RESEARCH ARTICLE

Ocean knowledge representation through integration of big data employing semantic web technologies

Anitha Velu[1] · **Menakadevi Thangavelu[1]**

Abstract

Implementation of ocean observation sensors are booming in recent years for encouraging research among coastal areas all over the world. This results in a copious amount of big data which makes it difficult for traditional data processing applications to manage them. The complexity in ocean observing community is heterogeneity and interpretation, which directs to a high-end information retrieval system. World Wide Web Consortium (W3C) spreads the usage of Semantic Web (SW) that provide easier way to search, reuse, combine and share information by integrating the data into a single platform. The use of semantic web in big data management helps to increase end-users ability for self-management of data from various sources, to handle the concepts and relationships of a domain and to manage the terminologies while connecting data from a varied data sources. This paper focuses on integrating big data with semantic web technology by developing a knowledge base system through ontology to solve the problem of heterogeneity in ocean observing communities. Ontology refers to a set of machine-readable controlled vocabularies which interprets big data by combining the data concepts with ontology classes. The proposed data model upgrades the information system in terms of improvising data analysis, discovery, retrieval and decision making. In addition to that, this paper also evaluates the quality of proposed ontology and found to be 39.28% improved in completeness, 45.29% reduced in structural complexity, 11% and 37.7% decreased in conciseness and correctness, respectively.

Keywords Big data · Web semantics · Machine-readable format · Knowledge representation · Ontologies · Information retrieval

Introduction

Geospatial data refers to the information which describes any feature or event with a location on or near the earth's surface. It consists of location information (co-ordinates on earth), attribute information (event or phenomena) and temporal information (the time at which the attributes exist). A typical geospatial data involves large sets of spatial data (big data) gathered from a diverse sources in varying formats (heterogeneous formats). For example, satellite image, weather data, social media data, census data etc. Every year several marine incidents happen across the world in oceans, coastal regions and near seashores (Shen et al. 2019) particularly

due to weather phenomenon, which causes severe damage to humans and marine lives. Hence, weather prediction and analysis among the coastal areas is an essential factor. Ocean observation sensors are implemented across the globe to collect weather information which includes physical, chemical and, biological factors of ocean. Due to the rapid increase in observation sensors a huge amount of weather data is generated for research among the specified areas. Moreover, the recorded information is said to hold heterogeneous formats and heterogeneous vocabularies that causes data exchange and integration cumbersome (Rio et al. 2018). Precise information retrieval from the available resources is challenging in Information Engineering (IE) techniques (Ali et al. 2017). Geospatial Information Systems (GIS) provides a better representation through physical mapping of data which is referred as Linked Data (LD).

Semantic Web (web 3.0) is an upgraded version of current web (web 2.0) which enables the data connectivity through LD for decision-making and sharing information across the

✉ Anitha Velu
   aniveluece@gmail.com

[1] Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering, Hosur 635 109, Tamil Nadu, India

application domain (Chughtai et al. 2017). In general, big data has three dimensions namely volume (copious data), velocity (handling real-time data) and variety (dealing with different sources of data). Most of the research works concentrates on handling big data through providing a solution for volume and velocity but the issues related to variety is also equally important to solve the real world problems. SW provides a path way for complications related to variety in big data analysis by integrating data from multiple sources of different formats into a single platform. The data in semantic web are published in machine-readable, understandable and processable format by providing knowledge about the data through defining ontology vocabularies (Nishanbaev et al. 2019).

Ontology is a W3C approved technology that provides an advantage of standard vocabulary with added robustness (Abburu et al. 2015). In computer science and information science, ontology refers to a formal framework for representing knowledge which defines the types, properties and inter-relationships of the entities in an application domain. Ontology is used to support interoperability and a common understanding of domain knowledge between users and wide spread application systems by enabling semantic interoperability between different web applications and services. Ontologies are developed to describe a domain's knowledge in order to make the machines understand the user requirements. Ontology Inference Layer (OIL), 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 (Kalibatiene and Vasilecas 2011) where OWL is widely preferred. Similarly WebOnto, NeOn, SWOOP, WebODE, OilEd, Protégé and OntoEdit are some of the platforms available for ontology development, where Protégé is analyzed to have various functionalities (Abburu and Golla 2013). When the data is published in machine-readable format with semantic descriptions, it is easy to search and access the data dynamically by writing Sparql Protocol and RDF Query Language (SPARQL). This means machines can infer the operations such as service request and response at run time.

Ontology is of three basic types namely; generic ontology, core ontology and domain ontology. Domain Ontology (DO) is built to represent a semantic relationship among the data for a particular application domain. Information and communication technology widely acknowledges the importance of DOs, predominantly for semantic web. DOs are constructed manually, which costs a significant workforce and involves manual invention by combining expert knowledge and domain knowledge results from Machine Learning (ML) (Mohsen et al. 2020). In this scenario, there is a need to build ontology addressing rules and axioms that give precise answers to the user queries supporting data identification and extraction (Hasany and Alwatban 2017). Certainly,

this task is lengthy, costly, and contentious in domain application since different researchers have different views about the same concept. In IE, ontology commits to learn the focus on concepts and relationship between the concepts of an application domain.

The remainder of this paper is structured as follows: Section 2 presents a detailed literature work and background of the research area. Section 3 proposes a four-phase ocean weather data model by integrating the copious sensor data with semantic web through building a new ontology for knowledge retrieval and some performance metrics considered for evaluating the quality of proposed ontology. The results and discussions of data integration and developed web service with quality analysis of the developed weather ontology are incorporated in Section 4. Finally, the paper concludes the work in Section 5 along with future enhancement of the research work.

Motivation and requirement

Background of the research area

Information Technology (IT) has provided extraordinary progress over the years, but in the field of environmental management many have not been addressed yet (Roy 2017). The proposed framework utilizes Geospatial Climatic Data (GCD) collected along south-eastern coastal areas of India where the data model have been implemented and tested. Approximately 1200 weather stations are deployed across India, and the data has been generated on a regular interval basis which results in voluminous data. The generated data is geographically referenced which includes location and time, resulting in a wide range of geospatial file formats that rely on new Information Retrieval (IR) approaches. Various websites like Indian Meteorological Department (IMD), Open weather map, Accu Weather, and others provide weather information by monitoring ocean regions. The provided datasets includes different data format, naming convention and units for different sources. To utilize the information effectively, semantic interoperability among the weather systems has to be addressed. Semantic web technology addresses these issues through domain knowledge representation using ontology.

India has a unique geo-climatic conditions and high socioeconomic vulnerability to calamities which are responsible for increased recurrences in natural disasters. Some weather phenomena like rainstorm, water spout, cyclone, marine heat waves and, storm are frequently reported along with India's coastal regions. According to recent research, Indian Ocean is the warmest among all five oceans, generating 7% of the total world's cyclones (Gupta et al. 2019). Climatic change leads to continuous warming of the Indian Ocean, resulting

in an increased number of severe cyclones on the east and west coasts of the Indian sub-continent (Sarkar 2020). The highest number of cyclones has been reported in the year 2020 ever since 1976 (Kambli 2020). The growing proposition of tropical cyclones, heating up of the atmosphere due to carbon dioxide emission and rising sea level due to global warming are the major causes of storms. The latest research shows that the strongest storm's proposition increases about 8% a decade (Kossin et al. 2020). Nearly eighteen storms of the Indian Ocean in 2020 with wind speed greater than 65 km/h have been reported by Accu Weather climatic web portal. Similarly, 341 weather stations reported extremely heavy rainfall, measuring above 20 cm in 2020, as compared to 554 stations in 2019; and 321 and 261 in 2018 and 2017, respectively. Very heavy rainfall in 2020, estimated between 11 and 20 cm, was recorded by 2,253 weather stations, as compared to 3,056 in 2019; and 2,181 and 1,824 in 2018 and 2017, respectively.

National Institute of Oceanography, suspects the theory and formation of water spouts. Similar spouts are reported over Nazare dam's water body in Pune, India, in 2018 (Khelkar 2018) and Kakinada and Yanam, Eastern coasts of India in 2020 (Naidu 2020). The Intergovernmental Panel on Climatic Change (IPCC) published a special report by discussing global warming on oceans which has a significant message for India (Koll and Murtugudde 2019). The report warns India about marine heat waves, which causes severe damage to marine sea lives and corals. Due to these heat waves, aquaculture industries along the Indian Ocean-rim have suffered severe damage in recent years. The causes and effects of these weather phenomena have led to incredible damage; hence, the research among the Indian Ocean is gaining consideration day by day. This research area focuses on weather phenomena and weather attributes related to those phenomena in and around the coastal regions of Indian Ocean to provide a better understanding of data for researchers, meteorologists and, other end-users. The ocean data analysis leads to many applications like marine safety, weather forecasting, fishing, aquatic lives, disasters and the rest.

Literature survey

The government and private sectors are increasingly getting committed to transparently managing all information regarding satellite data. This leads to face a lot of challenges and opportunities caused by a vast amount of datasets (big data) that are made available on web. In recent years, the Open Data (OD) approach initiated by W3C has received increasing attention (Ma 2017). In big data the structural and semantic heterogeneity is identified to be a great concern among the researchers all over the world as it causes many problems in data extraction, aggregation and integration.

Thus organization of heterogeneous big data leads to an efficient big data query engines (Bansal and Kagemann 2014). By integrating big data with semantic web it provides a better way to utilize and add capacities to existing frameworks. A survey for challenges and opportunities in big data and semantic web is presented by Ahmed (Ahmed and Ahmed 2018). Various observation sensors are implemented across the ocean which records the values of weather parameters in a successive interval of time which results in big data. The satellite data records consist of geospatial information like latitude, longitude, date, and time.

