

# Ontology Based Ocean Knowledge Representation for Semantic Information Retrieval

Anitha Velu\* and Menakadevi Thangavelu

Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Hosur, 635109, Tamil Nadu, India

\*Corresponding Author: Anitha Velu. Email: aniveluece@gmail.com

Received: 09 May 2021; Accepted: 18 July 2021

**Abstract:** The drastic growth of coastal observation sensors results in copious data that provide weather information. The intricacies in sensor-generated big data are heterogeneity and interpretation, driving high-end Information Retrieval (IR) systems. The Semantic Web (SW) can solve this issue by integrating data into a single platform for information exchange and knowledge retrieval. This paper focuses on exploiting the SW base system to provide interoperability through ontologies by combining the data concepts with ontology classes. This paper presents a 4-phase weather data model: data processing, ontology creation, SW processing, and query engine. The developed Oceanographic Weather Ontology helps to enhance data analysis, discovery, IR, and decision making. In addition to that, it also evaluates the developed ontology with other *state-of-the-art* ontologies. The proposed ontology's quality has improved by 39.28% in terms of completeness, and structural complexity has decreased by 45.29%, 11% and 37.7% in Precision and Accuracy. Indian Meteorological Satellite INSAT-3D's ocean data is a typical example of testing the proposed model. The experimental result shows the effectiveness of the proposed data model and its advantages in machine understanding and IR.

**Keywords:** Heterogeneous climatic data; information retrieval; semantic web; sensor observation services; knowledge representation; ontology

## 1 Introduction

Several marine disasters happen every year, mainly due to weather phenomena; hence, weather prediction and analysis among ocean areas are essential. Ocean observation sensors monitor the Coastal areas and record the values of weather parameters. The provided information is voluminous and heterogeneous, which causes precise Information Retrieval (IR), a challenging task [1]. A standardized framework should be created using semantic technologies and Linked Open Data (LOD) to represent and retrieve the knowledge. SW is an upgraded version of the current web, which publishes the data in a machine-readable, understandable, and processable format through defining ontologies [2]. Researchers propose a Semantic Web of Things (SWoT) framework that supports the collaborative discovery of sensors and actuators for semantic matchmaking [3] and



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

a data fusion framework for improving scalability. Semantic annotation of sensor data consists of metadata information in triples, making the data bulkier. Hence, semantic data management helps reduce the weight of such data by presenting latent context in the form of a triple [4]. SW collaborated with machine learning helps in Decision Support Systems (DSS) such as fake news detection in social media and applications like Service-Oriented Computing (SOC) and smart home systems. Key Information Extraction (KIE) helps to find the exact preferences for the user during semantic data analysis by considering the strength and influence of a term through combinational optimization techniques.

Ontologies help to describe knowledge of a domain and make machines understand the user requirement. Domain Knowledge Ontology (DO) represents a semantic relationship among the data constructed manually, involving a significant workforce and manual invention by combining expert knowledge and domain knowledge [5]. In general, ontology learning focuses on concepts and the relationship between ideas. Sensor-based knowledge acquisition has achieved a drastic development in recent years, where a tremendous amount of sensor data is processed using some computer algorithms. Ontology is a World Wide Web Consortium (W3C) approved technology that provides an advantage of standard vocabulary with added robustness. Ontology Inference Layer (OIL), The Defense Advanced Research Projects Agent Markup Language (DAML)+ OIL, Web Ontology Language (OWL), Resource Description Framework (RDF), and RDF-Schema (RDF-S) are some of the computer languages used to construct ontology. Among these languages, OWL is widely preferred [6].

Similarly, Web Ontology (WebOnto), NeOn, Semantic Web Ontology Overview and Perusal (SWOOP), Web Ontological Design Environment (ODE), OilEd, Protégé, and Onto Editor are some of the platforms available to create an ontology. Still, Protégé is accepted to be the best with various functionalities [7]. Suppose the information published is in machine-readable format with semantic descriptions, then it is easy to search and access the data dynamically by writing queries using Simple Protocol and RDF Query Language (SPARQL). These include service requests and responses at run time. Ontologies support a wide range of applications, namely climate, building environment, marine, e-learning, health care, video processing, social network analysis, web service detection, etc.

The paper is structured as follows: Section 2 presents the background of the research area, Section 3 deals with the literature review, and Section 4 deals with the proposed four-phase weather data model for ocean IR by building Offshore Wind Ontology (OWO) and expressions for quality measures. The results and discussions, along with quality analysis, are incorporated in Section 5. Finally, the paper concludes the work in Section 6 along with the future scope of research.

## 2 Research Background

Information Technology (IT) has provided extraordinary progress over the years, but many have not yet applied it to environmental management. The size, intricacy, and variety create a significant issue in dealing with data utilization. Creating an integrated interface over a given set of existing heterogeneous databases is a challenge many database administrators face today. In the early stages, developers used the Extensible Markup Language (XML) data model to experiment with the datasets. Since XML could not address the semantic heterogeneity, it relied on other computer-based approaches. The vision of the proposed work was on ontology, an evolving research area that provides a sample scope for addressing semantic heterogeneity. The research area focuses on India's south-eastern coastal regions on which the proposed method has tested.

Approximately 1200 weather stations have been established across India, regularly generating copious data. The generated datasets were referenced geographically in various geospatial file formats that rely on new IR approaches. Also, the data format, naming convention, and units were different for different sources. Ontology addresses the semantic interoperability among the weather systems to utilize the information effectively.

Frequently reported weather phenomena along India's coastal areas are cyclones, storms, water spouts, heavy monsoon rains, and marine heatwaves. According to recent research, the Indian Ocean is the warmest among all five oceans generating 7% of the world's cyclones [8]. Climatic change leads to continuous warming of the Indian Ocean, resulting in an increased number of severe storms on the east and west coasts of the Indian subcontinent [9]. Ever since 1976, the highest number of storms was reported in the year 2020 [10]. In the following year, the very first cyclone of 2021, Cyclone Tauktae, hit southern Gujarat on 17 May 2021. Another storm named Yaas was formed in the Bay of Bengal coasts on 26 May 2021, followed by Javid [11]. The latest research also shows that the strongest storm's proposition increases about 8% every decade [12]. Accu Weather climatic web portal reports nearly eighteen storms of the Indian Ocean in 2020 with wind speed greater than 65 km/h.

Among Nazare dam's water bodies in Pune and Yanam on the Eastern coasts of India, some water spouts were reported in 2018, and 2020 respectively [13]. Likewise, Heavy monsoonal rains across India caused floods and landslides, creating tremendous harm to individuals' life and assets. The India Meteorological Department cautioned about very heavy rains in the country's northwestern regions [14]. The IMD data show that 341 weather stations reported heavy rainfall, measuring above 20 cm in 2020, compared to 554 in 2019; and 321 and 261 in 2018 and 2017, respectively [15]. The Intergovernmental Panel on Climatic Change (IPCC) published a special report discussing global warming on oceans, a necessary message for India [16]. This report warns India about marine heatwaves, which cause severe damage to marine sea life and corals. The causes and effects of these weather phenomena have led to incredible damage in recent years; hence the Indian Ocean's research has been gaining consideration day by day. Due to these issues, the research area focuses on weather phenomena in and around the Indian Ocean. The ocean data analysis leads to many applications like marine safety, weather forecasting, fishing, aquatic lives, disasters, and the rest.

