produces detail information as an output.

In group search, there are five options types and can be combined any option types. If the relationship types exist with desired option types, the query engine displays the group information otherwise does not exist message. These options types are feeding habit, breeding habit, habitat, locomotion and body cover. Feeding habit option type has six subgroups such as Carnivore, Herbivore, Insectivore, Omnivore, Folivore and Frugivorous. Breeding habit option type has two subgroups such as Viviparous, Oviparous. Habitat option type has five subgroups such as Marine Water, Fresh Water, Terrestrial, Arboreal, and Tundra. Locomotion option type has five subgroups such as Fin, 4-Limbs, Wings, 2-Limbs, 4-Legs. Body cover option type has eight subgroups such as Scale, Scutes, Rings, Dermal Ring, Spine, Feather, Hair, and Shell.

**H. Graph Engine**

Graph engine component is based on neoclipse and input query type is cypher query and produces the graph node view as in Figure 4.

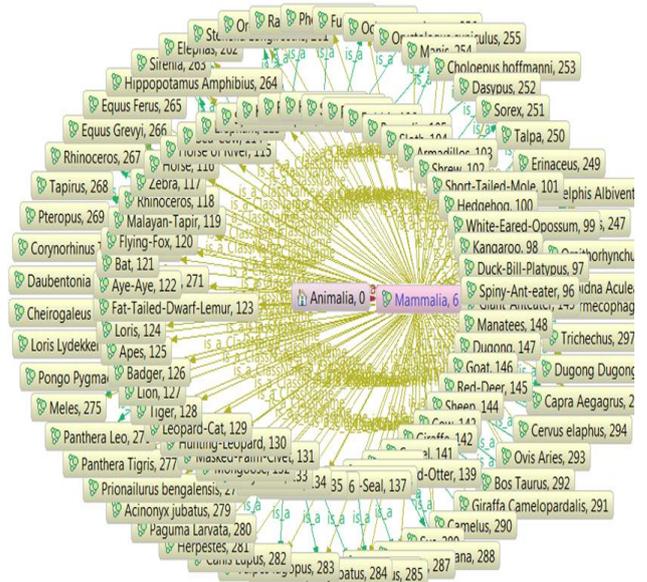


Figure 4 Classes, Property types and Properties of Biological Ontology

**IV. ANALYSIS OF BIOLOGICAL ONTOLOGY**

Biology is described in multiple kinds of entities and ontologies. Large ontologies provide categories and relations for the basic features found in databases of model organisms. Basic concept is used for integrating the generated data and interpreted by multiple heterogeneous groups and stored in distributed biological databases throughout the world. The use of a common vocabulary and common formal descriptions of the vocabulary's terms permit the comparison, retrieval and analysis of the data stored in these databases.

From a developer's perspective, not only performance, but also initial learning effort, code readability and maintainability are relevant when choosing a query language. In this chapter, Cypher is a declarative graph query language and efficient querying and updating of the graph store. Cypher is a relatively simple but still very powerful language. Very complicated database queries can easily be expressed through cypher. Cypher is designed to be a humane query language, suitable for both developers and operations professionals. Simple Cypher query syntaxes are discussed in this section.

Cypher is derived its structure from SQL queries which are built up using various clauses. Clauses are chained together, and they feed intermediate result sets between each other. For example, the matching identifiers from one MATCH clause will be the context that the next clause exists in.

The query language is comprised of several distinct clauses. Here are a few clauses used to read from the graph:

- **MATCH:** The graph pattern to match. This is the most common way to get data from the graph.
- **WHERE:** Not a clause in its own right, but rather part of MATCH, OPTIONAL MATCH and WITH. Adds constraints to a pattern, or filters the intermediate result passing through WITH.
- **RETURN:** What to return.