The provided meteorological information includes the values of ocean weather parameters like wind speed, sea surface temperature, air pressure, relative humidity, wind gust and other parameters in heterogeneous file formats. These data are usually collected and integrated from distinctive data sources that may comprise of structured, semi-structured and unstructured data (Bansal and Kagemann 2015; Ma et al. 2007) which is usually more than 85% (Bansal 2014). To establish data uniformity, interoperability and heterogeneity have to be addressed by presenting the data in a standard machine-understandable format like RDF. RDF has been widely accepted file format and has rapidly gained popularity in recent years. It helps to represent and share data in many application domains (Zhang et al. 2021). Atemezing (Atemezing et al. 2011) proposes a conversion approach through transforming meteorological data into RDF deploying python scripts. This work employs the data collected from the meteorological stations over Spain and published as Linked Data (LD) which supports modularity and reusability. Linked data is the practice of inter-connecting the assets through publishing, sharing and linking the domain's data that allows the sharing and re-use of scientific data (Pouchard et al. 2013).The linked open data feature of any data repositories allows interlinking of concepts both within and across the organizations and web sites (Wilson et al. 2015).

A research report for representation of meteorological data as RDF has been carried out in Irstea, France, for weather predictions in agricultural decisions (Catherine et al. 2014). This report provides information about the usage of weather ontologies associated with the generation and publication of LD from different weather stations. The lack of this work is mentioned that it provides difficulties in decision making about the instances or properties of a specific measurement of any parameter. Geospatial data integration is carried out by extracting geographical data from web and identifying the features and constructing a schema and instances using RDF (Cruz et al. 2013). Similarly, semantic web approach helps in extracting the unstructured geospatial data and transforming it into RDF, linking and integrating from heterogeneous sources (Zhang et al. 2013). However, the transformation of unstructured data file into RDF is still

rare in research. Semantic web helps in transforming different data that are aggregated from various sources, into useful information.

Ontologies are used to express geospatial information that has much heterogeneity. A platform is necessary to represent the knowledge of a particular domain in which one works. Accessing domain resources on web is difficult due to the heterogeneous problem. Ontologies are used to solve this problem and integrate the resources successfully (Xiong et al. 2014). For the problem of big data integration, representation and aggregation a semantic web based architecture has been proposed by providing a solution for heterogeneity (Saber et al. 2018). Data can be aggregated from various resources without any assumptions provides the retrieval of data in the way, place or time through combined data with semantic concepts (Gollapudi 2015). Big data semantic model can be processed through mapreduce framework to store the data semantically and overcome the problem of understanding, big data aggregating, linking, integrating and representing between heterogeneous data systems (Kang et al. 2014). But this system does not provide the process of data integration from existing databases.

Researchers have carried out a few works to present the content and structure of ontology creation to help developers build the DOs. Rudnicki (Rudnicki 2019) has developed Common Core Ontologies (CCO) that comprises eleven different ontologies by integrating the classes and relations among all interest domains. Although the developed CCO ontology provides interoperability and reduces the cost associated with enterprise information, it lacks ontology quality information. Some climatic ontology is created that reports about the prediction of solar irradiance (Kantamneni and Brown 2018). The proposed forecast model of solar irradiance has been validated and proved a high rate of completeness and accuracy but lacks predicting other different phenomena affected by the same weather parameters. Sensor observation for understanding the blizzard weather event has been modeled by developing ontology using the Canadian Climate Archives (CCA) datasets (Devaraju and Kauppinen 2012). This work exploits ontology vocabularies with a rule-based technique to represent the weather event's relations and detected properties. Ontology-based data access and integration have been developed using the weather datasets for farming in Nepal by converting them into RDF for data usage and knowledge retrieval (Pokharel et al. 2014). This work does not support non-experts to access the dataset to incorporate different and additional query cases using SPARQL.

Ontologies contribute to developing new technology for forecasting applications, enabling and supporting meteorological Decision Support System (DSS). A study has been reported by Bally (Bally et al. 2004) that presents a basic understanding of the existing weather forecasting systems

and their technologies supporting DSS. A study has been carried out to describe a method for representing the geoscience forecast data into ontology with existing metadata information (Chen and Plale 2013). This method represents the dataset with added semantics and other functionalities of data compared to the existing representation. Apart from manual construction, some researchers propose a semi-automatic way of developing ontology (Kaladevi et al. 2020). With all the available concepts, relations and attributes, the existing ontology is enhanced and extended using domain knowledge. With conceptual clustering, the concepts of weather ontology are collected based on their semantic data similarities to construct hierarchy. The background ontology has been developed for the weather domain using related knowledge sources and expert knowledge. This approach improves data retrieval and reduces search time.

Owing to the increasing popularity of the semantic web, researchers rely on measuring the quality of various aspects such as linked data, ontology, inference engine, data backing and user interface. Even though the ontologies are designed for a particular domain, determining their quality is challenging, including multiple works like the fact, quality of datasets, quality of search engine and quality of inference engine, etc. Research works have also been carried out in developing some software for ontology matching algorithms by considering the vocabularies of the ontology designed for a particular domain. Although some algorithms are proposed, it still lacks in terms of efficiency and accuracy. A novel iterative framework like RiMOM-Instance Matching (RiMOM-IM) has been proposed for matching instances in the ontology by discovering the corresponding instances in the knowledge base (Shao et al. 2019). RiMOM-IM considers a source knowledge base to the particular domain and matches the target knowledge base's instances to find the exact matching of ontology. Another framework named Data Mining for Ontology Matching (DMOM) based instances compares the instances and the data properties that have been matched and identified efficiently (Belhadi et al. 2020). Three stages have been examined in this work: exhaustive, statistical and Frequent Itemsets Mining (FIM) using the DBpedia ontology. DMOM is experimented with and efficient in factors such as execution time and quality of the matching process.

Similarly, a pattern-matching algorithm has been designed to solve ontology matching problems using a pattern mining approach (Belhadi et al. 2019). This method searches for the redundant patterns in the ontology database and matches the target ontology's relevant feature to find the ontology's efficient matching. Among all the above works, it is clear that a developed ontology's quality can be determined only by comparing the target ontology with the source ontology. Hence, the proposed method considers GS (Golden Standard) ontology to compare it with

the developed OWO ontology to measure its performance parameters.

Methodology

This section depicts the proposed semantic web-based weather data model to access, manipulate, store and provide knowledge of copious satellite data to end-users for coastal applications. This research aims to build ocean knowledge-base through ontology by providing suitable vocabularies related to the satellite data collected from various meteorological sources. The proposed method consists of four phases: Data integration, Knowledge representation, Semantic web processing, and Semantic query engine as represented in Fig. 1. The ocean field area is monitored by the sensors like Agro Floats, Buoys, Coastal Radars, Gliders and Sonde and others. Each sensor traces the information recorded in a successive interval of time and stores it in heterogeneous file formats. Sensor data is the major source

for any weather-related researches considered by various researchers and domain experts. The geospatial ocean data is aggregated from various bureaus like World Weather Online (WWO), Indian Meteorological Department (IMD), Accu Weather and similar web portals.

Data integration

Open Government Data (OGD) offers a huge anthology of real-time and catalogued datasets through distributed server websites from a variety of environmental information resources. The provided resources have been suffering from lack of uniformity, data interoperability, and data interpretation. The geospatial climatic data is aggregated from various bureaus is said to face two major issues: heterogeneous files and heterogeneous vocabularies, as shown in Fig. 2. The heterogeneity in sensor data makes it difficult to exchange, share and reuse the datasets. Comma Separated Values (CSV, *.csv), Totals file (TUV, *.tuv), Excel (*.xls), Network Common Data

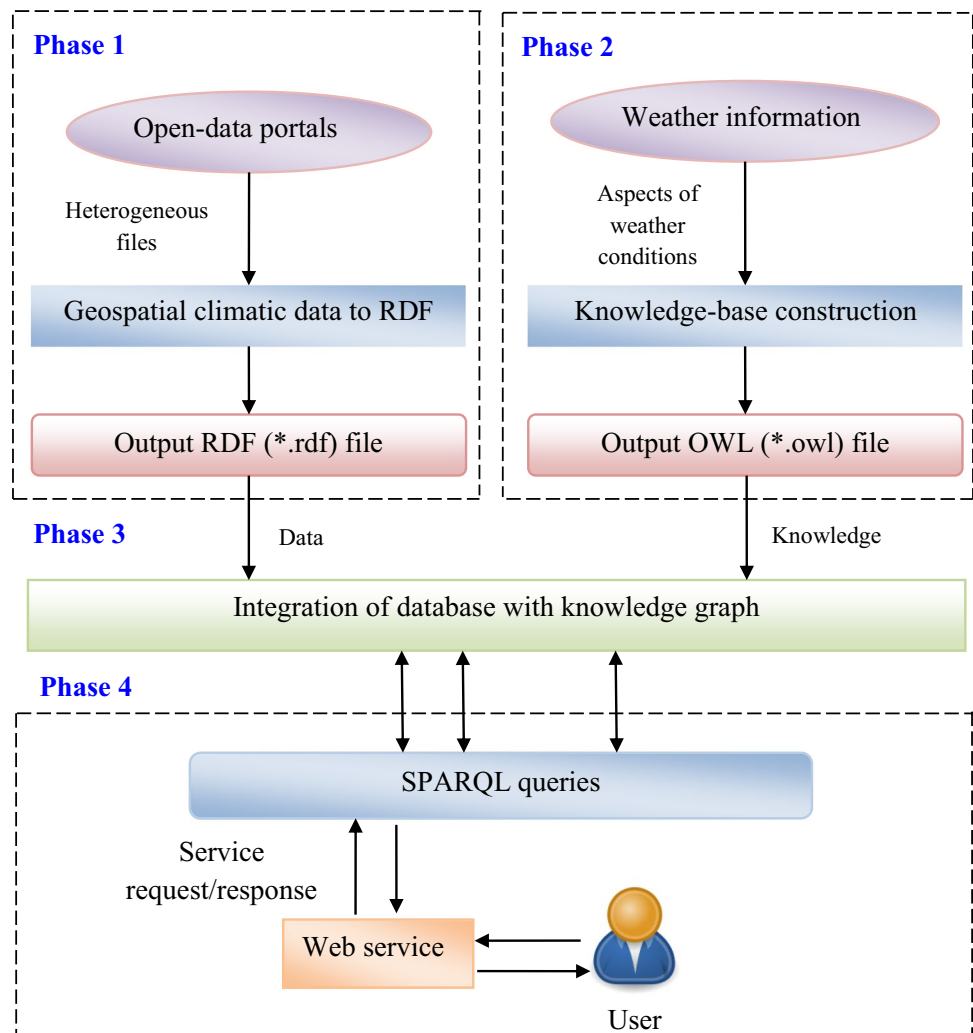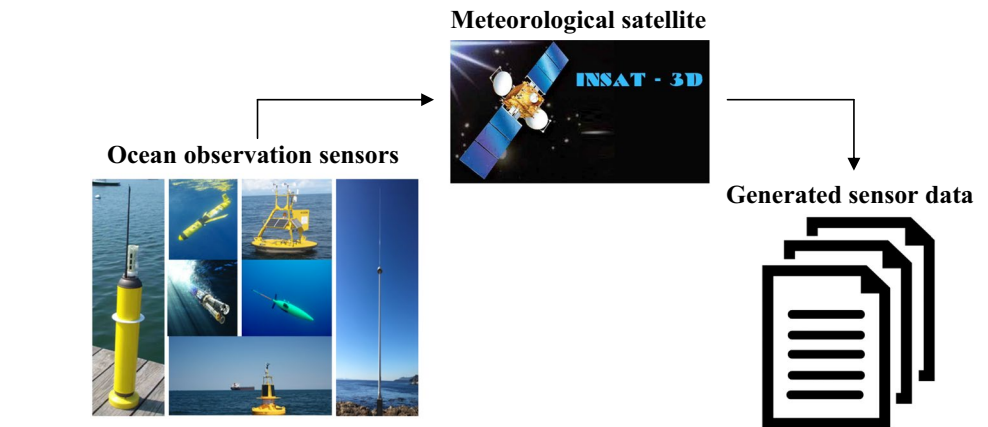**Fig. 1** Proposed approach: semantic web based satellite data integration

