# InBiodiv-O: An Ontology for Biodiversity Knowledge Management

Archana Patel, Eastern International University, Vietnam\* Sarika Jain, National Institute of Technology, Kurukshetra, India https://orcid.org/0000-0002-7432-8506

Narayan C. Debnath, Eastern International University, Vietnam Vishal Lama, Amdocs Development Centre India LLP, India

#### ABSTRACT

To present the biodiversity information, a semantic model is required that connects all kinds of data about living creatures and their habitats. The model must be able to encode human knowledge for machines to be understood. Ontology offers the richest machine-interpretable semantics that are being extensively used in the biodiversity domain. Various ontologies are developed for the biodiversity domain; however, these ontologies are not capable to define the Indian biodiversity information though India is one of the megadiverse countries. To semantically analyze the Indian biodiversity information, it is crucial to build an ontology that describes all the terms of this domain. Since the curation of the ontology depends on the domain where these are used, there is no ideal methodology defined yet. The aim of this article is to develop an ontology that semantically encodes all the terms of Indian biodiversity information in all its dimensions based on the proposed methodology. The evaluation of the proposed ontology depicts that ontology is well built in the specified domain.

#### **KEYWORDS**

Biodiversity, Evaluation, Methodology, Ontology, Semantic Model

#### 1. INTRODUCTION

Knowledge of the natural world is limited not just by the complexity of natural entities and processes, but also by the complexity of the data that describe them. Enhanced understanding of the natural world depends on our capacity to access and integrate data from the biological, physical, and social sciences; mine those data for new knowledge; and conveys new insights to decision-makers and the general public. The integration of scientific data has proven to be particularly challenging as in (Walls et al., 2014), e.g., due to the inherent complexity and variety of scientific data, and the "hidden" implicit

DOI: 10.4018/IJISMD.315021

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

semantics, which often are known only to domain experts or to the scientists who created the data. Standardization of domain knowledge of biodiversity provides an effective remedy for the problem of non-relevance and increases the participation of experts. The Semantic Web has emerged out as a dynamic field which is not only powerful in terms of the transformation it can bring in the web world, but also the realization of its capabilities can give a new face to the web structure. The Semantic Web adds semantics to the data of current web and thus transforms it into machine understandable and processable form. Whatever is the application, the web puts up to, the access to knowledge is always required and it all depends upon the representation and storage of knowledge structures in a manner that no piece of knowledge in the said domain remains untouched, elucidated or unspecified.

Semantic Technologies support management, reasoning, sharing and gathering of data from heterogeneous sources. This leads to the concept of ontology in which resources are defined in formal way. Ontology is a jargon from meta- physics which means- existence of being as such. In the domain of Artificial Intelligence, Ontology is the formalization of an abstract notion of domain knowledge (Mishra et al., 2021). It is explicit specification of the manifestation of the domain related concept, called conceptualization. Ontology engineering is the modeling of the domain related facts and knowledge. W3C has standardized Web Ontology Language (OWL) for the description of the facts in Resource Description Framework (RDF) which can be formatted with syntaxes like N- triple, TURTLE, RDF/XML, RDF/OWL etc. Ontology engineering takes following things into account: defining concepts in the domain, arranging the concepts in a hierarchy (superclass and subclass), defining attributes and properties, defining individual, defining the axioms, filling the value of attribute/ property.

Relevance of ontology:

- 1. **Consensus Understanding:** Ontology is proven to be a best way of sharing the common understanding of the domain related knowledge among people and among processing agents.
- 2. **Domain Knowledge Reusability:** Ontology enables the reuse of abstract knowledge of the domain. As introducing standards and the concept of upper ontology, interoperability can be achieved. Hence, avoids 're-inventing the wheel'.
- 3. **Inferring Implicit Knowledge:** The power of inference in ontology can be used to unleash the implicit knowledge, through SWRL i.e. Semantic Web Rule Language, from the explicit knowledge provided in knowledgebase through ontology representation.
- 4. Abstract Domain Knowledge Conceptualization: The manifestation of the tangible and intangible (abstract) knowledge can be well conceptualized or formalized with aid of ontology.

Therefore, nowadays ontologies are utilized in various applications to fulfill the need of the user. However, in the context of Indian biodiversity, available ontologies are lacking to provide meaningful information. Hence, it is required to build the semantic structure for Indian biodiversity that describes biodiversity data in all its dimension. This article has the following two contributions:

- The development of a comprehensive knowledge base for Indian Biodiversity (InBiodiv-O) based on the proposed methodology.
- The evaluation of the InBiodiv-O ontology using different evaluation approaches.

The rest of the article is divided as follows: Section 2 shows research work with respect to propose contributions. Section 3 mentions the development and detailed evaluation of Indian Biodiversity Ontology (InBiodiv-O) and the last section concludes the article.

# 2. RELATED RESEARCH WORK

The biodiversity domain is very popular for research and involves the growing interest of the research communities. Well-established efforts have been taken to extract and discover the biodiversity information. In the literature, various authors have developed ontologies for biodiversity data. However, the major problem of ontology development is to select the right methodology that builds correct and complete ontology as per application need. Therefore, ontology development methodologies are a strong foundation for the biodiversity domain as ontology provides semantic information of the imposed queries. In this section, we discuss the available methodologies for ontology development as well as analyse the biodiversity ontologies that contain biodiversity data and infer semantic information of the imposed queries.

# 2.1 Ontology Development Methodologies

Ontology development methodology describes the way to construct the well organize ontology (Jones et al., 1998). There are various methodologies proposed in the literature. However, the most famous used methodologies are TOVE, Enterprise Model Approach, METHONTOLOGY, and KBSI IDEF5. These methodologies provide various steps for the development of the ontology. All the steps of TOVE, Enterprise Model Approach, METHONTOLOGY and KBSI IDEF5 are presented in figure 1. The steps of all these methodologies are related to each other. For example, the step Motivation Scenarios of TOVE is related to step Identify Purpose of Enterprise Model Approach that is also related to step Specification of METHONTOLOGY as well as step Organising and Scoping of KBSI IDEF5.