## A. Creation of Ontology Based Bio Data

For the proposed system, firstly bio data are created as ontological commitments. For example, in this system Class Animalia that may be created based on ontology for teaching aid. Firstly, bio data was created in spreadsheet as CSV format file which supports that simply represents relationships. The CSV format can be obtained from any row data, databases or Excel export. Each line must contain at least two elements. Each row of the file is a node or an edge. Working with spreadsheets are comfortable can be manipulated data in this environment. That is why; this is often a very suitable way of conceptual data based on ontology.

First, create a node table (Table 1) which has a unique identifier (first column: node), a name (second column), and a type/label (third column) and as needed (e.g. Scientific Name, feeding habit, breeding habit, habit, location, locomotion, body cover). Among these columns, third column to the last column are properties or attributes. These data are imported in a meaningful way into a graph database management system like Neo4j, and it has some additional information about the relationships between these data. In Table 2, data are interconnected between their nodes vice versa.

The two CSV format files: nodes.csv and rels.csv are created with all nodes and relationships and described in Table 1 and Table 2. Sample query syntax for loading data.csv and rels.csv are described in Figure 1 and Figure 4.

## B. Analysis by Using Venn Diagrams

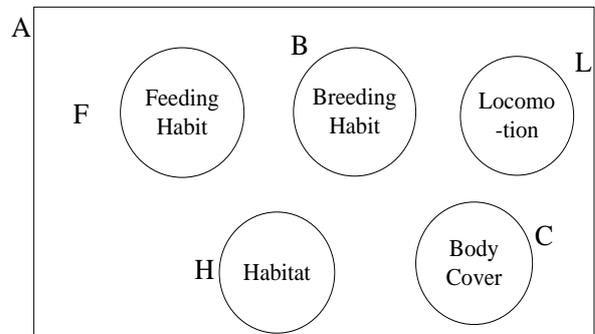
Diagrams make mathematics easier because they help the system to see the whole situation at a glance. These diagrams are now called Venn diagrams.

In most problems involving sets, it is convenient to choose a larger set that contains all of the elements in all of the sets being considered. This larger set is called the universal set. In a Venn diagram, the universal set is generally drawn as a large rectangle, and then other sets are represented by circles within this rectangle.

In this system, the universal set A {Animalia} has five sets. If B = { Breeding Habit }, F = { Feeding Habit } H = { Habitat }, L = { Locomotion }, C = { Body Cover }, then

$\{ \text{Breeding Habit} \} \cup \{ \text{Feeding Habit} \} \cup \{ \text{Habitat} \} \cup \{ \text{Locomotion} \} \cup \{ \text{Body Cover} \} = \{ \text{Animalia} \}$ . The set A {Animalia} s as shown in Figure 5.

In this system, the set F = {Feeding Habit} has six sets. If C = { Carnivore }, H = { Herbivore }, I = { Insectivore }, Fl = { Folivore }, O = { Omnivore } and R = { Frugivorous }, then  $\{ \text{Carnivore} \} \cup \{ \text{Herbivore} \} \cup \{ \text{Insectivore} \} \cup \{ \text{Folivore} \} \cup \{ \text{Omnivore} \} \cup \{ \text{Frugivorous} \} = \{ \text{Feeding Habit} \}$ . Figure 6 shows the Venn diagram of Feeding Habit.

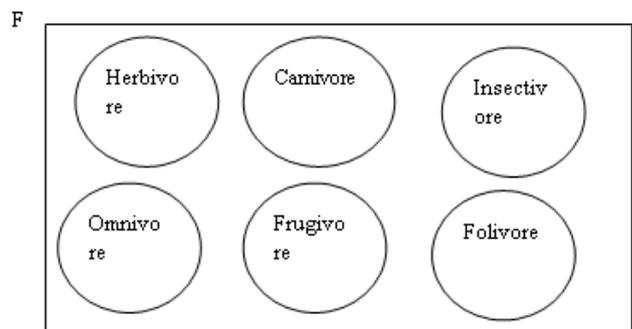


**Figure 5 Venn Diagram of Animalia**

Similarly, in this system, the set: (1) B = { Breeding Habit } has two sets. If V = { Viviparous }, O = { Oviparous }, then  $\{ \text{Viviparous} \} \cup \{ \text{Oviparous} \} = \{ \text{Breeding Habit} \}$ .