Fig. 2 Issues in sensor-generated ocean data

Meteorological satellite

Ocean observation sensors

Generated sensor data

Issues

- **Heterogeneous formats** (e.g, CSV, NetCDF, ASCII, Excel, TUV, HDF5 etc.)
- **Heterogeneous vocabulary** (e.g, Temp, Surface temperature, SST, Temperature, Sea surface temperature etc.)

Form (NetCDF, *.nc) and Hierarchical Data Format (HDF5, *.hdf) are some of the heterogeneous sensor data file formats. These formats are said to be semantically quite heterogeneous and leaves many ambiguities open, which makes explicating, balancing and visualising the available data difficult. This motivated a set of researchers and developers to provide a new representation of data which is understandable and usable by web agents.

The proposed research work includes four different sensor data files namely CSV (*.csv), Excel (*.xls), TUV (*.tuv) and NetCDF (*.nc). The data representation for each file format is different hence; these files are converted in to a standard machine understandable format namely RDF. The function of CSV file F_{csv} is represented as per Eq. 1 where R_a= rows, C_a= columns and S_a= special characters; Eq. 2 presents the function of Excel file F_{xls} where R_b= rows and C_b= cell; the totals file F_{tuv} is represented as given in Eq. 3 where F_{uv}= file information, H_{uv}= header and D_{uv}= data and Eq. 4 illustrates the function of NetCDF F_{nc} where H_{xy}= header and D_{xy}= description.

$$F_{csv}(a) = \{R_a, C_a, S_a\} \tag{1}$$

$$F_{xls}(b) = \{R_b, C_b\} \tag{2}$$

$$F_{tuv}(u, v) = \{F_{uv}, H_{uv}, D_{uv}\} \tag{3}$$

$$F_{nc}(x, y) = \{H_{xy}, D_{xy}\} \tag{4}$$

Semantic data integration enables blending of data together from disparate sources by employing a data centric architecture built upon RDF model. Semantic web has the ability to easily import and harmonize heterogeneous data from multiple sources and interlink it as Linked Data (LD) namely, RDF statements into an RDF triple store. Semantic web based data integration plays a vital role in the field of many knowledge management solutions, where the raw data (satellite data) is transformed into a machine-understandable format in order to facilitate efficient semantic retrieval. Heterogeneity among the sensor data have been addressed by using RDF, a data-modeling framework for semantic web technology.

The data integration layer converts the heterogeneous satellite data into graphical format called RDF to make them suitable for semantic processing. The conversion process is carried out using Apache Jena API, which is a
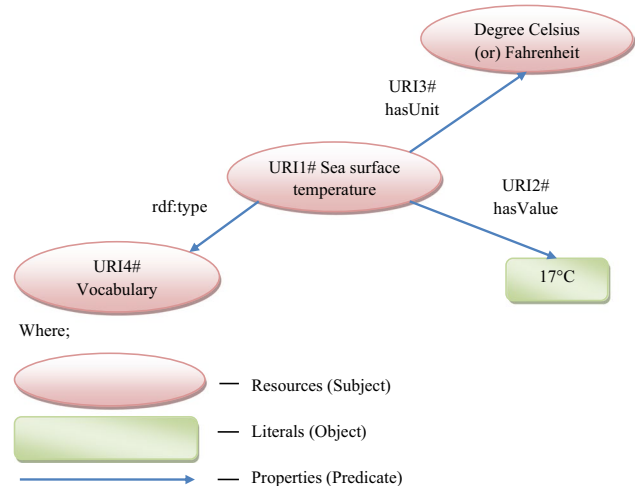
Fig. 3 Graphical representation of RDF

java framework used to build linked data for semantic web applications by transforming the flat-file controlled vocabulary into a standard RDF format. Apache Jena affords an Application Program Interface that extracts the data and presents it as RDF. Generally RDF is represented as an abstract "model". A model can be a basis with data files formats, databases, URLs or intermingle of all the three. Each record in the heterogeneous files are retrieved and written as RDF statements using Jena API through assigning a subject, predicate and an object. The "model.write" is an RDF writer API which generates the RDF resources providing a meaning for each data to address interpretation.

RDF data model defines the structure of RDF language (World Wide Web Consortium 2004). The basic RDF data model consists of three object types: (1) Resources – all data objects described by a RDF statement, (2) Properties – a specific aspect, characteristic or relation of a resource and (3) Statements – a statement that combines a resource with its describing property and the value of the property (Bonstrom et al. 2003). An RDF statement is typically expressed as "*resource → property → value*"– a triple and is commonly written as $P(R, V)$ where; a resource R has a property P with a value V. The resources and properties are expressed using statements (triples:$< s, p, o >$) that consists of three parts namely; subject (s), predicate (p) and object (o). RDF uses URI to identify the resources and properties, in case a resource does not have an identifier it is referred to as a blank node. For example a simple statement "Sea surface temperature is 17 °C" is represented as an RDF statement as per the Eq. 5.

$$< URI1\#Sea\ surface\ temperature >< URI2\#hasValue >< URI3\#17°C > \quad (5)$$

In the above RDF statement $\varepsilon < URI1\#Sea\ surface\ temperature > \varepsilon$ represents the subject, $\varepsilon < URI2\#hasValue > \varepsilon$ indicates the predicate and $\varepsilon < URI3\#17°C > \varepsilon$ refers to the object. Figure 3 depicts an example RDF graph for sea surface temperature where "hasUnit" is an object property which gives a relation between "Sea surface temperature" and "Degree Celsius". On the other hand "Sea surface temperature" consists of a data property "hasValue" which links to the literal value "17^^xsd:float" of data type xsd:float defined in XML schema. RDF defines a predicate type called "rdf:type" which refers to indicate that the thing is of certain type. For example, "Sea surface temperature" is of the type "vocabulary" which refers to the vocabulary of an ocean weather parameter.

Similarly, the weather parameter vocabularies of different observation sensors are recorded in different naming formats such as temp, temperature, sea surface temperature, SST, surface temperature, and others. During RDF conversion process, the parameters in heterogeneous files are represented in RDF using a standard naming convention for ocean vocabularies, such as sea_surface_temperature. The sensor data vocabularies from different data sources are integrated into a standard parameter vocabulary as depicted in Fig. 4. The Integrated Ocean Observing System (IOOS) standard vocabulary format represents the collected ocean data into RDF to overcome this issue. IOOS is a standard
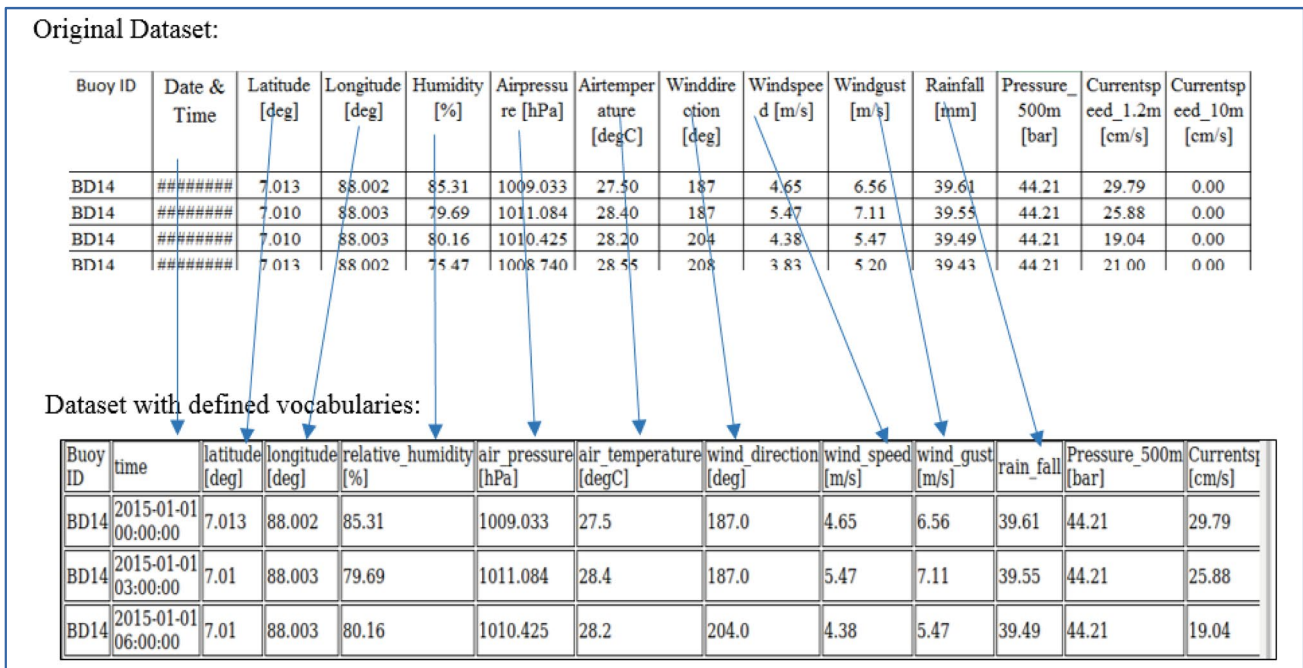