Basically, all these methodologies have the following mentioned steps:

- The most important step of ontology development is to identify the purpose and fix the scope of the ontology. For this purpose, users use motivation scenarios and write competency questions that provide a clear understanding of the purpose and scope of the ontology.
- Now, the next step is to collect and analyze the information as per the defined scope of the ontology. For this purpose, users can use different sources or repositories and can also conduct



#### Figure 1.



the interview with domain experts in order to collect the information. This step is also called knowledge acquisition.

- After collecting and analyzing all the required information, now need to formalize it. For this purpose, first, specify the classes and their properties. The formal competency questions can be utilized here to know that which concepts or entities will be treated as a class. Now, we start to build up the ontology by encoding the classes and properties. As per the need of the domain, we imposed constraints on the classes and their properties that act as axioms and help us to extract the semantic information from the ontology. It is good practice to use the existing concepts of the upper ontology or any standard ontologies so that data integration can be done smoothly and can avoid the multiple interpretations of the same entity.
- To check the quality of the developed ontology is a very vital step of ontology development methodologies. Ontology evaluation shows completeness (Ontology contains all the required information as per domain need), accuracy (ontology is able to infer the correct answer of all the imposed queries), Correctness (Ontology is well encoded or build). There are various tools available that can be used to check the quality of the ontology. The competency questions are also very helpful for the evaluation of ontology.
- The last step of ontology development methodologies is to document the ontology that can become the base of other activities.

Apart from these methodologies, two methodologies namely Neon (Suárez-Figueroa et al., 2015), and SAMOD (Peroni, 2016) are also available in the literature for the development of the ontology:

- Neon Methodology: This is a scenario-based methodology that focuses on the collaborative aspects of ontology development (Suárez-Figueroa et al., 2012). A set of nine scenarios is the key to this methodology. These nine scenarios are: (1) From the application specification: The ontology is developed from scratch and developers must specify the requirement of the ontology. (2) The reuse and reengineering of non-ontological resources: The developer must decide the reuse terms from the non-ontological resources and must return to the ontological reengineering process. (3) The reuse of ontological resource: The developer must collect the terms from the ontologies and reused them. (4) The reuse and re-engineering of ontological resources: Developers reuse the resources and reorganize. (5) The reuse and fusion of ontological resources: The developer can select more than one ontology of the same domain for the development of the new ontology. (6) Reuse, fusion, and re-engineering of ontological resources: Developers reuse, integrate, and rearrange the extracted ontological resources. (7) Reuse of ontology design patterns: Developers access the repositories for the reusing of ontology design patterns. (8) Restructuring of ontological resources (9) Location of ontological resources: Developers can adopt an ontology to other languages thus creating a multilingual ontology.
- SAMOD Methodology: Simplified Agile Methodology for Ontology Development (SAMOD) is a new agile methodology that permits to development of the ontology in an iterative workflow manner. It focuses on to create well developed and well-documented models beginning from domain description. This methodology has three iterative steps (1) ontology engineers collect all the essential information about a domain with the help of domain experts. The aim of this step is to build the model by using ontology development principles. If everything works fine then release the prototype and proceed further (2) Ontology engineers merge the model of new test cases with the old one that proceeds in the proceeding step. If everything works fine then release the prototype and proceed further (3) Ontology engineers refactors the current models with concentrating on the last added part. If everything works fine then release the prototype and if there is another scenario that needs to be addressed then iterate the process.

The ontology development process heavily depends on the requirement of the application for which it is developed. Despite these available methodologies, practice shows that most of the researchers create their own principle or method for the development of the ontology.

# 2.2 Biodiversity Ontologies

There are various ontologies about the biodiversity information is proposed whereas few of them are accessible and available online and others are available in the form of research articles:

- The authors Senderov et al. (2018) developed an OpenBiodiv ontology that contains structural information about the biodiversity domain. The aim of this ontology is to fill the gaps between biodiversity resources-related ontologies and semantic publishing ontologies. OpenBiodiv ontology focuses on the biological taxonomy and contains classes, relations, properties, and axioms. These classes are aligned with other relevant domain ontologies.
- The authors Albuquerque et al. (2015) developed a biodiversity ontology called OntoBio ontology. This ontology is divided into five sub-ontologies namely Sub-Ontology Collect, Sub-Ontology Material Entity, Sub-Ontology Ecosystem, Sub-Ontology Spatial Location, and Sub-Ontology Environment. These sub-ontologies are connected by the relationship between concepts and axioms.
- The authors Walls et al. (2014) have developed Biological Collections Ontology (BCO) that supports semantic sharing of biodiversity data. BCO includes the data from museum collections, environmental samples, and ecological surveys. BCO ontology offers a clear and well-structured logical definition of terms about the biological data and these terms are organized in a hierarchal manner. The latest version of BCO ontology is released on 31.03.2020 and openly available on various portals like bio portal and OBO foundry, for the public use.
- The author Franco has developed a Biodiversity Ontology (BOF) that contains information about the biodiversity of INPA (Bioportal, n.d.). The latest version of BOF ontology is released on 08.01.2012 and openly available on various portals like bio portal and AberOWL, for public use.

Although these ontologies have enough information about the biodiversity domain yet these ontologies cannot use in the Indian biodiversity domain because they do not offer complete information about the Indian biodiversity. A lot of data about Indian biodiversity is available on the web in different formats (like text, audio, videos, research articles) therefore it is essential to curate or develop the ontology from the available unstructured data so that users can get precise information of the imposed queries.

# 2.3 Ontology Evaluation

Ontology evaluation approaches are broadly classified into four categories (Brank et al., 2005; Tartir et al., 2010; Yu et al., 2007) namely the Technology-based approach, the Quality-based approach, the Data-driven approach, and the Application-based approach:

- The technology-based approach examines the structural characteristics of an ontology. The structural characteristics involve syntax, formal semantic, and consistency of an ontology. This approach focuses on the correctness and usability aspect of an ontology rather than applicability and content quality aspects.
- The quality-based approach focuses on the quality aspect of an ontology. It detects formal redundancies, semantic redundancies, missing definition, inconsistencies in the ontology. The quality measurement depends on the metrics that determine the value or magnitude of the attributes.
- The data-driven approach focuses on the usability aspect of an ontology. This approach checks the ontology has enough entities or not by comparing the ontology entities with other existing ontologies. This approach does not state anything about the clarity and correctness of the ontology. The overall aim of this approach is to check the coverage of the ontology in a particular domain.