(2) H = { Habitat } has five sets. If M = { Marine Water }, Fr = { Fresh Water }, T = { Terrestrial }, Ar = { Arboreal }, Tu = { Tundra },

then  $\{ \text{Marine Water} \} \cup \{ \text{Fresh Water} \} \cup \{ \text{Terrestrial} \} \cup \{ \text{Arboreal} \} \cup \{ \text{Tundra} \} = \{ \text{Habitat} \}$ .



**Figure 6 Venn Diagram of Feeding Habit**

**Table 1. Nodes. CSV File**

| Node | Name           | Label       | Scientific Name  | Feeding Habit | Breeding Habit | Habitat      | Locomotion | Body Cover |
|------|----------------|-------------|------------------|---------------|----------------|--------------|------------|------------|
| 0    | Animalia       | rootnode    | -                | -             | -              | -            | -          | -          |
| 1    | Chondrichthyes | ClassName   | -                | -             | -              | -            | -          | -          |
| 2    | Ostrichthyes   | ClassName   | -                | -             | -              | -            | -          | -          |
| 3    | Amphibia       | ClassName   | -                | -             | -              | -            | -          | -          |
| 4    | Reptilia       | ClassName   | -                | -             | -              | -            | -          | -          |
| 5    | Aves           | ClassName   | -                | -             | -              | -            | -          | -          |
| 6    | Mammalia       | ClassName   | -                | -             | -              | -            | -          | -          |
| 7    | Shark          | Common Name | <i>Scoliodon</i> | Carnivore     | Viviparous     | Marine Water | Fin        | Scale      |

|    |                     |             |                       |           |            |              |     |       |
|----|---------------------|-------------|-----------------------|-----------|------------|--------------|-----|-------|
| 8  | Dog-Fish            | Common Name | <i>Chiloscyllium</i>  | Carnivore | Oviparous  | Marine Water | Fin | Scale |
| 9  | Tiger-Shark         | Common Name | <i>Stegostoma</i>     | Carnivore | Viviparous | Marine Water | Fin | Scale |
| 10 | Hammer-Headed-Shark | Common Name | <i>Sphyrna Zygaen</i> | Carnivore | Viviparous | Marine Water | Fin | Scale |
| 11 | Guitar-Fish         | Common Name | <i>Rhinobatus</i>     | Carnivore | Viviparous | Marine Water | Fin | Scale |
| 12 | Sting-Ray           | Common Name | <i>Trygon</i>         | Carnivore | Viviparous | Marine Water | Fin | Scale |
| 13 | Eagle-Ray           | Common Name | <i>Myliobatis</i>     | Carnivore | Viviparous | Marine Water | Fin | Scale |

Table 2. Rels. CSV File

| From | Name           | Relationship   | To | Name                |
|------|----------------|----------------|----|---------------------|
| 0    | Animalia       | has            | 1  | Chondrichthyes      |
| 0    | Animalia       | has            | 2  | Osthychthyes        |
| 0    | Animalia       | has            | 3  | Amphibia            |
| 0    | Animalia       | has            | 4  | Reptilia            |
| 0    | Animalia       | has            | 5  | Aves                |
| 0    | Animalia       | has            | 6  | Mammalia            |
| 1    | Chondrichthyes | Is_a_ClassName | 7  | Shark               |
| 1    | Chondrichthyes | Is_a_ClassName | 8  | Dog-Fish            |
| 1    | Chondrichthyes | Is_a_ClassName | 9  | Tiger-Shark         |
| 1    | Chondrichthyes | Is_a_ClassName | 10 | Hammer-Headed-Shark |
| 1    | Chondrichthyes | Is_a_ClassName | 11 | Guitar-Fish         |
| 1    | Chondrichthyes | Is_a_ClassName | 12 | Sting-Ray           |
| 1    | Chondrichthyes | Is_a_ClassName | 13 | Eagle_ray           |

## V. CONCLUSION

In this paper, the model of ontology-based biological information system has been described and query can be processed by single search, group search, and graph view with cypher query with colorful graph nodes. Then biological ontology construction, query engine and graph engine component and graph database management module component have been proposed and these can be managed for creating, querying detailed description of bio data. Neo4j cypher query syntax is described. This model will be supported for the researchers and learners to search detail information of bio data with colorful images and conceptual relationship with between them.