Original Dataset:

Buoy ID	Date & Time	Latitude [deg]	Longitude [deg]	Humidity [%]	Airpressure [hPa]	Airtemperature [degC]	Winddirection [deg]	Windspeed [m/s]	Windgust [m/s]	Rainfall [mm]	Pressure_500m [bar]	Currentspeed_1.2m [cm/s]	Currentspeed_10m [cm/s]
BD14	########	7.013	88.002	85.31	1009.033	27.50	187	4.65	6.56	39.61	44.21	29.79	0.00
BD14	########	7.010	88.003	79.69	1011.084	28.40	187	5.47	7.11	39.55	44.21	25.88	0.00
BD14	########	7.010	88.003	80.16	1010.425	28.20	204	4.38	5.47	39.49	44.21	19.04	0.00
BD14	########	7.013	88.002	75.47	1008.740	28.55	208	3.83	5.20	39.43	44.21	21.00	0.00

Dataset with defined vocabularies:

Buoy ID	time	latitude [deg]	longitude [deg]	relative_humidity [%]	air_pressure [hPa]	air_temperature [degC]	wind_direction [deg]	wind_speed [m/s]	wind_gust [m/s]	rain_fall	Pressure_500m [bar]	Currents [cm/s]
BD14	2015-01-01 00:00:00	7.013	88.002	85.31	1009.033	27.5	187.0	4.65	6.56	39.61	44.21	29.79
BD14	2015-01-01 03:00:00	7.01	88.003	79.69	1011.084	28.4	187.0	5.47	7.11	39.55	44.21	25.88
BD14	2015-01-01 06:00:00	7.01	88.003	80.16	1010.425	28.2	204.0	4.38	5.47	39.49	44.21	19.04

Fig. 4 Sensor dataset represented using IOOS standard vocabularies

RDF vocabulary (e.g., IOOS Parameter Vocabulary, http://mmisw.org/ont/ioos/parameter) created to assist interoperability between the data catalogues published on the web (Haines et al. 2012). It allows any publisher to illustrate the datasets and services in an index using a standard vocabulary model that enables metadata consumption from numerous indexes. This can improve the recovery of data services and datasets from various sources by providing a user-friendly environment. It also facilitates the efficient search for datasets throughout different websites' catalogues using a similar query mechanism and structure.

Knowledge representation

The core terminology for semantic web representation of data is ontology construction which represents the knowledge of data. The term otology is introduced by semantic web technology which aims to establish meaning of the data such that it can be shared, reasoned, and reused through machine-readable applications. Interoperability among data is achieved by ontological representation, which addresses structural, syntactic and semantic heterogeneity. Ontology provides a common language for representing how data relates to the real world objects by allowing a person or a machine to understand the set of databases which are connected by being the same thing. Ontological representation of data is a vision of information that can be interpreted by machines, so that it can perform more of the tedious work involved in finding, combining, and acting upon information on the web. It enables the machine to understand and respond to the complex human requests based on the meaning of the data.

Ontology is the best technology to accomplish semantic concept-based data retrieval, which provides a meaningful representation of data. Ontology is built by describing the concepts and relations among them to the application domain. Hence, it is a major step towards achieving semantic interoperability in Information Systems (IS). The comprehensive structure of ontology is a 5-tuple composition (Neches et al. 1991) as described in Eq. 6.

$$5 - \text{tuple } O : (C, H_C, R, H_R, I) \tag{6}$$

where; '*C*' represents a set of concepts (i.e., instances of "rdf:Class") which are arranged with a corresponding hierarchy '*H_C*'. '*R*' represents a set of relations that relates each concepts to one another (i.e., instances of "rdf:Property"). $R_i \in R$ and $R_i \rightarrow C_1 \times C_2$. '*$H_C$*' represents the concept hierarchy in the form of a relation $H_C \subseteq C_1 \times C_2$ (i.e., a relation corresponding to "rdfs:subClassOf"). where, $H_C \subseteq C_1 \times C_2$ denotes that '*C_2*' is a sub-concept of '*C_1*'. '*H_R*' represents a relationship hierarchy in the form of a relation $H_R \subseteq R_1 \times R_2$ (i.e., a relation corresponding to "rdfs:subPropertyOf"). where, $H_R \subseteq R_1 \times R_2$ denotes that '*R_2*' is a sub-relation of '*R_1*'. '*I*' is the representation of instance of the concepts in a particular domain (i.e., "rdf:type").

Second phase provides knowledge representation of satellite ocean data through conceptualization definition by building a new ontology named OWO. This ontology is built using DO type for oceanographic weather domain to define how the weather attributes are related to different weather conditions. Ontologies have basic profiles on which it has been built, namely OWL-Full, OWL-EL, OWL-QL, and OWL-RL. The proposed OWO ontology aims to define a syntactic subset described in a suitable rule-based engine that requires scalable reasoning. Hence, OWO is developed based on the OWL-RL profile of the ontology. OWL-RL is the best profile that is widely used for domain applications based on the rule engine. Developing a new ontology for an application domain can't value the full potential of existing domain relevant knowledge. Thus it tends to follow the FAIR principle in reusing the existing ontologies of the same domain. Ontology reuse can be defined in two different categories (1) building ontology by extending, specializing, assembling and adapting other ontologies and (2) building ontology by merging different ontologies of same subject domain into a single one that unites all of them. This paper builds an ocean weather ontology based on the help of two different weather ontologies "weather phenomenon prediction using semantic web" (Roy 2017) and "weather ontology for predictive control" (Staroch 2013). The namespaces of the existing ontologies are qb:structure, qb:DatatSet, owl, core, schema, xsd, dct, nc, CF_INTERNAL, float, int, long, short, double, byte, string, char, rdf, rdfs, NS_STRUCT_INTERNAL and NS_DATA_INTERNAL.

Ontologies play a significant role in following FAIR data principles, particularly in relation to provide support for interoperability and reusability (Poveda-Villaló et al. 2020). The data principles mostly indicates on (1) usage of vocabularies that follows FAIR (Findable, Accessible, Interoperable and Reusable) principles; where the proposed method uses a standard IOOS parameter vocabulary, (2) use of a formal, accessible, shared and broadly applicable language for representing the domain knowledge; the proposed method uses OWL language to represent the knowledge about ocean domain, (3) meet domain-relevant community standard; which represents the easy reuse of data sets by providing it in an organized and standard way with sustainable file formats namely RDF in a common vocabulary namely IOOS, and (4) the presented metadata is retrievable from a unique identifier; ontologies provide unique identifier for representing the concepts and

relations in application domain that are accessible through SPARQL queries.

The existing satellite data retrieval systems effectively execute the query based on location, time, date, sensor ID, satellite, weather attribute etc. In that case, retrieval of additional domain-specific concept-based data from copious information generated by satellite is challenging. For example, retrieving the values of wind speed can be queried using API but retrieving a specific knowledge-based data such as fresh gale, no wind or fresh breeze is a complex task. To facilitate an efficient retrieval system and to fulfill user requirements, the system should be designed in such a way to support semantic concept or knowledge-based satellite data retrieval (Bai et al. 2012). The attributes associated with weather conditions are collected from various weather-related websites to build ontology. Second phase provides an ontological representation of knowledge for each data source from the data integration phase. The proposed OWO ontology consists of 37 concepts, 112 instances, 85 relations and 126 attributes. Some major categories of the developed OWO ontology are explained as illustrated in Table 1.

Weather concepts namely wind_speed, precipitation, relative_humidity, sea_surface_temperature, conductivity and aerosol_optical_thickness etc., are related to each other by incorporating 85 different relational links of different data types. The relations are: is-a, has-a, hasLong, hasLat, hasInterval, hasAttribute, hasCondition, and the rest. Each concept consists of a number of instances that hold a range of values of different data types through a data property. The concepts are related to sub-concepts or individuals by incorporating object properties. The range of values for instances of each concept namely wind_speed (Beaufort and Beaufort scale, n.d.), precipitation (Engineering Tool Box n.d.), barometric_pressure (Haby 2014), humidity (Measurement of Precipitation 2018) and sea_surface_temperature (Shenoi et al. 2009) and so on are collected from various weather prediction analysis reports. For instance, the data property,

measurement unit and value type of the concept weather_attributes are illustrated in Table 2.

Each sub-concept is further categorized into several instances/individuals according to a range of data values to the specific concept. For example, the individuals of the concept "wind speed" are represented, as shown in Table 3 with the range of data values. By incorporating these described specifications, the proposed OWO ontology is designed using the protégé tool. Apart from weather attributes the proposed OWO ontology includes eight different ocean weather phenomena namely; storm, cyclone, water spout, rainstorm, heat dome, humid weather, thunderstorm and marine heat waves. The weather attributes related to the phenomenon, along with the values are collected from the reports of various meteorologists, scientists etc., and included in the proposed ontology. The details of the weather attributes involved in various ocean weather phenomena are represented as illustrated in Table 4.