• The application-based approach evaluates the ontology by using it in a particular domain. This approach is not considered for the general evaluation because the results of this evaluation are completely dependent on the domain where ontology is used.

Although there are various approaches to evaluate the ontology yet there is no ideal approach that can be considered all aspects of ontology evaluation like clarity, correctness, completeness, consistency, usability, applicability, redundancies, and many more.

# 3. INDIAN BIODIVERSITY ONTOLOGY (INBIODIV-O)

In this section, we propose an ontology development methodology and on the basis of the proposed methodology, we curate an Indian Biodiversity Ontology (InBiodiv-O). We also provide a detailed evaluation of the curated InBiodiv-O ontology via various approaches.

## 3.1 Proposed Methodology

As we discussed earlier, there is no perfect methodology that can be used in all the applications. Every methodology has its own advantage and disadvantage. For example, TOVE and Enterprise Model Approach are stage-based models whereas METHONTOLOGY and KBSI IDEF5 are evolving prototype models. A stage-based model is good if the purpose and requirements for building the ontology are clear at the outset whereas evolving prototype model is applicable when the purpose for building ontology is not clear. We have developed a methodology on the basis of earlier discussed methodologies. Basically, our proposed methodology is a combined approach of the above-mentioned methodologies (the relationship among the existing methodologies is discussed in figure1) that can be used mostly for all types of applications by iterating some steps until all the requirements of the ontology (CUODM) contains six steps namely Scope Determination, Concept Extraction, Concept Organization, Encoding, Evaluation, and Documentation. The Concept Organization step has three sub-steps that execute sequentially. These steps are Analyse, Knowledge Synthesis, and Reuse Concept. Figure 2 shows all the steps of the CUODM methodology.

We use CUODM methodology for the development of the InBiodiv-O ontology. The detailed analysis of each step is mentioned below.

## 3.1.1 Phase I: Scope Determination

The first step of the CUODM methodology is to fix the scope of the ontology. The motivation scenarios and competency questions are used for this purpose. The motivation scenarios provide a

#### Figure 2.

#### Combined Upper Ontology Development Methodology (CUODM)



clear understanding of the boundary of the ontology and make a solid background for the competency questions that explain the requirement of the ontology.

**Motivation Scenario 1:** India secured 12<sup>th</sup> position among the mega diversity countries. However, In the last few decades, India has suffered a rapid decline in biodiversity.

With only 2.35% of the world's land area, India has 16.7% of the world's population and 18% livestock (Anil et al., 2014). Currently available data shows:

- For plant diversity: India is in 10<sup>th</sup> position in the world and 4<sup>th</sup> position in Asia.
- For Mammalian species: India is in 10<sup>th</sup> position in the world.
- For higher vertebrates' endemic species: India is 11<sup>th</sup> position in the world.

India is rich in biodiversity however it has suffered a rapid decline in the biodiversity. The major reason for the decline of biodiversity is listed below:

- The tropical forests of India are disappearing very fast (at a rate of about 0.6% per year). These forests are home to various species, so these species are getting destroyed along with the forests.
- Hunting is one of the major issues for the decline of biodiversity. Wild animals are hunted for their products like tusk, meat, pharmaceutical's purpose, and so on. The Indian government has launched a program to stop the poaching of Indian tigers however this program is failed to save the tiger's life due to the increasing demand for tiger's bone in the pharmaceuticals industry and many other reasons.
- Overexploitation is a serious concern for biodiversity. For example, excessive harvesting of marine organisms (like fish, turtles, sea cows and etc) has caused in loss of these animals.
- Animals and plants collection for zoo and research.
- Pollution that alters the natural habitat.
- Humans are continuously destroying natural habitats for their greedy needs. As a result, the species must move to other places or must adapt to the changes which create hurdles for species life.
   Motivation Scenario 2: How much percentage did India record of plant and animal species as compared to the world?

India is a rich country for flora and faunal species. India has about 45,944 plant species which accounts for 10.75% of the world's plants. The flowering plant species are 18000 however 36% of flowering plant species are endemic. India is about 89,317 animal species whereas about 4952 and 84365 are vertebrates and invertebrates respectively. Figures 3 (a) and (b) show the percentage of India with respect to the world for plant and animal species respectively.

On the basis of motivation scenarios, we have framed a list of 25 competency questions. To frame the competency questions, we have conducted an interview with the expert of the biodiversity domain as well as we contact the end-users. Some Competency questions are listed below and explain that what information should be encoded inside the ontology:

CQ1: How many biodiversity hotspots in India?

CQ2: In Which state of India has the highest biodiversity?

**CQ3:** Which one is the largest biodiversity hotspot in India?

CQ4: How many species are disappearing?

**CQ5:** In which area of India, biodiversity hotspots located?

CQ6: What steps have been taken so far to preserve Indian biodiversity?

CQ7: Which factor is responsible for the mass extinction of species in India?

International Journal of Information System Modeling and Design Volume 13 • Issue 7

#### Figure 3.





CQ8: What is the reason behind the vast biodiversity in Indian?

CQ9: How to find common features of the species?

CQ10: Which methods are used for the assessment of biodiversity in bacteria?

CQ11: What rank did India record in the world and Asia for biodiversity?

CQ12: How much % of the world's biodiversity is found in India?

CQ13: How many projects are planned for the management of endangered species in India?

CQ14: What is Ramsar site and how many Ramsar site is in India?

CQ15: What different types of Plants available in India according to the Plant family?

## 3.1.2 Phase II: Concept Extraction

Concept extraction means collecting the concepts from different places as per the specified domain. We have extracted the concept of the InBiodiv-O ontology from the different databases, ontologies, news articles, government websites as well as conducting the interview with the experts of the domain. Some selected sources are listed below:

- Global Biodiversity Information Facility (GBIF) is a well-known portal for biodiversityrelated information. This portal has 2228 datasets about the Indian biodiversity that clearly specify the name and number of occurrences of biodiversity like Animalia has 17,554,258 occurrences.
- Ontologies like biological collection ontology (BCO), Environment Ontology (ENVO), and Population & Community Ontology (PCO), Biodiversity Ontology (BOF) are used for the concept extraction. Only those concepts are extracted that are useful in the context of Indian biodiversity.
- Articles (research articles and articles that available on websites like blogs, Wikipedia, and so on) are also used for the collection of the concepts.
- The government websites like india.gov.in (national portal of India) are used to know the launched policies of the Indian government about biodiversity.