### 3 Literature Review

Government and private sectors are increasingly getting committed to transparency of information regarding satellite data [17]; hence, the publication of heterogeneous Open Data (OD) on the web is increasing. To establish the uniformity of data, interoperability, and heterogeneity, data must be presented in a standard machine-understandable format like RDF. Heterogeneous geospatial files are converted to RDF, making it easier to retrieve the information through the SPARQL protocol [18]. Some research works present the transformation of meteorological data into RDF over Spain [19] and Iran [20] for weather prediction analysis. But this method provides a problem in decision making about the cases or properties of a particular size for any parameter. Security and data access are significant problems in the data web. Satellite Awareness Technique (SAT) and Ontology-based Data Access Control (ODAC) improve cloud environments' secret sharing schemes. Research work focuses on providing Common Core Ontologies (CCO) [21] to help developers build Denial of Service based on the application domain. Though the developed ontologies reduce the cost associated with upgradation, they lack ontology quality information. Ontologies can be combined with Artificial Intelligence (AI) and Artificial Neural Networks

(ANN) to represent the knowledge base through modular structure but suffer some limitations like a knowledge construction task.

The research works [22,23] report about the prediction of weather events by developing an ontology, namely, a forecast model for solar irradiance and blizzard weather events by Canadian Climate Archives (CCA) datasets. Ontology-based data access and integration using the weather datasets of Nepal presents IR and data usage by converting them into RDF for agricultural applications [24]. Although these methods result in a high rate of completeness and accuracy, they lack predicting other different phenomena affected by the same weather parameters. The study has reported a basic understanding of the existing weather forecasting systems and their technologies supporting meteorological Decision Support Systems (DSS). It also explains the efforts in developing new weather ontology but lacks in testing and quality assessment of ontology. AEMIX, a new project, has been designed to quantify the accuracy of weather forecasts provided by various websites based on Natural Language Processing (NLP) [25]. It obtains weather information from forecasting systems with the help of ontologies. Some work proposes creating ontology from XML by mapping two layers, namely the concept layer, which includes the ontology and instance layer, which addresses the instances of those concepts [26]. Copious data and semantics lead to the loss of information during the XML to the ontology conversion process. Apart from manual construction, ontology is created in a semi-automatic way [27] through conceptual clustering. The concepts of weather ontology are collected based on their semantic data similarities to construct a hierarchy using related knowledge sources and expert knowledge. This approach improves data retrieval and reduces search time.

Owing to the increasing popularity of SW, researchers rely on measuring the quality of various aspects such as linked data, ontology, inference engine, data backing, and user interface. Even though ontologies are designed for a particular domain, determining their quality is a challenging task. Researchers have developed software for ontology matching algorithms but lack in terms of efficiency and accuracy. Quality measurement frameworks by various researchers like Semantic Web Application Quality Evaluation (SWAQ), which includes the fuzzy logic for evaluating SW applications' quality attributes, are proposed in [28]. Similarly, Zhu et al. [29] proposed some performance parameters to evaluate the developed ontology for the application domain by choosing the respective Golden Standard (GS) ontology. A novel iterative framework like RiMOM-Instance Matching (RiMOM-IM) helps to match instances in the ontology by discovering the corresponding examples in the knowledge base [30]. It has a source knowledge base to the application domain and the target knowledge base's samples to find the exact ontology matching. Data Mining for Ontology Matching (DMOM) framework compares the cases and the data properties through three stages: exhaustive, statistical, and Frequent Itemset Mining (FIM) using the DBpedia ontology [31]. Similarly, a pattern-matching algorithm proposes to solve ontology matching problems using a pattern mining approach [32]. This method searches for the redundant patterns in the ontology database and matches the target ontology's relevant feature to find efficient matching.

#### **4 Ontology-Driven Information Retrieval for Ocean Data**

This section depicts a detailed explanation of the proposed SW-based approach for ocean Weather Phenomenon Prediction (WPP). It consists of four phases: Input data processing, Ontology creation, SW processing, and Semantic query engine, as shown in Fig. 1. Observation sensors monitor the field area, trace the information in a successive interval of time, and store it in heterogeneous file formats. These data are available to users and researchers by various bureaus

like World Weather Online (WWO), the Indian Meteorological Department (IMD), Accu Weather, and similar climatic web portals.

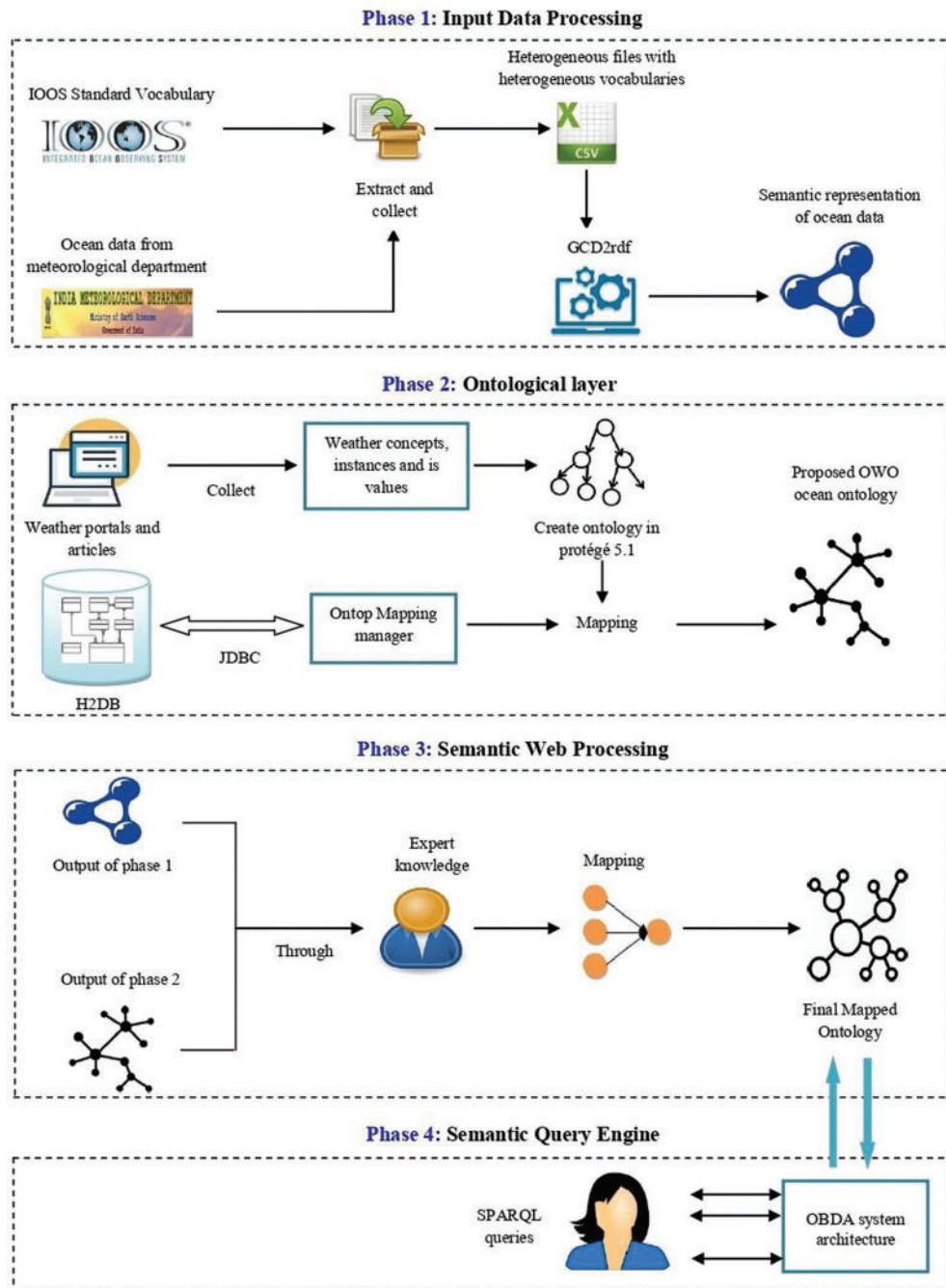


Figure 1: Proposed approach: semantic web-based ocean weather phenomenon prediction

#### ***4.1 Proposed Four-Phase Weather Data Model***

Sensor data, a major data resource, is used by many researchers and domain experts for any type of weather-related researches. It faces two major issues: heterogeneous files and heterogeneous vocabularies, making it difficult to exchange, share, and reuse. Data from different observation sensors are in various formats like Comma Separated Values (CSV, \*.csv), Total's file (TUV, \*.tuv), Excel (\*.xls), Network Common Data Form (NetCDF, \*.nc) and Hierarchical Data Format 5 (HDF5, \*.h5) etc. In the first phase, these data have been presented in a standard machine-understandable format, namely RDF, to facilitate efficient semantic retrieval. Heterogeneous-Geospatial Climatic Data to RDF (Hetero-GCD2RDF) assists the transformation of simple vocabularies into an SW format.