Ontologies developed and used in online systems are larger; hence a database is a mandatory for storage and efficient and optimal utilization (Morsey et al. 2012; Stegmaier et al. 2009). Research works have been carried out for developing a tool for converting ontology into relational tables (Zidan et al. 2019). Relational databases support performance, robustness, reliability, availability, legacy data, legacy applications and large scale ontologies. In this paper H2 DataBase (H2DB) console is considered a relational database management system is written in java. This database is developed and tested on Linux OS (Ubuntu16.04 version) using java (jdk 1.8.0_181). H2DB is preferred over other databases since it is extremely fast, open-source, contains a scrollable result set and browser-based application. It supports web server, Transmission Control Protocol (TCP) and JDBCAPI, which connects it to the ontology. A relational database is created in H2 Engine that depicts the internal relations between the sub-concepts developed in the OWO ontology. The developed OWO ontology is mapped to the database through an ontop mapping manager. Mappings are hypotheses that are used to relate the data in RDBMs to the vocabulary of ontology. Hence, the mapped ontology

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

S.No	Top-level concepts	Respective sub-concepts
1	GeoLocation	latitude, longitude
2	Date	instances, value, year
3	Time	interval, hours, minutes, seconds
4	WeatherAttributes	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, lightning, thunder, dew_point
5	WeatherCondition	good, bad, severe
6	WeatherPhenomenon	cyclone, storm, water_spout, marine_heat_waves, thunderstorm, humid_weather, heat_dome, rainstorm

Table 2 Relations and data properties of weather_attributes

S.No	Concept name	Data property name	Data type	Unit
1	precipitation_rate	hasPrecipitation	xsd:decimal	mm/h
2	relative_humidity	hasHumidity	xsd:decimal	%
3	wind_speed	hasWindSpeed	xsd:decimal	m/s
4	cloud_cover	hasCloudCover	xsd:integer	%
5	barometric_pressure	hasPressure	xsd:integer	hPa
6	sea_surface_temperature	hasSurfaceTemp	xsd:float	℃
7	aerosol_optical_thickness	hasOpticalThickness	xsd:decimal	μm
8	fog	hasFogIndex	xsd:decimal	mm/μm
9	conductivity	hasConductivity	xsd:decimal	S/mho/ cm
10	air_temperature	hasTemperature	xsd:float	℃
11	wind_from_direction, wind_to_direction	hasDirection	xsd:decimal	deg
12	air_pressure	hasAirPressure	xsd:integer	hPa
13	wind_gust	hasWindGust	xsd:decimal	m/hr
14	solar_irradiance	hasRadiance	xsd:decimal	W/m^2
15	uv_index	hasUVIndex	xsd:integer	Units
16	lightening	hasLighteningPresence	xsd:boolean	Y/N
17	thunder	hasThunderOccurrence	xsd:boolean	Y/N
18	dew_point	hasDewPoint	xsd:float	℃

Table 3 Instances of wind speed with the range of data values

S.No	Sub-concept	Individuals	Data property	Data value
1	wind_speed	NoWind	hasWindSpeed	[< 0.3]
2	wind_speed	LightAir	hasWindSpeed	[> = 0.3, < = 0.5]
3	wind_speed	LightBreeze	hasWindSpeed	[> = 1.6, < = 3.3]
4	wind_speed	GentleBreeze	hasWindSpeed	[> = 3.4, < = 5.5]
5	wind_speed	ModerateBreeze	hasWindSpeed	[> = 5.6, < = 7.9]
6	wind_speed	FreshBreeze	hasWindSpeed	[> = 8.0, < = 10.7]
7	wind_speed	StrongBreeze	hasWindSpeed	[> = 10.8, < = 13.8]
8	wind_speed	ModerateGale	hasWindSpeed	[> = 13.9, < = 17.1]
9	wind_speed	NearGale	hasWindSpeed	[> = 13.9, < = 17.1]
10	wind_speed	FreshGale	hasWindSpeed	[> = 17.2, < = 20.7]
11	wind_speed	StrongGale	hasWindSpeed	[> = 20.8, < = 24.4]
12	wind_speed	WholeGale	hasWindSpeed	[> = 24.5, < = 28.4]
13	wind_speed	Storm	hasWindSpeed	[> = 24.5, < = 28.4]
14	wind_speed	ViolentStorm	hasWindSpeed	[> = 28.5, < = 32.6]

Table 4 Weather attributes involved in each phenomenon

S.No	Weather Phenomenon	Weather Attributes
1	Cyclone	barometric_pressure, relative_humidity, wind_speed, precipitation_rate, sea_surface_temperature
2	Storm	barometric_pressure, precipitation_rate, wind_speed, cloud_cover
3	Water spout	relative_humidity, wind_speed, air_temperature, water_temperature
4	Marine heat waves	relative_humidity, sea_surface_temperature, water_temperature_100m, water_temperature_200m, cloud_cover, precipitation_rate
5	Rainstorm	cloud_cover, relative_humidity, precipitation_rate, thunder, lightening, dew_point
6	Heat dome	uv_index, barometric_pressure, solar_irradiance, sea_surface_temperature
7	Humid weather	relative_humidity, air_temperature, dew_point
8	Thunderstorm	wind_speed, precipitation_rate, cloud_cover, lightening, thunder

makes it easier for the user to retrieve the data through the query engine.

Semantic processing and query engine

The semantic web layer is directly interacted with the domain ontology through Jena API in the third phase. There are two categories of information in logic terminologies: T-Box contains axioms defining classes and relations, and A-Box contains assertions about individuals in the domain. Generally, RDFS based ontologies like RDFS, OWL, DAML + OIL don't distinguish between these categories where the terminology and instance data can be freely mixed. In this paper, the OWL document does not directly import the relevant information; hence, the capability of reasoner architecture is used to bind T-Box separately and A-Box data sources separately. The reasoner is associated with the T-Box (classes and properties), then applied to A-Box (instances). This can be done by defining a separate model factory for both the cases to hold the data then the reasoner is created to use the declarations. Finally, the ontology model specification, including the reasoner, is designed and used to build an ontology model with A-Box as a base model.

Jena's model factory can create an inference graph by connecting datasets with a reasoner; it also supports a general-purpose rule engine. The reasoner's main aim is to answer the queries by transforming them into queries over the source. This project implements an ontop reasoner to answer a particular application domain by considering the developed OWO ontology as an input. The ontop reasoner is always connected to the source, and the data is not duplicated, which is up-to-date, as shown in Fig. 5. Considering the developed OWO ontology and mapping as input reasoner provides the answer for particular application domain. This platform allows any user to uniformly access the data stored in heterogeneous sources, which are renovated into RDF. The data is said to be incomplete. Hence, the conclusion cannot be drawn until it extends the data with that particular domain's knowledge. This can be achieved through an ontology that is developed and made available to the end-users. Then the system connects the concepts in the ontology with the sources. This platform uses mappings to allow any user to access the data from multiple resources through a single interface. The OWO ontology and the mapping present a virtual RDF graph, which can be queried by SPARQL, a standard query language for semantic web communities.

The final phase grants information retrieval through SPARQL queries for any weather-related applications. A typical query of OBDA is generally expressive, meaning it describes the user's desire instead of training the system how to answer. This allows the query to be independent of the data source and uniform access to heterogeneous sources. Using the developed ontology, the system finds out and executes the necessary query for any end users as per their requirement. 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) and others. But the most commonly used query language for ontology is SPARQL (Peng et al. 2016). The concept of writing a SPARQL query is to match its triples with the RDF triples and retrieve the queried information. The user can access more information by querying an integrated database and its relations built and saved in the ontology. Even though various environments are available for developing ontology, evaluating the developed ontology's quality is still a challenging issue. Some performance parameters are discussed in Section 3.4 using which the quality of developed OWO ontology is evaluated.
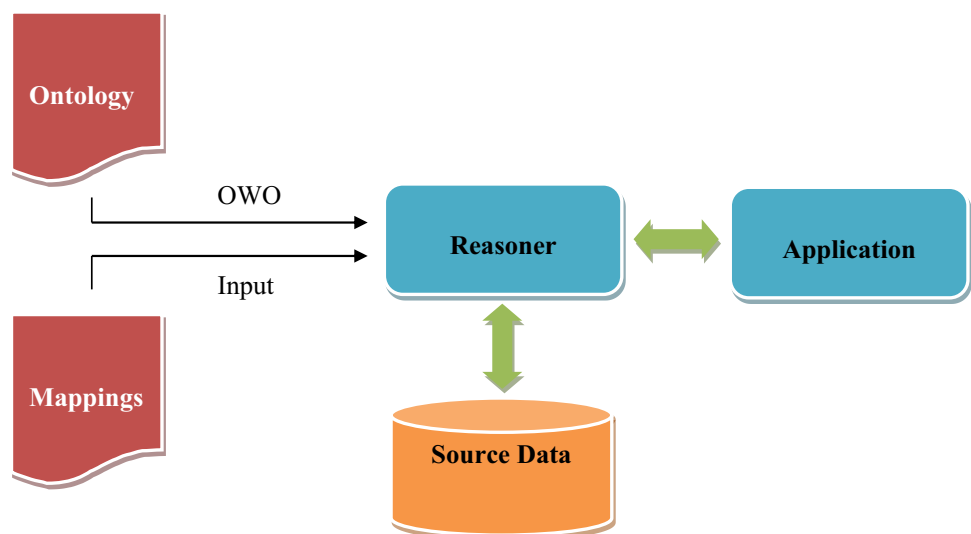

Fig. 5 Ontology-Based Data Access

Table 5 Characteristics of input datasets

Dataset (File formats)	No of files	Total Size (MB)	Total no of records
CSV (*.csv)	45	128.42	312,431
Excel (*.xls)	39	167.89	282,028
Totals (*.tuv)	41	137.18	324,712
NetCDF (*.nc)	47	213.24	358,374

Performance metrics of ontology

The quality of developed ontology has been evaluated by calculating some performance metrics, as depicted in this paper. The software quality check can be done based on two models: hierarchical and relational (Gillies 1997). This paper follows the hierarchical model to evaluate the analysis of the proposed ontology. Very little research has gone through evaluating the parametric measures of ontology; a survey has been proposed by Raad & Cruz (Raad and Cruz 2015) based on the evaluation methods. The quality of any ontology is evaluated against another ontology called GS ontology (Zavitsanos et al. 2011). Hong Zhu (Zu et al. 2017) proposed some performance parameters to evaluate the developed ontology based on the equations illustrated in this paper.