These sources generate a list of relevant concepts (that can be classified as classes, properties, and instances) and also assist to the know the irrelevant concepts. All the extracted concepts are stored in the excel sheet for further analysis. Initially, we found 1056 concepts after examining the synonyms of the concepts. Some of the most significant concepts are listed below that extracted from the biodiversity-related ontologies and datasets.

### 3.1.2.1 Extracted Concepts From Ontologies

ClimaticCondition, Cloudy, Rainy, Sunny, Ecosystem, MacroEcosystem, Biome, TypeBiome, Location MesoEcosystem, MicroEcosystem, Environment, MacroEnvironment, MicroEnvironment, Faun, Air, Sex, TypeMacroEnvironment, TypeMicroEnvironment, AffluentEnvironment, AmbientedeFuro, Flora, Soil, BayEnvironment, BeachEnvironment, CanopyEnvironment, CaveEnvironment, CreeksEnvironment, Person DeforestedAreaEnvironment, Litolitic, Rocky, Sandy, BrakishWater, SalineWater, SweetWater, Animal

#### 3.1.2.2 Extracted Concepts From Databases

National Park, Sanctuaries, Biosphere Reserves, Ashtamudi Wetland, Kolleru Lake, Loktak Lake, Nalsarovar Bird Sanctuary, Rudrasagar Lake, Ropar, Tsomoriri, Surinsar-Mansar Lakes, Vembanad-Kol Wetland, Wular Lake, Gulf of Katchh, Vasai-Manori, Shrivardhan, Vasai-Manori, Kundalika-Ravdana, Devgarh-Vijay Dur, Achra-Ratnagiri, Dakshin Kannada/Honnavar, Coondapur, Vembanad, North Andamans, East Godavari, Mangrove Areas, Keoladeo National Park, Manas National Park, Kaziranga national Park, Nanda Devi Natinal

#### 3.1.3 Phase III: Concept Organization

The aim of this phase is to provide a hierarchical model (Parent-Child relationship) for the origination of the extracted concepts. The preceding step provides all the essential concepts that are required to encode the domain knowledge. Now, we follow the following mentioned steps to organize the concepts:

- Analysis: The extracted concepts are analyzed on the basis of their characteristics and breaking them into element levels. All the concepts are filtered into three groups namely class, property, and instance on the basis of their properties. For example, the concept Animal must be a class rather than instance or property because all the animal has some common properties that grouped them into a class. We also identify the connecting points that connect the concepts with each other.
- Knowledge Synthesis: All the element-level concepts are organized in a hierarchical manner by defining the relationship between the concepts (Jain & Patel, 2019). For example, class *Biosphere Reserves* relates with class *Protected area categories* via is\_a relationship. We also impose the constraints on the concepts that restrict the interpretation of that concept like class *Biosphere Reserves* should be disjoint with class *National\_Parks*. All the constraints and axioms are imposed on the basis of competency questions and motivation scenarios that help in the building of the correct ontology as per need.
- **Reuse concept:** To reuse the concepts assist to avoid multiple interpretations of the concepts. Few ontologies about the biodiversity information are available publicly so it is required to reuse or import the required concepts in the InBiodiv-O ontology. We have imported the concepts from BCO, BOF, and ENVO ontologies. These concepts are imported in InBiodiv-O ontology via IRI (internationalized resource identifier). InBiodiv-O ontology has 628 classes whereas 132 classes are imported (125 classes are imported from BOF, 5 classes are imported from ENVO and 2 classes are imported from BCO). We use prefix bioC: for BCO ontology; prefix bioO: for BOF ontology; prefix bioE: for ENVO ontology; prefix bioIn: for the InBiodiv-O ontology.

The prefix declarations are as follows:

- bioC: http://rs.tdwg.org/dwc/dwctype/
- bioO: http://www.owl-ontologies.com/BiodiversityOntologyFull.owl#
- bioE: http://purl.obolibrary.org/obo/
- bioIn: http://www.semanticweb.org/mca/ontologies/2018/8/untitled-ontology-47#

## 3.1.4 Phase IV: Encoding

Encoding of the ontology means to represent the concepts that are identified in the preceding phase. The InBiodiv-O ontology has schema-level information about the Indian biodiversity domain that is encoded in web ontology language (OWL) by the protégé 5.5.0 tool. Protégé is an open-source tool and allows the export of the ontology in various data formats (like RDF/XML, OWL/XML, OBO, and so on), and these formats are compatible with each other. By default, the protégé tool supports Pellet and Hermit reasoners in order to check the consistency of the ontology. The ontology knowledge base consists of schema and instance-level information. We have encoded schema-level information inside InBiodiv-O ontology by defining classes and their properties. The development of ontology is not the same as an object-oriented programming language. In ontology, the design decision is made on the basis of the structural properties of a class whereas in the object-oriented programming language, the design decision made on the basis of the operational properties of a class. Therefore, in the same domain, the class structure and relations (that hold between the classes) might not be the same in ontology and object-oriented programming.

**Defining Classes and Properties:** The classes are the main entity of the ontology and it defines the reality of a domain. Thing is a main class of ontology and all the classes become the subclasses of the Thing class. All the classes are organized in a hierarchical manner. Three approaches are defined to build the ontology hierarchy. These approaches are (a) Top-down approach: define the most general classes and later on break them into more specific classes (b) Bottom-up approach: define most specific classes and later on group them into more general classes. (c) Combination approach: define more silent classes and later on make these classes specific and general as per need. We follow the combination approach to build the class hierarchy of InBiodiv-O ontology. Classes can have restrictions on themselves, for example, a class can be a SubClassof, Disjoint with, disjoint union, and many others.