Similarly, the Integrated Ocean Observing System's (IOOS) standard vocabulary format represents the heterogeneous parameters as RDF. IOOS is a common RDF vocabulary created to provide interoperability between the data catalogues published on the web. IOOS standard vocabularies can improve data recovery from various data sources by providing a user-friendly environment using a similar query mechanism.

The second phase provides a conceptualization definition for WPP through ontology named OWO. The proposed ontology follows the concept of DO for the oceanographic weather domain to define how the attributes are related to different weather conditions. Existing satellite data retrieval systems help retrieve information effectively based on geolocation, date and time, sensor, satellite, weather attribute, etc. In that case, retrieval of additional, domain-specific, concept-based data generated by satellites from copious information is challenging. For instance, retrieving the values of weather parameters like wind speed, precipitation or temperature can be queried using API. But the retrieval of specific knowledge-based data such as fresh gale, no wind, or fresh breeze is a complex task. Hence, the proposed IR approach should support semantic concepts or knowledge-based satellite data retrieval to facilitate an efficient retrieval system and fulfill the user requirements. Ontology-based web service helps in efficient knowledge representation and data access by enhancing existing satellite data retrieval methods. Once the ontology is created for a particular application domain, updating and maintenance become the developer's responsibility to produce the accuracy of IR. The proposed OWO ontology consists of 37 concepts, 112 instances, 85 relations, and 126 attributes. [Tab. 1](#) illustrates some preliminary ideas and sub-concepts of the developed OWO ontology.

Each sub-concept is categorized further into several instances/individuals. For example, the individuals of the concept Wind\_Speed are NoWind, LightAir, LightBreeze, GentleBreeze, ModerateBreeze, FreshBreeze, StrongBreeze, ModerateGale, NearGale, FreshGale, StrongGale, WholeGale, Storm, ViolentStorm, and Hurricane. Each instance holds a range of data values of different data types through a data property. The value of any attribute of a concept has a data value through an object property. Various weather sites and analysis reports provide details about the classifications of ontology instances and their values. The weather concepts are related to each other by 85 different relational links. The relations are: is-a, has-a, has long, has an attribute, condition, etc. By incorporating all the specifications, the protégé tool assists in developing the proposed OWO ontology. In addition to that, the OWO ontology includes four different ocean weather phenomena: storm, cyclone, water spout, and marine heatwaves. The weather factors related to these phenomena, along with the values, are included in the ontology. [Tab. 2](#) illustrates the details of the weather factors included in each weather phenomenon.

**Table 1:** Top-level concept and sub-concept information of the developed OWO ontology

Top-level concepts	Respective sub-concepts
GEO_LOCATION	Latitude, longitude
DATE	Instances, date, year
TIME	Interval, hours, minutes, seconds
WEATHER_ATTRIBUTES	Fog, Aerosol_Optical_Thickness, Cloud_Cover, Precipitation_Rate, Relative_Humidity, Wind_Speed, Wind_Direction, Wind_Gust, Barometric_Pressure, Atmospheric_Pressure, Sea_Surface_Temperature, Air_Temperature, Conductivity, Solar_Irradiance, UV_Index
WEATHER_CONDITION	Good, bad, severe
WEATHER_PHENOMENON	Cyclone, storm, water spout, marine heat waves

**Table 2:** Weather attributes involved in each phenomenon

Weather phenomenon	Weather attributes
Cyclone	Barometric_Pressure, Relative_Humidity, Wind_Speed, Precipitation_Rate, Sea_Surface_Temperature
Storm	Barometric_Pressure, Precipitation_Rate, Wind_Speed, Cloud_Cover
Water Spout	Relative_Humidity, Wind_Speed, Air_Temperature, Water_Temperature
Marine Heat Waves	Relative_Humidity, Sea_Surface_Temperature, Water_Temperature_100 m, Water_Temperature_200 m, Cloud_Cover, Precipitation_Rate

Ontologies developed and used in online systems are larger; hence a database is mandatory for storage, efficiency, and optimal utilization [33]. The advantage of storing ontology in databases is that it supports performance, robustness, reliability, availability, legacy, and large-scale ontologies. In this paper, the H2 Database (H2DB) console is considered a Relational Database Management System (RDBMS) written in java. H2DB is preferred over other databases since it is speedy, Open-Source, contains a scrollable result set and browser-based application. It supports Transmission Control Protocol (TCP), Java Database Connectivity (JDBC), and a web server that connects it to the ontology. This research develops a relational database with the help of the H2 Engine that depicts the internal relations between the sub-concepts in OWO ontology. An on top mapping manager is used to map the proposed OWO ontology with the database.

Mappings are hypotheses that are used to relate the data in RDBMS to the vocabulary of ontology. Hence, the mapped ontology makes it easier for the user to retrieve data through the query engine. In the third phase, the semantic layer directly interacts with DO through Jena API. Jena's model factory can create an inference graph by connecting datasets with a reasoner supporting a general-purpose rule engine. The main aim of the reasoner is to answer the queries by transforming them into questions over the source. This platform uses mappings by allowing users to access the data from multiple resources through a single interface. W3C provides a

standard called RDB to RDF Mapping Language (R2RML) for the specification of mapping in Ontology-Based Data Access (OBDA). The IOOS standard vocabularies are incorporated in RDF vocabulary and ontology concepts as well. Hence, it makes it easier for the query engine to map the data, which returns the data value and knowledge.

The final phase grants IR through queries for any weather-related applications. A typical question of OBDA is generally expressive in that it describes the user's desire instead of training the system on answering. The OBDA allows the query to be independent of the data source and uniform access to heterogeneous sources. There are numerous query languages designed for RDF databases, namely SPARQL protocol, RDF Data Query Language (RDQL), RDF Query Language (RQL), Versa and Sesame RDF Query Language (SeRQL), etc. Still, SPARQL is widely preferred [34].

The concept of writing a SPARQL query is to match its triples with the RDF triples and retrieve the queried information. Users can access more information by querying an integrated database built and saved in the ontology. Queries can be made user-friendly by creating a web page using Java Servlet Pages (JSP) to retrieve the information dynamically. JSP pages help to create a dynamic website that is easier to maintain compared to other servlets. It adds Java code inside Hypertext Markup Language (HTML) code using JSP tags. Using JSP, one can easily separate Presentation, and Business logic as a web designer can design and update JSP pages creating the presentation layer. Java developers can write server-side complex computational code without concerning the web design. And both the layers can easily interact over HyperText Transfer Protocol (HTTP) requests. Even though various environments are available for developing an ontology, evaluating the ontology's quality is still challenging. Hence, some performance metrics are discussed in Section 4.2 to assess the quality of developed OWO ontology.