Definition 1: Model of ontologies

Let the developed weather ontology be O which holds a record (C, I, A, R). Where; C, I, A, R is a finite set of classes, instances, attributes, and relations defined in the ontology. $c \in C, a \in I^c$ and $\Psi \in A^c$ are the elements of each record. $R = \{r_1, r_2, r_{3,\ldots,}, r_n\}$ is n number of relations where each r defines a relation between the concepts c. Size of ontology $Size(O)$ is defined using Eq. 7.

$$Size(O) = Size_C(O) + Size_I(O) + Size_A(O) + Size_R(O) \quad (7)$$

The size of classes $Size_C(O)$, individuals $Size_I(O)$, attributes $Size_A(O)$ and relations $Size_R(O)$ are determined by the individual expression given in Eq. 8 using the mathematical operator ‖ (norm), which measures a linear map's size. In a linear map, mapping is represented between two modules; for instance, $X \to Y$ where X denotes the class and Y indicates sub-class.

$$Size_C(O) = \|C\|, Size_I(O) = \sum_{c \in C} \|I^C\|, Size_A(O) = \sum_{c \in C} \|A^C\|,$$
$$Size_R(O) = \sum_{r \in R} \|r\| \quad (8)$$

Definition 2: Vocabulary coverage

Let us define the GS ontology as ω. Vocabulary is the name defined to the classes, individuals, attributes, and
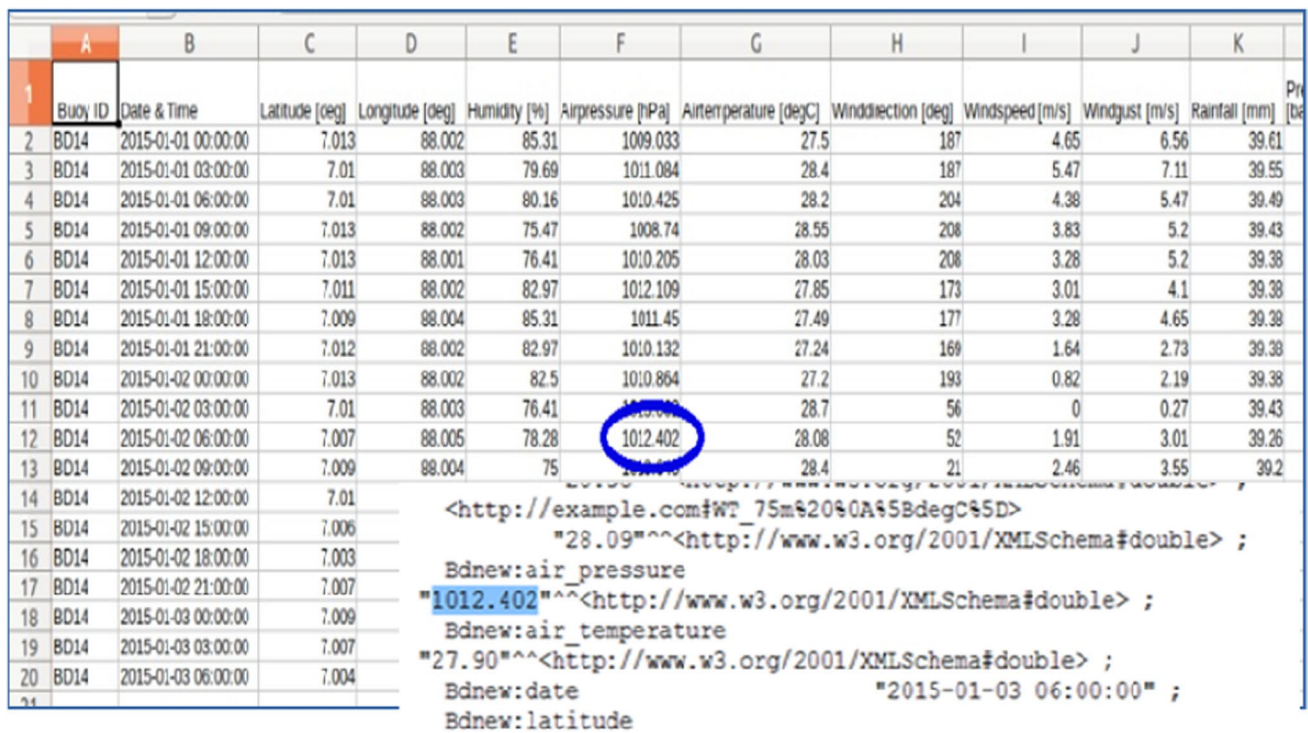
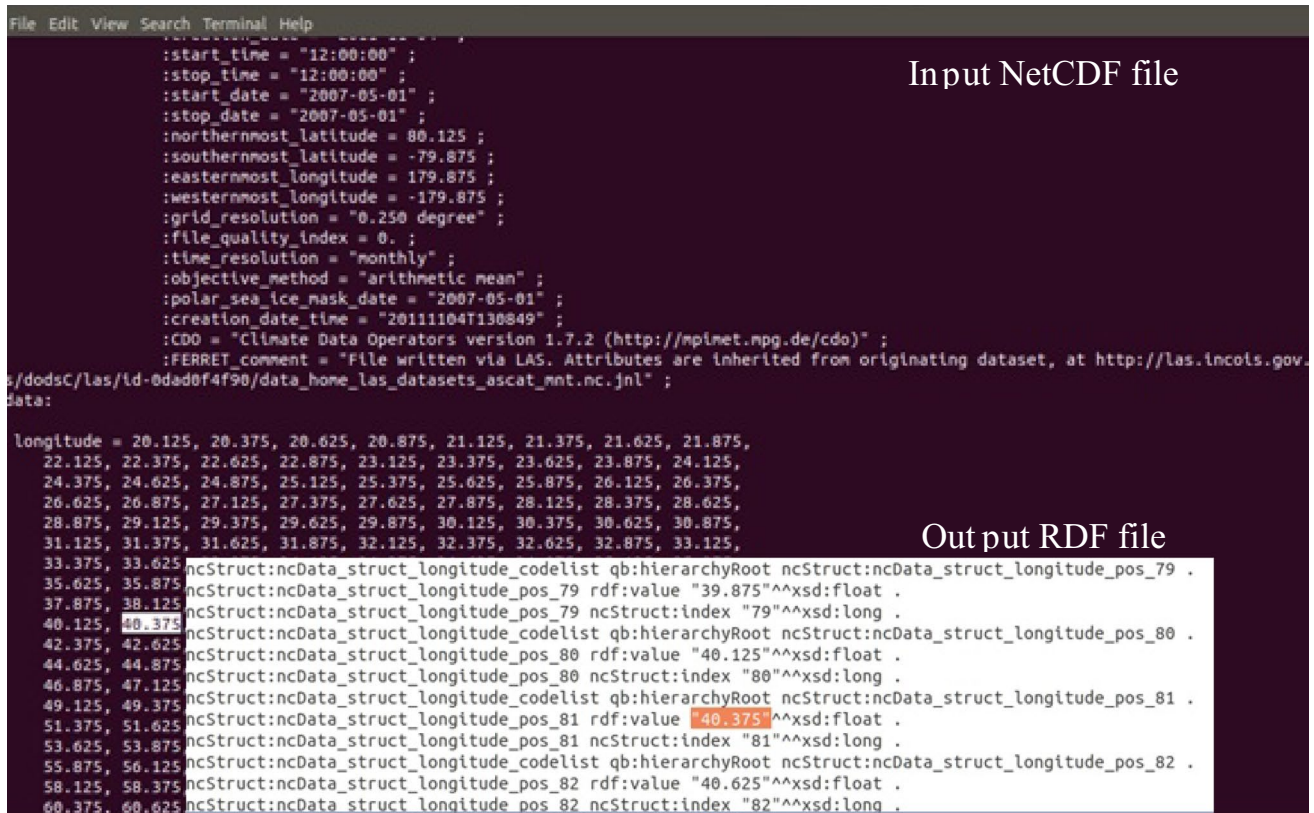**Fig. 6** Representation of CSV data file into RDF

Fig. 7 Conversion output of the NetCDF file into RDF

relations in the proposed ontology. Vocabulary coverage of ontology has to be calculated individually using the respective expressions in Eq. 9.

$$Cov_C^\omega(O) = \frac{\|C \cap C'\|}{Size_C(\omega)}, Cov_I^\omega(O) = \frac{\sum_{c \in C} \|I^c \cap I'^c\|}{Size_I(\omega)},$$
$$Cov_A^\omega(O) = \frac{\sum_{c \in C} \|A^c \cap A'^c\|}{Size_A(\omega)}, Cov_R^\omega(O) = \frac{\sum_{r \in R} \|r \cap r'\|}{Size_R(\omega)} \tag{9}$$

where, C', I', A' and r' are the classes, individuals, attributes, and relations of the GS ontology ω. The overall coverage of ontology is derived by using Eq. 10.

$$Cov^\omega(O) = \frac{\| C \cap C' \| + \sum_{c \in C} \| I^c \cap I'^c \| + \sum_{c \in C} \| A^c \cap A'^c \| + \sum_{r \in R} \| r \cap r' \|}{Size(\omega)} \tag{10}$$

Definition 3: Semantic coverage

Semantic coverage metrics $SCov_C^\omega, SCov_I^\omega(O), SCov_A^\omega(O)$ and $SCov_R^\omega(O)$ are evaluated similar to the above expressions, but the coverage is calculated, including the classes defined in ω which can be derivable from proposed ontology 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 class "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 evaluated using the expression mentioned in Eq. 11.