Classes alone do not provide enough information needed to answer the competency questions. Therefore, we must define the internal structure of the classes by defining properties. We already made a list of classes, now the remaining concepts become properties as per need. In ontology modeling, there are two types of properties: data properties (it is used to relate the instances of the classes with literal data value), object properties (it is used to relate the pair of instances of the classes). All the properties have domain and range as well as some additional properties like symmetric, transitive, reflexive, and many more. These properties provide additional semantics by which the machine infers hidden knowledge like a human.

The latest version of Covid19-IBO ontology is available on the Bio portal, link: https://bioportal. bioontology.org/ontologies/INBIODIV/?p=summary.

## 3.1.5 Phase V: Evaluation

The detailed evaluation of InBiodiv-O ontology is mentioned in section 3.2.

## 3.1.6 Phase VI: Documentation

It is a good particle to provide a human-readable description of the ontology. This is possible by taking into account the ontological axioms as well as annotations and order these with respect to the functionality of W3C recommendations web pages (HTML pages) by means of embedded links for ease of browsing and navigation.

# 3.2 Evaluation of InBiodiv-O Ontology

We use a combined approach for the evaluation of the InBiodiv-O ontology proposed by authors Mehla and Jain (2019). They have divided the evaluation approaches into two groups (a) Verification:

building an ontology correctly (b) Validation: building the correct ontology. The detailed analysis of the evaluation of InBiodiv-O ontology is listed below:

- 1. **Verification of InBiodiv-O Ontology:** We use quantitative and qualitative approaches to verify the InBiodiv-O ontology. The quantitative approach checks the quantity of the ontology whereas the qualitative approach measures the quality of the ontology. The metric-based method is widely used to check the quantity of the ontology whereas the criteria-based method is famous to reveal the quality of the ontology:
  - a. **Metric-Based Method:** We use a metric-based method to evaluate the InBiodiv-O ontology quantitatively. OntoMetric tool measures the quantity of the ontology on the basis of various metrics (Lozano-Tello & Gómez-Pérez, 2004). It sets all the features of an ontology under five metrics:
    - i. **Base Metrics:** These metrics show the quantity of the ontology element and they consist of simple metrics like a number of classes, axioms, data property and object property, instances, etc.
    - ii. **Schema Metrics:** These metrics are used to describe the design of the ontology by indicating attribute richness, relationship richness, and inheritance richness, etc.
    - iii. **Knowledge Metrics:** It measures the amount of data that are encoded inside the ontology. Basically, it explains the effectiveness of the ontology design by counting instances of an ontology.
    - iv. Class Metrics: It measures the classes and relationships of the ontology.
    - v. **Graph or Structure Metrics:** It explains the structure of the ontology by examining the cardinality, depth, breadth, total number of paths, etc.

We have run the OntoMetric tool on InBiodiv-O ontology along with the existing biodiversity ontologies namely BCO, ENVO, and BOF. Table 1 depicts the value of the metrics of these ontologies. The metrics value shows the richness of the ontologies.

Table 1 depicts that InBiodiv-O ontology has a good level of information. Although the available ontologies namely BCO, ENVO, and BOF contain enough information of the biodiversity (in some cases these ontologies have more metric values as compared to InBiodiv-O ontology) yet these ontologies cannot use in the context of the Indian biodiversity domain. The proposed InBiodiv-O ontology describes the complete information about the Indian biodiversity domain and imported some concepts from the available ontologies as needed. Thus InBiodiv-O ontology is itself unique in scope.

- b. Criteria-based Method: We use a criteria-based method to evaluate the ontology qualitatively. Ontology Pitfalls Scanner (OOPs) is a criteria-based tool and available on the web (Poveda-Villalón et al., 2014). This tool shows the errors or anomalies of the ontology under three categories namely minor pitfalls (these are not serious concerns), important pitfalls (these are a moderate pitfall and it is good practice to remove these pitfalls from the ontologies), and critical pitfalls (these pitfalls must be removed from the ontologies). OOPs, the tool shows 41 types of pitfalls (from P01 to P41) as well as their descriptions. Table 2 depicts the pitfalls that available in the BCO, BOF, and InBiodiv-O ontologies. The description of the obtained pitfalls is mentioned below:
  - i. **Minor pitfalls P02:** Shows that creating synonyms as classes, P04: shows that ontology has unconnected elements, P07: shows that concepts are merged in the same class, P08: shows that annotation properties are missing, P13: shows that the property inverse is not defined explicitly, P20: shows the properties annotations are not used properly, P22: shows that ontology has different naming conventions.
  - ii. **Important Pitfalls P11:** Shows that domain and range properties are missing, P24: shows that ontology uses recursive definitions, P30: shows that equivalence classes are

#### Table 1.

Metrics value of the available ontologies with respect to InBiodiv-O

Metrics	Ontologies			
	ВСО	ENVO	BOF	InBiodiv-O
Axioms	8661	49216	1993	1418
Logical axioms count	1024	12270	1378	739
Class count	351	6,566	186	628
Object property count	48	135	57	18
Data property count	159	1	37	27
Attribute Richness	0.452	1.52E-4	0.198	0.042
Inheritance Richness	1.344	1.657	0.924	0.979
Relation Richness	0.195	0.107	0.628	0.1176
Absolute cardinality root leaf sibling	6 175 286	456 3870 6565	14 149 186	21 539 628
Depth: Absolute Average Maximal	2720 8.16 35	234 12.280 36	603 3.24 7	2282 3.62 8
Breadth: Absolute Average Maximal	333 2.25 38	19081 2.080 456	186 4.89 22	630 7.0 51

not defined explicitly, P34: shows that ontology has untyped class, P41: shows that no license declares.

iii. Critical Pitfalls - P06: Shows that class hierarchy has cycles, P19: shows that ontology has multiple domains and ranges in properties, P28: shows that ontology defined wrong symmetric relationships, P29: shows that ontology defined wrong transitive relationships, P31: shows that ontology defined wrong equivalence relationships.

The number (like 1, 3, 4,..) in table 2 shows the total number of cases corresponding to the mentioned pitfalls for example BCO ontology has 9 cases of pitfall PO4. The sign depicts that there are no pitfall cases available.