#### 4.2 Performance Metrics of Ontology

The quality of any ontology is evaluated against another ontology called GS ontology. Let the developed weather ontology be  $O$ , which holds a record  $(C, I, A, R)$ .  $C, I, A, R$  is a finite set of classes, instances, attributes, and relations defined in the ontology.  $c \in C$ ,  $\alpha \in I^c$  and  $\Psi \in A^c$  are the elements of each record.  $R = \{r_1, r_2, r_3, \dots, r_n\}$  is  $n$  number of relations where each  $r$  defines a connection between the concepts  $c$ . Size of ontology  $Size(O)$  is defined using Eq. (1).

$$Size(O) = Size_C(O) + Size_I(O) + Size_A(O) + Size_R(O) \quad (1)$$

The size of classes  $Size_C(O)$ , individuals  $Size_I(O)$ , attributes  $Size_A(O)$  and relations  $Size_R(O)$  is determined by the individual expression given in Eq. (2) using the mathematical operator  $\|$  (norm), which measures a linear map's size. In a linear map, mapping is represented between two modules; for instance,  $X \rightarrow Y$  where  $X$  denotes the class and  $Y$  indicates sub-class.

$$Size_C(O) = \|C\|, \quad Size_I(O) = \sum_{c \in C} \|I^c\|, \quad Size_A(O) = \sum_{c \in C} \|A^c\|, \quad Size_R(O) = \sum_{r \in R} \|r\| \quad (2)$$

Let us define the GS ontology as  $\omega$ . Vocabulary is the name specific to the classes, individuals, attributes, and relations in the proposed ontology. Vocabulary coverage of ontology has to be calculated individually using the respective expressions in Eq. (3).



$$\begin{aligned} \text{Cov}_C^\omega(\mathbf{O}) &= \frac{\|C \cap C'\|}{\text{Size}_C(\omega)}, & \text{Cov}_I^\omega(\mathbf{O}) &= \frac{\sum_{c \in C} \|I^c \cap I'^c\|}{\text{Size}_I(\omega)}, \\ \text{Cov}_A^\omega(\mathbf{O}) &= \frac{\sum_{c \in C} \|A^c \cap A'^c\|}{\text{Size}_A(\omega)}, & \text{Cov}_R^\omega(\mathbf{O}) &= \frac{\sum_{r \in R} \|r \cap r'\|}{\text{Size}_R(\omega)} \end{aligned} \quad (3)$$

where,  $C', I', A'$  and  $r'$  are the classes, individuals, attributes, and relations of the GS ontology  $\omega$ . The overall coverage of ontology is derived by using Eq. (4).

$$\text{Cov}^\omega(\mathbf{O}) = \frac{\|C \cap C'\| + \sum_{c \in C} \|I^c \cap I'^c\| + \sum_{c \in C} \|A^c \cap A'^c\| + \sum_{r \in R} \|r \cap r'\|}{\text{Size}(\omega)} \quad (4)$$

Semantic coverage metrics  $SCov_C^\omega, SCov_I^\omega(\mathbf{O}), SCov_A^\omega(\mathbf{O})$  and  $SCov_R^\omega(\mathbf{O})$  are evaluated similar to the above expressions, but the coverage is calculated, including the classes defined in  $\omega$ , which can be derivable from proposed ontology  $\mathbf{O}$ . Where  $D_C, D_I, D_A$  and  $D_R$  are the classes, instances, attributes, and relations, including derived elements of the ontology. For example, ontology has a named type, “weather phenomenon,” but the same information can be retrieved from the classes “weather state” and “weather attributes”; hence, the class “weather phenomenon” is said to be a derivable class. The overall semantic coverage is given by Eq. (5).

$$SCov^\omega(\mathbf{O}) = \frac{D_C + D_I + D_A + D_R}{\text{Size}(\omega)} \quad (5)$$

Ontology is said to be semantically compatible only if the contents are reliable to GS ontology  $\omega$ . Compatibility metrics  $RCC^\omega, ARCI^\omega, ARCA^\omega$  and  $ARCR^\omega$  are calculated as per Eqs. (6) and (7).

$$RCC^\omega(\mathbf{O}) = \frac{\|c \in C \mid \omega\|}{\text{Size}_C(\mathbf{O})}, \quad ARCR^\omega(\mathbf{O}) = \frac{\sum_{r \in R} \|(x, y) \in r \mid \omega\|}{\|R\|} \quad (6)$$

$$ARCI^\omega(\mathbf{O}) = \frac{\sum_{c \in C} \frac{\|\alpha \in I^c \mid \omega\|}{\|I^c\|} \forall c}{\text{Size}_I(\mathbf{O})}, \quad ARCA^\omega(\mathbf{O}) = \frac{\sum_{c \in C} \frac{\|\psi \in A^c \mid \omega\|}{\|A^c\|} \forall c}{\text{Size}_A(\mathbf{O})} \quad (7)$$

An element of ontology is said to be redundant if it can be derived from other factors. For instance, a concept is defined in the ontology, which can also be derived from other concepts; hence, the idea is redundant. Expressions in Eq. (8) are the redundant metrics of ontology.

$$CR(\mathbf{O}) = \frac{\|C_R\|}{\text{Size}_C(\mathbf{O})}, \quad IR(\mathbf{O}) = \sum_{c \in C} \frac{\|I_R^c\|}{\text{Size}_I(\mathbf{O})}, \quad AR(\mathbf{O}) = \sum_{c \in C} \frac{\|A_R^c\|}{\text{Size}_A(\mathbf{O})}, \quad RR(\mathbf{O}) = \sum_{r' \in R'} \frac{\|r'\|}{\text{Size}_R(\mathbf{O})} \quad (8)$$

where  $C_R, I_R, A_R$  and  $r'$  are said to be the most extensive set of redundant elements, respectively.

The graph of any ontology is denoted by  $G(\mathbf{O}) = (N, E)$  where  $n \in N$  and  $e \in E$  are nodes and edges. A node  $n$  is a root node if no edge  $e$  enters the node or leaf node or if no edge  $e$  leaves it. For any relation  $r \in R$  the relation-based structural metrics are  $NRN, NLN, MaxSPL, NIC, TNRNR$  and  $ANRNR$ .  $NRN$  and  $NLN$  are root nodes and leaf nodes,

respectively, as per Eq. (9). The isolated nodes  $NIC$  specify the node not linked to any other node in the graph as expressed in Eq. (10).

$$NRN(O) = \|Rootnodes(O)\|, NLN(O) = \|Leafnodes(O)\| \quad (9)$$

$$NIC(O) = \|Rootnodes(O) \cap Leafnodes(O)\| \quad (10)$$

The length of a path  $p$  from node  $a$  to node  $b$  is specified by the number of nodes in the path from  $a$  to  $b$ . The maximum size of the ontology graph is denoted as  $MaxSPL$ , as expressed in Eq. (11).

$$MaxSPL(O) = \underset{p \in path(O)}{Max} (Length(p)) \quad (11)$$

Set of reachable nodes  $Reachable^O(c)$  from root node  $c \in Root(O)$  is denoted as  $TNRNR$ . Similarly, the average number of reachable nodes  $ANRNR$  from the root node  $c$  is estimated in Eq. (12).

$$TNRNR(O) = \sum_{c \in Root(O)} \|Reachable^O(c)\|, \quad ANRNR(O) = \frac{TNRNR(O)}{\|NRN(O)\|} \quad (12)$$