$$SCov^\omega(O) = \frac{D_C + D_I + D_A + D_R}{Size(\omega)} \tag{11}$$

Definition 4: Semantic compatibility

Ontology is said to be semantically compatible only if the contents are reliable to GS ontology ω. Compatibility metrics $RCC^\omega, ARCI^\omega, ARCA^\omega$ and $ARCR^\omega$ are evaluated for the proposed ontology by using the expression given in Eq. 12 and 13.

$$RCC^\omega(O) = \frac{\| c \in C \mid \omega \|}{Size_C(O)}, ARCR^\omega(O) = \frac{\sum_{r \in R} \| (x, y) \in r \mid \omega \|}{\| R \|} \tag{12}$$

$$ARCI^\omega(O) = \frac{\sum_{c \in C} \frac{\|a \in I^c \mid \omega\|}{\|I^c\|} \forall c}{Size_I(O)}, ARCA^\omega(O) = \frac{\sum_{c \in C} \frac{\|\Psi \in A^c \mid \omega\|}{\|A^c\|} \forall c}{Size_A(O)} \tag{13}$$
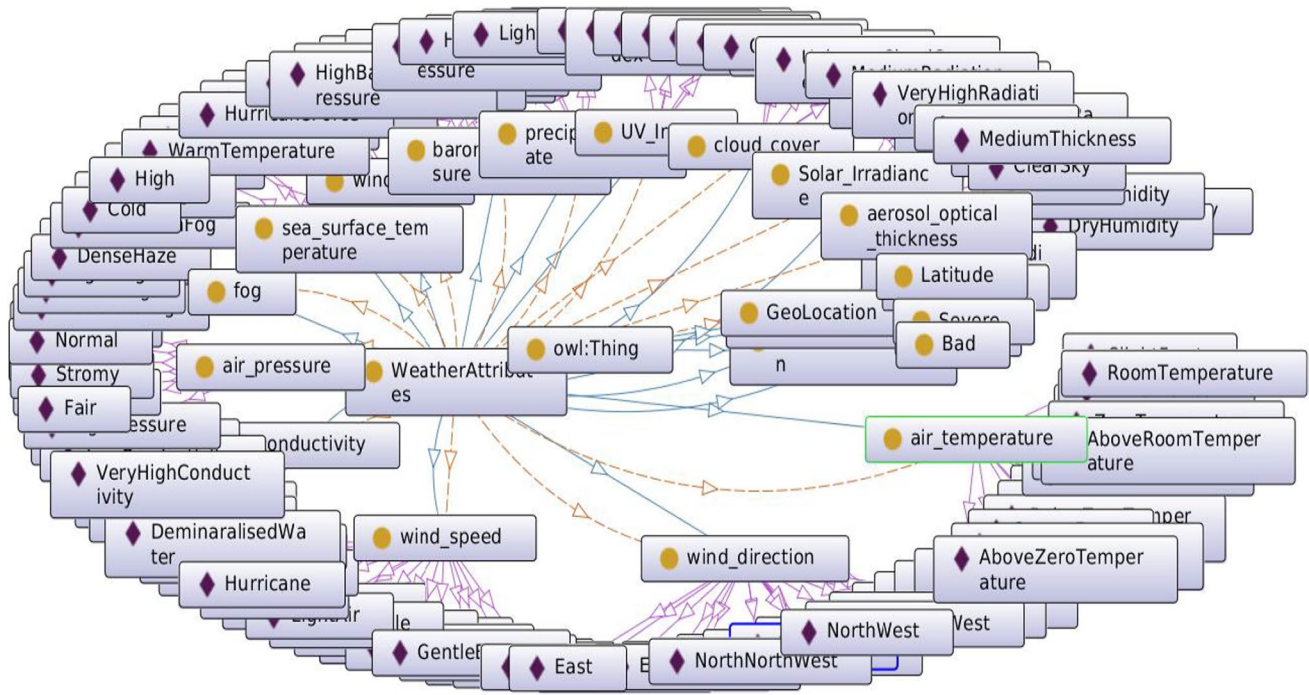
Fig. 8 The snippet of the onto-graph of proposed OWO ontology

Definition 5: Redundant elements

An element of ontology is said to be redundant if it can be derived from other elements. For instance, a concept is defined in the ontology, which can also be derived from other concepts; hence, the concept is redundant. Expressions in Eq. 14 are the redundant metrics of ontology.

$$CR(O) = \frac{\| C_R \|}{Size_C(O)}, IR(O) = \sum_{c \in C} \frac{\| I_R^C \|}{Size_I(O)}, AR(O) = \sum_{c \in C} \frac{\| A_R^C \|}{Size_A(O)},$$
$$RR(O) = \sum_{r' \in R'} \frac{\| r' \|}{Size_R(O)} \tag{14}$$

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

Definition 6: Cohesion (Relation based metrics)

The graph of any ontology is denoted by $G(O) = (N, E)$ 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 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. 15. Whereas the isolated nodes NIC specify the node that is not linked to any other node in the graph as expressed in Eq. 16.

$$NRN(O) = \| Rootnodes(O) \|, NLN(O) = \| Leafnodes(O) \| \tag{15}$$

$$NIC(O) = \| Rootnodes(O) \cap Leafnodes(O) \| \tag{16}$$

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 length of the ontology graph is denoted as $MaxSPL$ as expressed in Eq. 17.

$$MaxSPL(O) = Max p \in path(O) (Length(p)) \tag{17}$$

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. 18.

$$TNRNR(O) = \sum_{c \in Root(O)} \| Reachable^O(c) \|, ANRNR(O) = \frac{TNRNR(O)}{\| NRN(O) \|} \tag{18}$$

Definition 7: Cohesion (Metrics for Acyclic Relations)

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 Eq. 19 and 20.

Ocean data set in Excel (*.xls) file

	A	B	C	D	E	F	G	H	I	J	K	
1	Buoy ID	Date & Time	Latitude [deg]	Longitude [deg]	Humidity [%]	Airpressure [hPa]	Airtemperature [degC]	Winddirection	Windspeed [m/s]	Windgust [m/s]	Rainfall [mm]	F
2	BD14	1/1/2015 0:00	7.013	88.002	85.31	1009.033	27.5	187	4.65	6.56	39.61	
3	BD14	1/1/2015 3:00	7.01	88.003	79.69	1011.084	28.4	187	5.47	7.11	39.55	
4	BD14	1/1/2015 6:00	7.01	88.003	80.16	1010.425	28.2	204	4.38	5.47	39.49	
5	BD14	1/1/2015 9:00	7.013	88.002	75.47	1008.74	28.55	208	3.83	5.2	39.43	
6	BD14	1/1/2015 12:00	7.013	88.001	76.41	1010.205	28.03	208	3.28	5.2	39.38	
7	BD14	1/1/2015 15:00	7.011	88.002	82.97	1012.109	27.85	173	3.01	4.1	39.38	
8	BD14	1/1/2015 18:00	7.009	88.004	85.31	1011.45	27.49	177	3.28	4.65	39.38	
9	BD14	1/1/2015 21:00	7.012	88.002	82.97	1010.132	27.24	169	1.64	2.73	39.38	
10	BD14	1/2/2015 0:00	7.013	88.002	82.5	1010.864	27.2	193	0.82	2.19	39.38	
11	BD14	1/2/2015 3:00	7.01	88.003	76.41	1013.062	28.7	56	0.27	39.43		
12	BD14	1/2/2015 6:00	7.007	88.005	78.28	1012.402	28.08	52	1.91	3.01	39.26	
13	BD14	1/2/2015 9:00	7.009	88.004	75	1010.645	28.4	21	2.46	3.55	39.2	
14	BD14	1/2/2015 12:00	7.01	88.003	75.47	1011.523	28.02	25	4.38	6.02	39.08	
15	BD14	1/2/2015 15:00	7.006	88.005	81.09	1013.501	27.9	52	3.83	5.2	39.43	
16	BD14	1/2/2015 18:00	7.003	88.007	83.91	1013.501	27.61	72	3.83	5.47	39.43	
17	BD14	1/2/2015 21:00	7.007	88.004	84.84	1012.109	27.25	84	2.46	3.83	39.43	
18	BD14	1/3/2015 0:00	7.009	88.003	85.31	1012.842	27.29	83	2.19	3.01	39.49	
19	BD14	1/3/2015 3:00	7.007	88.005	81.56	1013.721	27.79	46	2.46	3.55	39.32	

Knowledge base weather ontology in protégé

Query results in SOS services

Fig. 9 Query output in Sensor Observation Services

$$ADLN(O) = \frac{\sum_{C \in Leaf(O)} Depth^O(C)}{NLN(O)}, MaxDepth(O) = \underset{c \in Leaf(O)}{Max} \left(Depth^O(C)\right) \tag{19}$$

$$AWNLN(O) = \frac{\sum_{C \notin Leaf(O)} Width^O(C)}{NAN(O) - NLN(O)}, MaxWidth(O) = \underset{c \notin Leaf(O)}{Max} \left(Width^O(C)\right) \tag{20}$$

Definition 8: Efficiency of information retrieval

The general purpose of semantic-based information retrieval system is to retrieve the relevant information based on the user query or context of user queries. The performance of an information retrieval system is carried out through the standard measures namely; precision, recall and F-measure. Precision of a system is evaluated by identifying the number of relevant data retrieved with respect to total number of data retrieved according to the user query. Precision refers to the ability of a system to screen out irrelevant information and calculated using the expression mentioned in Eq. 21. Recall is calculated using the number of relevant data retrieved for the user query with respect to the total number of the relevant data in the data base. Recall refers to the proportion of required data that are retrieved in a search and evaluated using the expression given in Eq. 22. In contrast F-measure is a harmonic average of precision and recall which is calculated using the expression presented in Eq. 23. Greater the precision leads to more relevant data that are retrieved through the search engine. On the other hand lower the recall rate represents the less coverage of concepts. An ontology-based semantic search is said to be efficient when the coverage of the concepts of particular domain is high which leads to higher recall rate.