The BCO and InBiodiv-O ontologies have critical pitfalls that need to remove from the ontologies. We have removed all critical pitfalls P28, and P29 from the ontology by defining the correct symmetric and transitive relationship between the concepts as well as we have also removed all the minor and important pitfalls from the InBiodiv-O ontology. However, existing ontologies are still having these pitfalls.

c. Validation of InBiodiv-O Ontology: We use competency questions and validators to validate the InBiodiv-O ontology. The competency questions check that ontology is well built as per the specified domain whereas validators ensure that ontology is well-coded means ontology's syntax is written as per W3C standard.

Ontologies	Pitfalls (Minor)						
	P02	P04	P07	P08	P13	P20	P22
BCO		9		11	7	1	1
BOF		8		289	58		1
InBiodiv-O	7	18	6	671	11		1
	Pitfalls (Important)						
	P11		P24	P30	P34	P41	
BCO	183			3			
BOF	8		33	4	8	1	
InBiodiv-O	31			6		1	
Pit		ritfalls (Critical)					
	P06		P19	P28	P29	P31	
BCO	4		1		1	4	
BOF							
InBiodiv-O				3	1		

#### Table 2. Obtained Pitfalls in Ontologies

# 3.3 Competency Questions

Competency questions describe the requirement of the ontology and are framed on the basis of motivation scenarios. A set of 45 questions that cover the Indian biodiversity domain are provided in natural language and are translated into SPARQL queries. SPARQL builds a graph pattern to extract the knowledge from the ontology. SPARQL has the power to integrate the data from diverse data sources. The SPARQL query starts with a list of namespaces found in IRI. All the query is run on the InBiodiv-O ontology via protégé tool and all the missing information is added into ontology on the basis of failed queries. These queries determine that ontology is complete in itself or not. Table 3 shows few queries that translated into SPARQL and run on InBiodiv-O ontology. Queries 1 and 2 are the subject of new concepts whereas queries 3 and 4 are related to the imported concepts taken from the BOF and ENVO ontologies respectively. All the imported concepts are extracted from the InBiodiv-O ontology via SPARQL queries that show that reused concepts are encoded inside ontology successfully. The competency questions are successfully executed on the developed ontology that reflects that ontology is well built within the proposed scope.

# 3.4 RDF Triple-Checker Validator

We have used the RDF Triple-Checker validator to check the deferenceability of the namespaces in the RDF document. To retrieve a representation of a resource identified by a URI (unique resource identifier) is term as dereferencing that URI. The process of deferencing a URI is crucial on the web. RDF Triple-Checker validator checks the typos error in the RDF document, as well as this tool, investigates the dereferenceability of the namespaces associated with the classes and properties in a document. It examines that the URI of the classes and properties (object and data properties) are defined in the namespace and the namespace corresponds with a well-known prefix. Figure 4 shows that all the 2095 triples are extracted successfully and all the classes and properties have a well-defined prefix.

Volume 13 · Issue 7

#### Table 3. SPARQL queries and their results

Queries Related to new Concepts				
1. List the name of Botanical Gardens that a	vailable in India			
prefix bioIn: <http: <="" td="" www.semanticweb.org=""><td>/mca/ontologies/2018/8/untitled-ontology-47#&gt;</td></http:>	/mca/ontologies/2018/8/untitled-ontology-47#>			
SELECT ?BG	5 6			
WHERE {?BG rdfs:subClassOf bioIn:Botar	nical Gardens }			
"Pilikula Arboratum (Pilikula Botanical Garden)"	IFGTB Botanical Garden			
The_Garça_Branca_Ayurvedic_Botanical_Garden	Lloyd's_Botanical_Garden			
Regional_Museum_of_Natural_History	'Garden_of_Medicinal_Plants_North_Bengal_University'			
'Government_Botanical_Gardens,_Ootacamund'	Odisha_State_Botanical_Garden_Nandankannan			
Jhansi Botanical Garden	Acharva Jacadish Chandra Bose Indian Botanic Garden			
Lalbagh	Saharanpur_Botanical_Garden			
RBBotanical_Garden_and_Amusement_Park	'Botanical_Garden,_Near_Sarangpur'			
Jammu_and_Kashmir_Medicinal_Plants_Introduction_Centre	Curzon_Park Retained Carden of Forest Research Institute			
Agri Horticultural Society of India	Auroville Botanical Garden			
Vellayani_Agricultural_College	Assam_State_Zoo-cum-Botanical_Garden			
Empress_Garden	Malampuzha_Garden			
University_of_Mysore_Botanical_Garden				
2. What is Ramsar site and how many Rams	ar site is in India			
prefix bioIn: <http: <="" td="" www.semanticweb.org=""><td>/mca/ontologies/2018/8/untitled-ontology-47#&gt;</td></http:>	/mca/ontologies/2018/8/untitled-ontology-47#>			
SELECT distinct 2n 20 2RS	8			
WHERE {				
thioIn:RamsarSite 2n 20				
2 n ndfitring ovul A materian Dramatte				
During OwnAnnotationProperty.				
} Union				
{?RS rdfs:subClassOf bioIn:RamsarSite	.} }			
o "A Ramsar site is a wetland site designated to be of international importance under the R	Ramsar Convention"			
RS				
ThrissurKole_Wetlands Hokersar Wetland				
DeeporBeel				
Wular_Lake Loktak_Lake				
Surinsar-Mansar_Lakes Chilika Lake				
Kanjii_Wetland				
Upper_Ganga_River Chandra_Taal				
Bhitarkanika_Mangroves Benuka Wetland				
Rudrasagar_Lake				
Tsomorin Point_Calimere_Wildlife_and_BirdSanctuary				
Vembanad-Kol_Wetland				
Sambhar_Lake				
Pong Dam Lake				
Kolleru_Lake				
Ashtamudi_Wetland Nalsarovar_Bird_Sanctuary				
Bhoj_Wetland Keoladeo National Park				
Sasthamkotta_Lake				
Queries related to imported entities				
3. How many types of climates are in India?				
Prefix bioO: <http: td="" www.owl-ontologies.co<=""><td>om/BiodiversityOntologyFull.owl#&gt;</td></http:>	om/BiodiversityOntologyFull.owl#>			
SELECT ?CD				
WHERE {				
?CD rdfs:subClassOf bioO:Climatic_Condit	tion .}			
Warm				
Moderate				
Rainy				
Cloudy				
Sunny				
Cold				
Composite				
4. Plant name according to the Plant family				
PREFIX : <http: obo#="" purl.obolibrary.org=""></http:>				
SELECT ?class				
WHERE { ?class rdfs:subClassOf+ :disease}				
	,			
watermelon_plant				
cucumis_species				
squash_family				
meion_plant				