## REFERENCES

- Aitken, S. and S. Reid, Evaluation of an Ontology-Based Information Retrieval Tool, AIAI, University of Edinburgh, December 2014, <http://citeseerx.ist.psu.edu>
- Antezana, E., M. Kuiper and V. Mironov, Biological Knowledge Management: The Emerging Role of the Semantic Web Technologies, Briefings in Bioinformatics, Advance Access publication, Oxford University Press, Vol. 10, No 4, pp 392-407, May 2009
- Batra, S. and C. Tyagi, Comparative Analysis of Relational And Graph Databases, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Vol. 2, Issue 2, May 2012
- Bordoloi, S. and B. Kalita, Designing Graph Database Models from Existing Relational Databases, International Journal of Computer Applications (0975-8887), Vol. 74, No. 1, July 2013
- Bouramoul, A., M. K. Kholadi and B. L. Doan, An Ontology-based approach for semantic ranking of the web search engines results, <http://www.semantic-web-journal.net/sites/default/files/swj248.pdf>
- Brank, J., M. Grobelnik and D. Mladenic, A Survey Of Ontology Evaluation Techniques, Jozef Stefan Institute, Slovenia, <http://eprints.pascal-network.org>
- Calado, P. et al 9, Searching Web Databases by Structuring Keyword-Based Queries, ACM 1-58113-492-4/02/0011, CIKM'02, Mclean, Virginia, USA, November 2002
- Chen, Y. et al, Keyword Search on Structured and Semi-Structured Data, SIGMOD'09, Providence, Rhode Island, USA, ACM 978-1-60558-551-2/09/06, July 2009
- Ciravegna, F., (LP)2: Rule Induction for Information Extraction Using Linguistic Constraints, Technical Report, CS-03-07, Computer Science, University of Sheffield, September 2003
- Coto, A. L., The Use of Ontology for Improving Image Retrieval and Annotation, Technical Report, kmi-08-08, KMi Institute, September 2008.
- Franz, M. N. and D. Thau, Biological Taxonomy and Ontology Development: Scope and Limitations, Biodiversity Informatics, Vol. 7, 2010, pp. 45-66
- Jakoniene, V. and P. Lambrix, Ontology-based Integration for Bioinformatics, Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005, <https://www.ida.liu.se/~patla/publications/05odbis.pdf>.
- Joshi, M. D. et al, Image Retrieval and Re-Ranking Techniques-A Survey, Signal and Image Processing: An International Journal (SIPIJ), Vol. 5, No. 2, April 2014
- Krishnan, A., Knowledge bases, Ontologies and Key-Value Stores, [www.cbrg.ethz.ch/education/SDB/L4.pdf](http://www.cbrg.ethz.ch/education/SDB/L4.pdf)
- Malafsky, G.P. and B. D. Newman, Organizing Knowledge with Ontologies and Taxonomies, TECh2, Fairfax, VA, <http://www.techi2.com>
- Schulz, S., H. Stenzhorn and M. Boeker, The Ontology of Biological Taxa, Bioinformatics, Vol. 24, ISMB, 2008, pp 1313-1321, doi:10.1094/bioinformatics/btn158, 2008
- Sklyar, N., Survey of Existing Bio-Ontologies, September, 2001, [citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.226.2803...](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.226.2803...)
- Vogt, L., The Future Role of Bio-Ontologies for Developing a General Data Standard in Biology: Chance and Challenge for Zoo-Morphology, © Springer-Verlag, doi: 10.1007/s00435-008-008-5, October 2008