Relation  $r$  of any ontology is said to be acyclic. In an acyclic graph, the depth of a node  $n$  is denoted by the longest path of the root node. Similarly, the width of a node  $n$  is indicated by the number of nodes it is related to through a relation  $r$ . The average depth  $ADLN(O)$ , maximum depth  $MaxDepth(O)$ , average width  $AWNLN(O)$  and maximum width  $MaxWidth(O)$  are calculated using Eqs. (13) and (14).

$$ADLN(O) = \frac{\sum_{C \in Leaf(O)} Depth^O(C)}{NLN(O)}, \quad MaxDepth(O) = \underset{c \in Leaf(O)}{Max} (Depth^O(C)) \quad (13)$$

$$AWNLN(O) = \frac{\sum_{C \notin Leaf(O)} Width^O(C)}{NAN(O) - NLN(O)}, \quad MaxWidth(O) = \underset{c \notin Leaf(O)}{Max} (Width^O(C)) \quad (14)$$

## 5 Evaluation Results and Discussion

The proposed data model experimented using Indian Meteorological Satellite INSAT-3D's ocean data collected along with the south-eastern coastal areas of India. Observation sensors like Agro Floats (NetCDF), Buoys (CSV), Coastal Radars (TUV), Gliders (NetCDF), Sonde (Excel/CSV), and others monitor the field area. The proposed work used a 64-bit Intel Core i5 processor with 4 GB of RAM, 2 TB hard disk and deployed in LINUX (Ubuntu16.04 version) system. The java (JDK 1.8.0\_181) code for research work developed on Eclipse 5.0, Apache Jena with Apache Tomcat 9.0.14, JSP as a server, and Internet Explorer or any web browser as a client.

### 5.1 Framework Implementation

A sensor observed data are in heterogeneous format. Before dealing with the sensor followed data, it is harmonized by converting it into an SW supportable RDF file format. The input data processing phase is essential to produce a machine-readable format for a computer to search and understand how the terms of a particular domain are related to each other. Once the datasets are presented as RDF, many tools are available for visualizing and working with the

data stored in them. In the second phase of OWO, the weather ontology is developed using the protégé 5.1 tool by deploying the specifications mentioned in Section 4.1. It serves as a knowledge base for ocean WPP and includes the hierarchy of weather conditions, attributes related to weather conditions, and their relationship. Elements of WPP consist of concepts, sub-concepts, attributes, instances, and relations between them. For example, individuals of the sub-concept Relative\_Humidity are Dry\_Humidity, Optimum\_Humidity, and Moist\_Humidity related to the top-level concept Weather\_Attribute through the link named “Has\_humidity,” which holds the range of data values <35, 35 to 60, and > 60 respectively of data type “xsd: decimal” measured in “%.” The developed OWO ontology incorporates the respective values, for instance, men, and the same information has been deployed in the H2 relational database as tables by describing their relationship. The third phase on the top mapping tool helps map the input data and the proposed ontology through the domain expert’s knowledge. Ontology mapping is a critical phase of the knowledge-building process. The current research work creates instances from sensor data and maps with ontology concepts concerning ontology vocabulary and sensor data vocabulary. RDF represents instance mapping data in a homogeneous machine-understandable format, supporting inference and enabling semantic IR. The proposed OWO ontology includes the concepts of weather phenomenon illustrated in Tab. 2, helping the easy recovery of data and knowledge for any end-user through a querying protocol.

After completing the ontology mapping, the SPARQL query is applied to the RDF graph to extract the information in the fourth phase of the proposed data model. Fig. 2 represents the overall IR process of the proposed satellite ocean data and knowledge retrieval. An Excel (\*.xls) file is taken, which stores the weather parameters’ values and the geographical location recorded by the buoy named BD14. The next stage represents the developed weather ontology describing the concepts and their relationships of the ocean application domain. The final section of Fig. 2 describes that the user has queried for the value of parameter “wind\_speed” recorded by buoy “BD14” from the Sensor Observation Services (SOS) tab. The user can choose the location, date, and time from which the data has to query output format such as XML, JavaScript Object Notation (JSON), and prefer a table. For instance, user has chosen parameter as “wind\_speed”, given “latitude” value as “7.007”, “longitude” value as “88.005”, range of date and time as “from 01/01/2015, 00:00 IST to 03/01/2015, 06:00 IST”, and the output format in “JSON”. First, the server queries for the value of “wind\_speed” recorded in specified date and time with latitude value “7.007” and longitude value “88.005” from input Excel file converted as RDF. According to user preferences in the query, the value of “wind\_speed” is “1.91” retrieved from the dataset. Then server maps extract value “1.91” from the knowledge base and identify the range under which this value falls and the ontology instance. The value “1.91” falls in the range “1.6 to 3.3 ms” of the concept “wind\_speed” and the ontology instance “Light breeze.” The final results are displayed in the SOS services tab in “JSON” format with latitude value, longitude value, date and time, the resultant value “1.91,” and ontology instance “Light breeze.”