#### Figure 4. Result of Covid19-IBO ontology over RDF Triple-Checker Validator

		RDF Tripl	e-Checker	
URI/I		os://data.bioontology.org/ontologies/INBIODIV/submissions/1/downl	ad?apikev=c1855b5e-ec88-4604	4-bbf5-54d1b8591f7f
		Output: @ H		
		Check		
			(Again	
		Loaded 2	095 triples	
		Namespac	es	
Namesp	bace	Looks Legit?		
http://w	ww.w3.org/	1999/02/22-rdf-syntax-ns# Matched common nam	espace. Prefix rdf	
http://w	ww.w3.org/	2000/01/rdf-schema# Matched common nan 2002/07/ow/# Matched common nan	lespace. Prefix rdfs	
ALC PLAY AV		Terms		
0		· · · · · · · · · · · · · · · · · · ·		
Cour	it Type	Namespac	e lem	LOOKS Legit?
749	property	http://www.w3.org/1999/02/22-rdf-syntax-ns#	type	OK - Looks good.
308	property	http://www.w3.org/1999/02/22-rdf-syntax-ns#	first	OK - Looks good.
308	property	http://www.w3.org/1999/02/22-rdf-syntax-ns#	rest	OK - Looks good.
15	property	http://www.w3.org/2000/01/rdf-schema#	domain	OK - Looks good.
5	property	http://www.w3.org/2000/01/rdf-schema#	t comment	OK - Looks good.
11	property	http://www.w3.org/2000/01/rdf-schema#	range	OK - Looks good.
9	property	http://www.w3.org/2000/01/rdf-schema#	subPropertyOf	OK - Looks good.
615	property	http://www.w3.org/2000/01/rdf-schema#	subClassOf	OK - Looks good.
1	class	http://www.w3.org/2002/07/owl#	Ontology	OK - Looks good.
17	class	http://www.w3.org/2002/07/owl#	DbjectProperty	OK - Looks good.
5	class	http://www.w3.org/2002/07/owl#	FunctionalProperty	OK - Looks good.
4	class	http://www.w3.org/2002/07/owl#	InverseFunctionalProperty	OK - Looks good.
4	class	http://www.w3.org/2002/07/owl#	SymmetricProperty	OK - Looks good.
1	property	http://www.w3.org/2002/07/owl#	inverseOf	OK - Looks good.
2	class	http://www.w3.org/2002/07/owl#	AsymmetricProperty	OK - Looks good.
2	class	http://www.w3.org/2002/07/owl#	IrreflexiveProperty	OK - Looks good.
1	class	http://www.w3.org/2002/07/owl#	TransitiveProperty	OK - Looks good.
1	class	http://www.w3.org/2002/07/owl#	ReflexiveProperty	OK - Looks good.
26	class	http://www.w3.org/2002/07/owl#	DatatypeProperty	OK - Looks good.
628	class	http://www.w3.org/2002/07/owl#	Class	OK - Looks good.
6	class	http://www.w3.org/2002/07/owl#	Restriction	OK - Looks good.
6	property	http://www.w3.org/2002/07/owl#	onProperty	OK - Looks good.
6	property	http://www.w3.org/2002/07/owl#	someValuesFrom	OK - Looks good.
8	property	http://www.w3.org/2002/07/owl#	disjointWith	OK - Looks good.
7	property	http://www.w3.org/2002/07/owl#	equivalentClass	OK - Looks good.
3	class	http://www.w3.org/2002/07/ow#	MamedIndividual	OK - Looks good.
46	class	http://www.w3.org/2002/07/owl#	# AllDisjointClasses	OK - Looks good.
46	property	http://www.w3.org/2002/07/owl#	# members	OK - Looks good.

## 4. CONCLUSION

A lot of data about biodiversity is available on the web in unstructured manner. However, this data is trivial until user is not able to infer the unseen patterns, and derive meaningful information. To create intelligence in data, it is necessary to build an ontology that allows the machine-readable representation of data. Although there are various ontologies are proposed in the literature for the

biodiversity domain yet these ontologies are not able to provide the Indian biodiversity information. The proposed InBiodiv-O ontology has rich knowledge about the Indian biodiversity domain and it is defined as per W3C standard. The comprehensive evaluation of the proposed ontology shows that ontology is well built as per the specified domain.

# **CONFLICT OF INTEREST**

The authors declare that they have no conflict of interest.

# FUNDING INFORMATION

No funding (Nil)

#### REFERENCES

Albuquerque, A. C., dos Santos, J. L. C., & de Castro, A. N. (2015, January). OntoBio: A biodiversity domain ontology for Amazonian biological collected objects. In 2015 48th Hawaii International Conference on System Sciences (pp. 3770-3779). IEEE. doi:10.1109/HICSS.2015.453

Anil, M. N. V., Kumari, K., & Wate, S. R. (2014). Loss of biodiversity and conservation strategies: An outlook of Indian scenario. *Asian Journal of Conservation Biology*, *3*(2), 105–114.

Bioportal. (n.d.). https://bioportal.bioontology.org/ontologies/BOF/?p=summary

Brank, J., Grobelnik, M., & Mladeni, D. (2005). A survey of ontology evaluation techniques. Academic Press.

Jain, S., & Patel, A. (2019, October). Smart Ontology-Based Event Identification. In 2019 IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC) (pp. 135-142). IEEE. doi:10.1109/MCSoC.2019.00027

Jones, D., Bench-Capon, T., & Visser, P. (1998). Methodologies for ontology development. Academic Press.

Lozano-Tello, A., & Gómez-Pérez, A. (2004). Ontometric: A method to choose the appropriate ontology. *Journal of Database Management*, 15(2), 1–18. doi:10.4018/jdm.2004040101

Mehla, S., & Jain, S. (2019). Development and evaluation of knowledge treasure for emergency situation awareness. *International Journal of Computers and Applications*, 1–11.

Mishra, A. K., Patel, A., & Jain, S. (2021, January). Impact of Covid-19 Outbreak on Performance of Indian Banking Sector. *International Semantic Intelligence Conference*.

Peroni, S. (2016). SAMOD: An agile methodology for the development of ontologies. Academic Press.

Poveda-Villalón, M., Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2014). Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems*, *10*(2), 7–34. doi:10.4018/ijswis.2014040102

Senderov, V., Simov, K., Franz, N., Stoev, P., Catapano, T., Agosti, D., Sautter, G., Morris, R. A., & Penev, L. (2018). OpenBiodiv-O: Ontology of the OpenBiodiv knowledge management system. *Journal of Biomedical Semantics*, 9(1), 1–15. doi:10.1186/s13326-017-0174-5 PMID:29347997

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2012). The NeOn methodology for ontology engineering. In *Ontology engineering in a networked world* (pp. 9–34). Springer. doi:10.1007/978-3-642-24794-1\_2

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernandez-Lopez, M. (2015). The NeOn Methodology framework: A scenario-based methodology for ontology development. *Applied Ontology*, *10*(2), 107–145. doi:10.3233/ AO-150145

Tartir, S., Arpinar, I. B., & Sheth, A. P. (2010). Ontological evaluation and validation. In *Theory and applications* of ontology: Computer applications (pp. 115–130). Springer Netherlands. doi:10.1007/978-90-481-8847-5\_5

Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P. L., Davies, N., Endresen, D., Gandolfo, M. A., Hanner, R., Janning, A., Krishtalka, L., Matsunaga, A., Midford, P., Morrison, N., Tuama, É. Ó., Schildhauer, M., & Wooley, J. et al. (2014). Semantics in support of biodiversity knowledge discovery: An introduction to the biological collections ontology and related ontologies. *PLoS One*, *9*(3), e89606. doi:10.1371/journal.pone.0089606 PMID:24595056

Yu, J., Thom, J. A., & Tam, A. (2007, November). Ontology evaluation using wikipedia categories for browsing. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 223-232). ACM. doi:10.1145/1321440.1321474 Archana Patel is currently a faculty of the Department of Software Engineering, School of Computing and Information Technology, Binh Duong Province, Vietnam. She has completed Postdoc from the Freie Universität Berlin, Berlin, Germany. Dr. Patel is MyNEP Ambassador, National Education Policy 2020, India. She has filed a patent titled "Method And System For Creating Ontology of Knowledge Units In A Computing Environment" in November 2019. She has received Doctor of Philosophy (Ph.D.) in Computer Applications and Post Graduate (PG) degree both from the National Institute of Technology (NIT) Kurukshetra, India in 2020 and 2016 respectively. She has qualified GATE and UGC-NET/JRF exams in year 2017. Dr. Patel has also contributed in research project funded by Defence Research and Development Organization (DRDO) for the period of two year. Her contribution in teaching is also remarkable. Dr. Patel is an author or co-author of more than 22 publications in numerous refereed journals and conference proceedings. She is two times best paper award winner in the international conferences. Dr. Patel has received various awards for presentation of research work at various international conferences, teaching, and research institutions. Her primary research interests are in Ontological Engineering, Semantic Web, Big Data, Expert System and Knowledge Warehouse.

Sarika Jain graduated from Jawaharlal Nehru University (India) in 2001. Her doctorate, awarded in 2011, is in the field of knowledge representation in Artificial Intelligence. She has served in the field of education for over 19 years and is currently in service at the National Institute of Technology Kurukshetra (Institute of National Importance), India. Dr. Jain has authored or co-authored over 100 publications including books. Her current research interests are knowledge management and analytics, the semantic web, ontological engineering, and intelligent systems. Dr. Jain serves as a reviewer for journals published by IEEE, Elsevier, and Springer. She has completed two funded research projects: one funded by CSIR TEQIP- III and worth Rs 2.58 lakhs, and the other funded by the Defense Research and Development Organization, India, and worth Rs 40 lakhs. She has continuously supervised DAAD interns from different universities of Germany and many interns from India every summer. She is a senior member of the IEEE, a member of ACM, and a Life Member of CSI. For more and most updated information, see https:// sites.google.com/view/nitkkrsarikajain/.

Narayan C. Debnath is currently the Founding Dean of the School of Computing and Information Technology and Head of the Department of Software Engineering at Eastern International University, Vietnam. Dr. Debnath has been the Director of the International Society for Computers and their Applications (ISCA) since 2014, and a member of the ISCA Board of Directors since 2001. Formerly, Dr. Debnath served as a Full Professor of Computer Science at Winona State University, Minnesota, USA for 28 years (1989-2017). He was elected as the Chairperson of the Computer Science Department at Winona State University for 3 consecutive terms and assumed the role of the Chairperson of the Computer Science Department at Winona State University for 7 years (2010-2017). Dr. Debnath earned a Doctor of Science (D.Sc.) degree in Computer Science and also a Doctor of Philosophy (Ph.D.) degree in Physics. In the past, he served as the elected President for 2 separate terms, Vice President, and Conference Coordinator of the International Society for Computers and their Applications. Dr. Debnath received numerous honors and awards while serving as a Professor of Computer Science, including the Best Paper Award in the field of Networking at the 2008 IEEE International Symposium on Computers and Communications. Professor Debnath has made significant contributions in teaching, research, and services across the academic and professional communities. He has made original research contributions in Software Engineering Models, Metrics and Tools, Software Testing, Software Management, and Information Science, Technology and Engineering, Dr. Debnath is an author or co-author of over 425 publications in numerous refereed journals and conference proceedings in Computer Science, Information Science, Information Technology, System Sciences, Mathematics, and Electrical Engineering. Professor Debnath has made numerous teaching and research presentations at various national and international conferences, industries, and institutions in Africa, Asia, Australia, Europe, North America, and South America. He has offered courses and workshops on Software Engineering and Software Testing at universities in Asia, Africa, Middle East, and South America. Dr. Debnath has been a visiting professor at universities in Argentina, China, India, Sudan, and Taiwan and has maintained an active research and professional collaboration with many universities, scholars and practitioners across the globe. Dr. Debnath has been an active member of the ACM, IEEE Computer Society, Arab Computer Society, and a senior member of the ISCA.