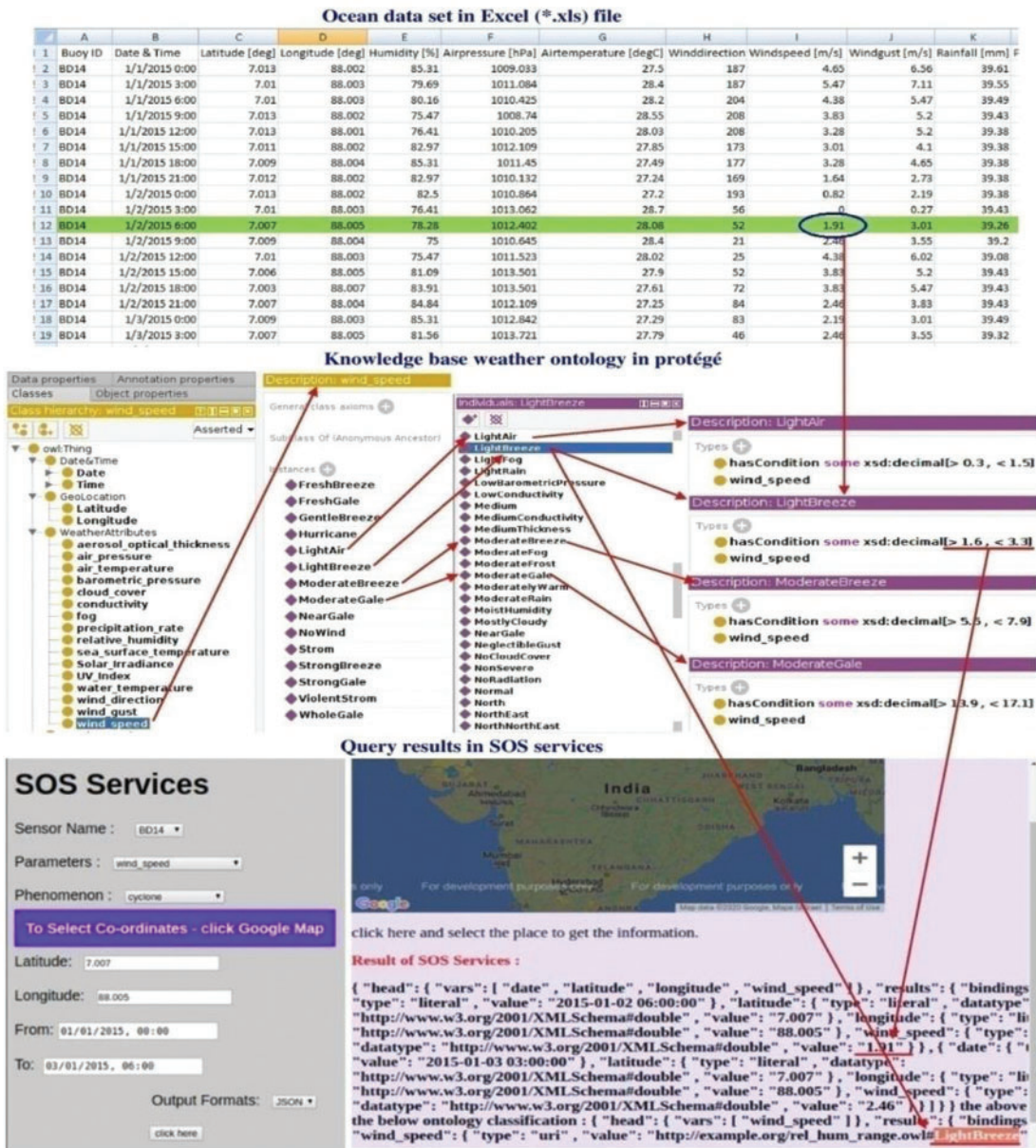


Figure 2: Query output in sensor observation services

### 5.2 Performance Evaluation

An experimental study demonstrates the proposed ontology's quality with the other state-of-the-art ontologies. Some existing weather ontologies O1 [35], O2 [36], O3 [37] and O4 [38] are considered for performance evaluation. Proposed OWO Ontology's (PO) performance is evaluated against another ontology called the GS ontology [39], as described in Section 4.2. Ontology is superior and efficient only by calculating some performance metrics to prove its quality over other ontologies. The quality of an ontology is evaluated by calculating the metrics against other ontologies of the same application domain, namely, GS. Researchers have already implemented

parametric measures to find the semantic similarity and coverage of the proposed ontology with state-of-the-art techniques [40]. The ontology elements like classes, instances, attributes and relations are extracted from the above ontologies as illustrated in Tab. 3. In some cases, classes defined in the ontology can also be derived from other elements defined in the ontology, as shown in Tab. 4.

**Table 3:** Characteristics of ontologies

Ontology		Classes	Instances	Attributes	Relations
Golden standard ontology (GS)		15	11	18	27
Weather domain ontologies	O1	13	72	31	11
	O2	24	78	37	89
	O3	19	11	52	37
	O4	35	49	114	84
	PO	37	112	126	85

**Table 4:** Classes defined in ontologies to GS

Defining classes	O1 (WO <sub>1</sub> )	O2 (WO <sub>2</sub> )	O2 (WO <sub>3</sub> )	O4 (WO <sub>4</sub> )	PO (WO <sub>5</sub> )
Defined both in GS and WO <sub>n</sub>	7	9	6	11	12
Defined in GS and derivable from WO <sub>n</sub>	1	1	0	0	1
Defined in GS but not in WO <sub>n</sub>	7	6	9	4	2
Defined in WO <sub>n</sub> but not in GS	4	8	13	24	24

The study evaluates the performance metrics of ontologies expressed in Section 4.2 on the chosen ontologies, metrics  $Cov_C^\omega$ ,  $Cov_I^\omega$ ,  $Cov_A^\omega$ ,  $Cov_R^\omega$  and  $Cov^\omega$  comes under vocabulary coverage and metrics  $SCov_C^\omega$ ,  $SCov_I^\omega$ ,  $SCov_A^\omega$ ,  $SCov_R^\omega$  and  $SCov^\omega$  comes under semantic coverage. Metrics  $RCC^\omega$ ,  $ARCI^\omega$ ,  $ARCA^\omega$  and  $ARCR^\omega$  evaluates semantic compatibility of the proposed ontology for the GS ontology. Similarly,  $CR$ ,  $IR$ ,  $AR$ , and  $RR$  evaluates the conciseness of the proposed ontology. The cohesion factor has two categories: (1) relation-based metrics, namely  $NRN$ ,  $NLN$ ,  $MaxSPL$ ,  $NIC$ ,  $TNRNR$  and  $ANRNR$ , and (2) metrics for acyclic relations, namely  $ADLN$ ,  $AWNLN$ ,  $MaxDepth$  and  $MaxWidth$ . This paper evaluates four major quality factors of ontology, such as completeness (COM), correctness (COR), conciseness (CON), and Structural Complexity (SC), where the metrics involved are shown in Tab. 5.

**Table 5:** Quality factors of ontologies

Quality factors	Performance metrics involved
Completeness	$Cov_C^\omega$ , $Cov_I^\omega$ , $Cov_A^\omega$ , $Cov_R^\omega$ , $Cov^\omega$ , $SCov_C^\omega$ , $SCov_I^\omega$ , $SCov_A^\omega$ , $SCov_R^\omega$ , $SCov^\omega$
Correctness	$RCC^\omega$ , $ARCI^\omega$ , $ARCA^\omega$ , $ARCR^\omega$
Conciseness	$CR$ , $IR$ , $AR$ , $RR$
Structural complexity	$NRN$ , $NLN$ , $MaxSPL$ , $NIC$ , $TNRNR$ , $ANRNR$ , $ADLN$ , $AWNLN$ , $MaxDepth$ , $MaxWidth$

The evaluated results of performance metrics of proposed ontology are illustrated further in this paper. Completeness metrics of proposed ontology against the GS ontology evaluates by considering either vocabulary coverage or semantic coverage. The experimental results show that the PO is highly correlated with the GS ontology than the existing ontologies. It can be noted from Fig. 3a that the proposed ontology results in the highest score of vocabulary coverage with  $Cov^{\omega}(O)$  as 0.77 and semantic  $SCov^{\omega}(O)$  as 0.79, which is high among the other ontologies. Similarly, an ontology of the same domain consists of redundant elements, which illustrate the conciseness of ontology as presented in Fig. 3b. Ontology is said to be efficient and concise only if it has minor redundancy elements. Results show that redundant features are present only in the relations  $r$  of ontologies. The value of  $RR$  is estimated to be very low in the proposed ontology, i.e., 0.89, compared to O1, O2, O3, and O4, which scores 1, 0.97, 1, and 0.93, respectively. Hence the proposed ontology is efficient and concise.

The compatibility factor generally indicates the correctness of ontology to GS ontology. However, weather ontologies include many elements: some don't have domain knowledge but are subject to a particular web service. Hence, all the elements mentioned are not equivalent to GS ontology. Fig. 4 illustrates that the compatibility of PO is high compared to the other ontologies. Where,  $RCC^{\omega}$ ,  $ARCI^{\omega}$ ,  $ARCA^{\omega}$  and  $ARCR^{\omega}$  scores are 0.66, 0.75, 1 and 0.62 respectively. Similarly, cohesion metrics evaluates ontology's structural complexity, including relation-based metrics and metrics for acyclic relations. The biggest difference in the metrics is  $TNRNR$  and  $NLN$ , which belong to relation-based cohesion, as illustrated in Fig. 5. Moreover, all ontologies include low scores in coupling metrics; thus, the ontologies are well-structured. PO ontology holds the values 3.05, 2.09, 3, and 8 for the coupling metrics  $ADLN$ ,  $AWNLN$ ,  $MaxDepth$ , and  $MaxWidth$  respectively.

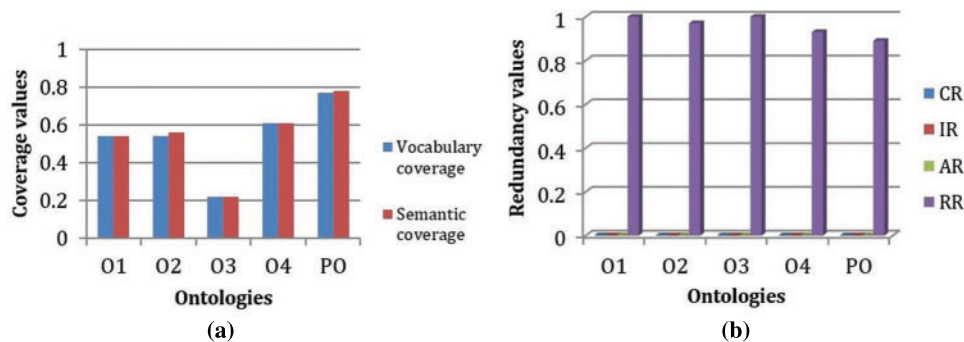


Figure 3: (a) Coverage and (b) redundancy metrics of ontologies

The overall comparison of quality factors, including the corresponding metrics defined in Tab. 5, is aggregated into a single value by taking an average and plotted in Figs. 6a and 6b. The completeness and correctness metrics of the proposed ontology are high at 0.77 and 0.75, respectively; hence, it covers most of the weather function and is compatible with GS ontology. Similarly, the redundancy score is 0.89, which is a minor case of PO. Hence, it is distinct from using redundant elements. Finally, the structural complexity of PO was concluded to be the least complex 10.58, compared to O2, O3, and O4. The structural metrics of O1 is low because it uses the minimum concepts and attributes to develop the ontology. Even though the structural complexity is low, O1 cannot provide accurate results compared to other state-of-art and proposed

ontologies. This paper’s main contribution is to design a practical ontology for ocean WPP based on service-oriented computing. A study has demonstrated the weather domain’s existing ontologies and concluded that the proposed OWO ontology is efficient. The result shows that the proposed ontology is adequate in terms of completeness and uniqueness by scoring least in redundant elements and structurally fewer complexes than other ontologies that hold a larger size. Thus, the quality of the proposed ontology is high compared to the existing ontologies.

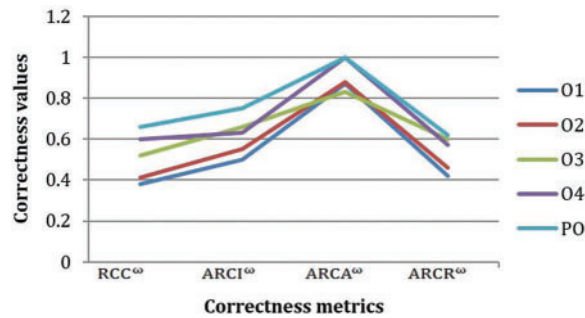


Figure 4: Compatibility chart of ontologies

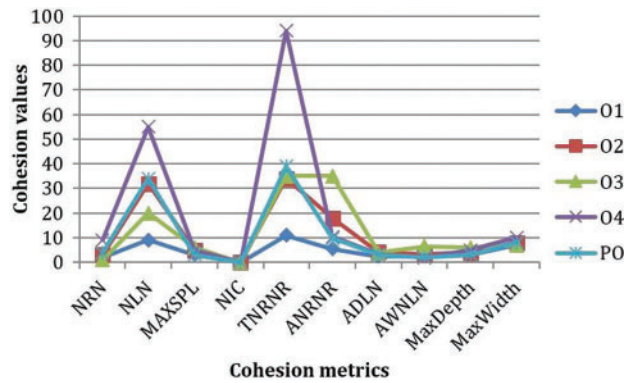


Figure 5: Distribution of coherence metrics of weather ontologies

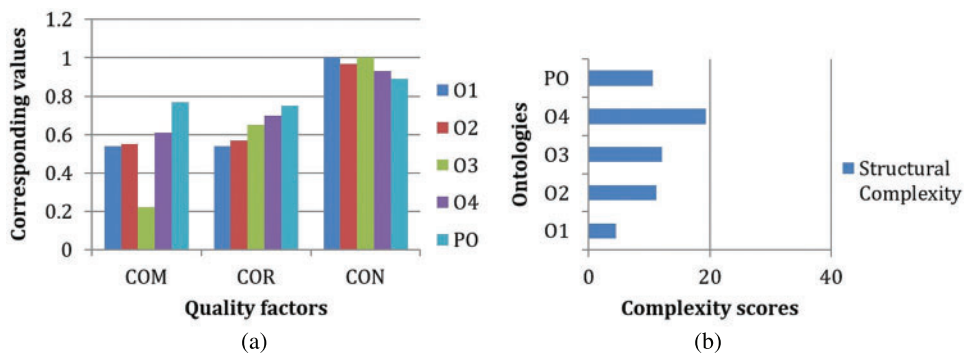


Figure 6: Comparison plot of ontologies (a) Quality factors and (b) Complexity scores

## 6 Conclusions and Future Work

This paper confers a climatic information data model dependent on OWL ontology by coordinating satellite information using the semantic web for ocean applications. The fundamental objective of this data model is to predict the climate marvel through ontology models by presenting information about different climate conditions influenced by weather attributes. The proposed data model has reduced much effort since information translation from the information sources causes versatility issues while processing computing applications. The proposed ontology's performance metrics have improved by 26.23% and 7.14% in completeness and correctness. Likewise, it has diminished by 4.3% and 45.32% in conciseness and structural complexity, respectively, with O4 ontology. The experimental outcomes show that the proposed technique outperforms the current strategies in terms of scalability and IR. The ontological approach sustains various scientific domains, research data sharing, semantic query execution, and efficient visualization. The future work of this research is to develop a web application to deduce weather prediction based on user-defined queries and to develop a more comprehensive ontology to infer any kind of weather forecast. The outlook of this research is to implement various SPARQL queries implementing arithmetic and logical calculations according to the user requirement. A proposed ocean ontology can be developed by including many concepts and supporting attributes for more efficient and accurate knowledge retrieval in the ocean weather domain.

**Acknowledgement:** The authors would like to express their sincere thanks to MoES, Govt of India for their financial support, and Adhiyamaan College of Engineering for their moral support for completing the project successfully.

**Funding Statement:** This work is financially supported by the Ministry of Earth Science (MoES), Government of India, (Grant. No. MoES/36/OOIS/Extra/45/2015), URL: <https://www.moes.gov.in>

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] J. Rio, D. M. Toma, E. Martinez, T. C. O'Reilly, E. Delor *et al.*, "A sensor web architecture for integrating smart oceanographic sensors into the semantic sensor web," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 4, pp. 830–842, 2018.
- [2] I. Nishanbaev, E. Champion and D. A. McMeekin, "A survey of geospatial semantic web for cultural heritage," *Heritage*, vol. 2, no. 2, pp. 1471–1498, 2019.
- [3] I. A. Al-Baltah, A. A. A. Ghani, G. M. Al-Gomaei, F. H. Abdulrazzak and A. A. A. Kharusi, "A scalable semantic data fusion framework for heterogeneous sensors data," *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, pp. 30, 2020.
- [4] G. Urkude and M. Pandey, "Contextual triple inference using a semantic reasoner rule to reduce the weight of semantically annotated data on fail-safe gateway for WSN," *Journal of Ambient Intelligence and Humanized Computing*, vol. 54, pp. 2787, 2021.
- [5] W. Mohsen, M. Aref and K. ElBahnasy, "Cooperative domain ontology reduction based on rough sets," *International Journal of Computer Science and Mobile Computing*, vol. 9, pp. 163–174, 2020.
- [6] D. Kalibatiene and O. Vasilecas, "Survey on ontology languages," in *Proceedings of 10th Int. Conf. on Perspectives in Business Informatics Research*, Riga, Latvia, 2011.
- [7] S. Abburu and S. B. Golla, "Survey on ontology construction tools," *International Journal of Scientific & Engineering Research*, vol. 4, pp. 1748–1752, 2013.



- [8] S. Gupta, I. Jain, P. Johari and M. Lal, "Impact of climate change on tropical cyclones frequency and intensity on Indian coasts," in *Proc. of Int. Conf. on Remote Sensing for Disaster Management*, Visakhapatnam, India, 2017.
- [9] S. Sarkar, "Cyclones rise as climate change heats up Indian ocean," *India Climate Dialogue*, 2020. [Online]. Available: <https://indiaclimatedialogue.net/2020/06/05/cyclones-rise-as-climate-change-heats-up-indianocean/>.
- [10] K. Kambli, "Throwback to a stormy year: A look at the 8 north Indian ocean cyclones of 2019," *The Times of India*, 2019. [Online]. Available: <https://weather.com/en-IN/india/news/news/2020-01-03-stormy-year-8-north-indian-ocean-cyclones-2019>.
- [11] A. Javaid, "List of cyclones that hit India in 2020–2021," *Jagran Josh*, 2021. [Online]. Available: <https://www.jagranjosh.com/general-knowledge/list-of-cyclones-in-india-1591178815-1>.
- [12] K. P. Kossin, K. R. Knapp, T. L. Olander and C. S. Velden, "Global increase in major tropical cyclone exceedance probability over the past four decades," in *Proc. of the National Academy of Sciences of the United States of America*, USA, pp. 11975–11980, 2020.
- [13] T. A. Naidu, "Waterspout leaves locals puzzled in Yanam," *The Hindu*, 2020. [Online]. Available: <https://www.thehindu.com/news/national/andhra-pradesh/waterspout-leaves-locals-puzzled-in-yanam/article32118473.ece>.
- [14] BBC, "Heavy monsoon rains trigger flood chaos across India," *BBC News*, 2020. [Online]. Available: <https://www.bbc.com/news/av/world-asia-india-53831594>.
- [15] J. Nandi, "After 61 years, India records two consecutive years of 'above normal' rain," *Hindustan Times*, 2020. [Online]. Available: <https://www.hindustantimes.com/india-news/after-61-years-india-records-two-consecutive-years-of-above-normal-rain/story-VZXoL0iupT5YCZUBRM9I1L.html>.
- [16] R. M. Koll and R. Murtugudde, "New IPCC report warns of marine heatwaves, extremely severe cyclones," *India Science Wire*, 2019. [Online]. Available: <https://vigyanprasar.gov.in/isw/New-IPCC-report-warns-of-marine-heat-waves-extremely-severe-cyclones.html>.
- [17] A. Judie, O. Fabrizio, S. Simon and A. Sören, "A systematic review of open government data initiatives," *Government Information Quarterly*, vol. 32, no. 4, pp. 399–418, 2015.
- [18] A. Velu and M. Thangavelu, "Hetero-GCD2RDF: An interoperable solution for geospatial climatic data by deploying semantic web technologies," *Wireless Personal Communications*, vol. 117, no. 3, pp. 3527–3551, 2021.
- [19] G. Atemezing, O. Corcho, D. D.Garijo, J. Mora, M. Poveda-Villalon *et al.*, "Transforming meteorological data into linked data," *Semantic Web*, vol. 4, no. 3, pp. 285–290, 2013.
- [20] C. Roussey, S. Bernard, G. André, O. Corcho, G. Sousa *et al.*, "Weather station data publication at Irstea: An implementation report," *Context-Aware Approach for E-Agriculture*, vol. 1401, pp. 89–104, 2014.
- [21] R. Rudnicki, "An overview of the common core ontologies," *CUBRC Inc*, 2019, [Online]. Available: [https://www.nist.gov/system/files/documents/2019/05/30/nist-ai-rfi-cubrc\\_inc\\_004](https://www.nist.gov/system/files/documents/2019/05/30/nist-ai-rfi-cubrc_inc_004).
- [22] A. Kantamneni and L. E. Brown, "An ontology for solar irradiation forecast models," in *Proc.10th Int. Joint Conf. on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Seville, Spain, 2018.
- [23] A. Devaraju and T. Kauppinen, "Sensors tell more than they sense: Modeling and reasoning about sensor observations for understanding weather events," *International Journal of Sensors, Wireless Communications and Control*, vol. 2, no. 1, pp. 14–26, 2012.
- [24] S. Pokharel, M. A. Sherif and J. Lehmann, "Ontology-based data access and integration for improving the effectiveness of farming in Nepal," in *Proc. of WI-IAT '14: IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Technologies*, Warsaw, Poland, 2014.
- [25] A. L. Garrido, M. G. Buey, G. Munoz and J. L. Casado, "AEMIX: Semantic verification of weather forecasts on the web," in *Proc. 12th Int. Conf. on Web Information Systems and Technologies*, Rome, Italy, 2016.

- [26] M. Chen and B. Plale, "From metadata to ontology representation: A case of converting severe weather forecast metadata to an ontology," in *Proc. Am Soc Info Sci-Tech.*, Maryland, US, pp. 1–4, 2012.
- [27] A. Revathi, "A development of background ontology for weather systems through ontology learning," *International Journal of Recent Technology and Engineering*, vol. 8, pp. 168–172, 2020.
- [28] N. Baliyan and S. Kumar, "SWAQ: A semantic web application quality evaluation framework," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 30, no. 6, pp. 953–972, 2018.
- [29] H. Zhu, D. Liu, L. Bayley, A. Aldea, Y. Yang *et al.*, "Quality model and metrics of ontology for semantic descriptions of web services," *Tsinghua Science and Technology*, vol. 22, no. 3, pp. 254–272, 2017.
- [30] C. Shao, L. Hu, J. Li, Z. Wang, T. Chung *et al.*, "RiMOM-IM: A novel iterative framework for instance matching," *Journal of Computer Science and Technology*, vol. 31, no. 1, pp. 185–197, 2016.
- [31] H. Belhadi, K. Akli-Astouati, Y. Djenouri and J. C. Lin, "Data mining-based approach for ontology matching problem," *Applied Intelligence*, vol. 50, no. 4, pp. 1204–1221, 2020.
- [32] H. Belhadi, K. Akli-Astouati, Y. Djenouri and J. C. W. Lin, "Exploring pattern mining for solving the ontology matching problem," in *Proc. of European Conf. on Advances in Databases and Information Systems*, Bled, Slovenia, 2019.
- [33] M. Morsey, J. Lehmann, S. Auer, C. Stadler and S. Hellmann, "DBpedia and the live extraction of structured data from Wikipedia," *Program: Electronic Library and Information Systems*, vol. 46, no. 2, pp. 27–181, 2012.
- [34] P. Peng, L. Zou, M. T. Özsü, L. Chen and D. Zhao, "Processing SPARQL queries over distributed RDF graphs," *The VLDB Journal*, vol. 25, no. 2, pp. 243–268, 2016.
- [35] H. Chabert, Open weather map ontology, T research sector in satellite data processing. London, UK, 2020. [Online]. Available: <https://openweathermap.org>.
- [36] R. K. Roy, "Weather phenomenon prediction using semantic web," Masters Dissertation. East-West University, Bangladesh, 2017.
- [37] L. Denis, World weather online developer, World wide marine online. London, UK, 2019. [Online]. Available: <https://www.worldweatheronline.com>.
- [38] C. Wood, Yahoo weather developer network, Weather developer ontology. London, UK, 2020. [Online]. Available: <https://developer.yahoo.com/weather>.
- [39] W. Kastner, Weather ontology developed by automation systems group, The Institute of computer-aided automation, 2019. [Online]. Available: <https://algo4rail.auto.tuwien.ac.at/~k/thinkhome/ontologies.html>.
- [40] S. Thenmalar, "Enhanced information retrieval and information extraction using ontology," PhD. Dissertation. Anna University, India, 2016.