$$Precision = \frac{No : of : relevant\ data\ retrieved}{No : of : total\ data\ retrieved} x100 \tag{21}$$

$$Recall = \frac{No : of : relevant\ data\ retrieved}{Total\ no : of : relevant\ data\ in\ the\ data\ base} x100 \tag{22}$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{(precision + recall)} \tag{23}$$

Results and discussions

Framework implementation

The proposed data model has been experimented on real-time satellite data collected along the south-eastern coastal areas of India. The field area is monitored by sensors like Agro Floats (NetCDF), Buoys (CSV), Coastal Radars

(TUV), Gliders (NetCDF), Sonde (Excel/CSV) and others. Weather data is observed and recorded in a successive interval of time which results in big data that comprises of structured, semi-structured and unstructured data. The problems in big data which involves multiple data sources are semantic heterogeneity and structural heterogeneity. Semantic heterogeneity refers to data that is inconsistent with each other and unable to link the understanding between the data sets. Structural heterogeneity refers to different data stored in different model or structure. The proposed method solves the heterogeneity problem through data integration where, the heterogeneous data files are integrated into a machine-readable standard format called RDF by following a standard naming convention for ocean parameter vocabularies namely IOOS. The research has been performed on a 64-bit Intel Core i5 processor with 4 GB of RAM with 2 TB hard disk and deployed in the LINUX (Ubuntu16.04 version) system. The java (jdk 1.8.0_181) code is developed on Eclipse 5.0, Apache Jena with Apache Tomcat 9.0.14, Java Servlet Pages (JSP) as a server and Internet Explorer or any web browser as a client.

The ocean datasets are collected from the Indian Meteorological Department (IMD) to implement the proposed framework whose characteristics are mentioned in Table 5. If one has to deal with the data, it must be harmonized before using it. Hence, the collected heterogeneous data files are integrated into a semantic web supportable format called RDF. The data integration phase is important 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. For instance, the conversion output of a CSV file into RDF is represented in Fig. 6 and the NetCDF file into RDF is illustrated in Fig. 7. It can be noticed that each data is represented as RDF statements by adding triples to it ($<$s,p,o$>$). Once the datasets are presented in a standard RDF, there are many tools available for visualizing and working with the information stored in it.

After the integration of heterogeneous data into RDF, a knowledge representation is built in the second phase with the specifications mentioned in Section 3.2 that describes the meaning of the data. This paper uses protégé 5.1 tool to build the proposed OWO ontology and H2 database as the data source for storing the attribute values for ocean weather applications. The developed ontology follows FAIR data principles by using IOOS standard vocabularies that follows FAIR data principles, using formal and broadly applicable language namely OWL, providing domain-relevant standard to represent the data namely RDF and accessible or retrievable through a unique identifier using SPARQL queries. OWO includes the hierarchy of weather conditions, time, and the attributes related to the weather conditions and the relationship between them.

The elements of ocean weather phenomenon are described by a knowledge graph which provides a better representation of data through concepts like weather attributes, weather condition, geolocation etc., that are further classified into sub-concepts like wind_speed, humidity, precipitation_rate, latitude, longitude etc. Further the sub-concepts are classified into number of instances related to that particular attribute which carries a range of data values through data property. Data property relates the instance to its literal values defining a data type; for example, the instance "light breeze" of the concept "wind_speed" holds the literal value ranges from 1.6 to 3.3 m/s of the decimal data type. Onto graph of the proposed OWO ontology is illustrated in Fig. 8. The same information has been deployed in the H2 relational database as tables by describing their relationship. Then ontop mapping tool is used to map the data from the H2 relational database to the developed OWO ontology.

Data interpretation through databases is a time consuming process and creates scalability issues in terms of information retrieval hence, ontology is a best way to represent the knowledge of the data. Before building ontology the developer should generate consent about one conceptualization of the application domain. The conceptualization is created by choosing the lexical terms that supports the entity and lexical relations in ontology. Further, the constraints, rules and procedure are essential to achieve an understanding about the domain's semantics. This expresses the agreement that how such applications, implemented as software agencies may commit to ontology. Furthermore, maintaining the consistency in ontology will be the responsibility of application domain assumed with the help of ontology. For instance, it is easy to agree that "wind has a pressure", while difficult to agree that the pressure value is "high" or "low" and whether the range of values "affects the weather condition or not".

While developing the OWO ontology first a formal ontology is defined in a logic sense that consists of all the possible conceptualizations of the real world application domain. Then a formal ontology base is created with rues and commitments in the commitment layer which has a set of context-specific facts called lexons. Both layer together forms a scalable ontological model. This also leads to add new information sources without affecting any substantial changes in the ontological components. For instance, the developed weather ontology consists of concepts and relation between the concepts whereas the commitment layer consists of the conditions of weather parameters that affect the weather condition. Hence the ontology provides a naturally extended database modeling theory and practice that leads to scalable solution for ontology-based systems. The layered architecture improves scalability, where the rules and constraints are moved to the commitment layer that makes the developer easy to add the lexons to the ontology base without affecting the ontological commitments.

In the third phase, the input data and the proposed ontology are mapped with the knowledge of the domain experts. Once the mapping is completed, the SPARQL query is applied to the RDF graph to extract the information from the stored dataset to the user-defined query. The triples in the SPARQL syntax are to match with the triples of RDF to provide the output value. The satellite ocean data set's overall IR process is illustrated, as shown in Fig. 9. Here, an Excel (*.xls) file is taken, which stores the weather parameters' values along with the geographical location recorded by the buoy named BD14. The next stage represents the developed weather ontology describing the concepts and their relationships with the application domain.

In the final section of Fig. 9, it is noted that the user has queried for the value of the parameter "wind_speed" recorded by the buoy "BD14" from the Sensor Observation Services (SOS) tab. The user has the facility to choose the location, date and time in which the data has to be queried and the output format such as XML, JSON and Table. For instance, the user has given the "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 the specified date and time with latitude value "7.007" and longitude value "88.005" from the input Excel file. It has been illustrated that the value of "wind_speed" is found to be "1.91"according to the user preferences in the query. Then the server maps the extracted value "1.91" with the knowledge base and identifies the range under which this value falls and instances of the particular range. As explained before, the value "1.91" falls under the range "1.6 to 3.3 m/s" of the concept "wind_speed", and the name of the ontology instance has been specified as "Light breeze". Finally, the result will be displayed in "JSON" format with latitude value, longitude value and date and time along with the resultant value "1.91" and the instance name "Light breeze".

In big data information retrieval system the evaluation of system performance refers to the critical assessment of the degree to which a service fulfills the stated goals' of any end user. The two basic parameters that are defined to measure the performance of an IR system are effectiveness and efficiency. Effectiveness defines to the level up to which the given system attains the objective. On the other hand, efficiency refers to how well the system helps in achieving the user objectives. The factors for evaluating information systems include coverage, precision, recall, F-measure and presentation of results to the user. The drastic increase in the use of big data applications leads the developer to write efficient search queries for information retrieval systems. Ontologies help in data representation through

Table 6 Sample Ontologies (Weather Domain)

Ref ID	Weather Ontologies
GS	WeatherOntology.owl
O1	Open Weather Map
O2	Weather Phenomenon Prediction
O3	Weather Developer
O4	World Weather Marine Online
PO	Proposed OWO Ontology

knowledge graph and interactive query generation which provides an interface between the data and search requests. Moreover, the ontology-based information retrieval, database-to-ontology transformations and ontology-to-database mappings enhances the searching capabilities for massively loaded information management systems. The coverage of ontology refers to provide sufficient concept coverage of the domain knowledge which is determined by considering the vocabulary coverage (Cov_C^ω, Cov_I^ω, Cov_A^ω, Cov_R^ω and Cov^ω) and semantic coverage ($SCov_C^\omega$, $SCov_I^\omega$, $SCov_A^\omega$, $SCov_R^\omega$ and $SCov^\omega$) factors. Greater the coverage of concepts of an application domain the ontology is said to have a higher coverage factor and is said to be complete.

Similarly in precision-recall matrix there are four types of information namely; hints – retrieved relevant information (let us consider it as 'p'), noise – retrieved non-relevant information (let us consider it as 'q'), misses – non-retrieved relevant information (let us consider it as 'r'), and rejected – non-retrieved non-relevant information (let us consider it as 's'). The precision factor includes the ratio of relevant data retrieved to that of total data retrieved as given in Eq. 21 which is represented as $P = [p/(p+q)] \times 100$. Similarly, the recall factor includes the ratio of relevant data retrieved to that of total relevant data in the database as given in Eq. 22 which is represented as $R = [p/(p+r)] \times 100$. Whereas the F-measure is defined as the harmonic average of the precision and recall as mentioned in Eq. 23. The value of recall is increased with the increase in the value of p which refers to the retrieval of greater number of relevant data.

Various search engines are in use for the purpose of information retrieval from a complex and copious data sets namely; keyword based search, Universal Networking Language (UNL) based search, conceptual based search and ontology based search and so on. Keyword based information is not able

to incorporate the semantic of the queries hence; the process of relevant information retrieval is made difficult. Whereas UNL and conceptual based search provides a semantic link of the queries but achieves lesser rate in terms of precision and recall values than the ontology-based search engines. According to Thenmalar and Geetha (2014) ontology-based search engine achieves 79.54% improvement in precision, 73.68% improvement in recall and 73.17% improvement in F-measure and when compare to keyword based search engine. Similarly, in comparison with UNL based search engine it achieves 27.41% improvement in precision, 57.14% improvement in recall and 42% improvement in F-measure. Eventually when compared to the conceptual based search engine it achieves 21.53% improvement in precision, 29.41% improvement in recall and 24.56% improvement in F-measure. The ontology-based search system shows better results, due to the expanded concepts with ontological relations, and enhanced query cases for obtaining more relevant information. Finally, the presentation of the query results is provided to the users in human readable formats such as XML, JSON and Table. The proposed ontology is analyzed to have higher coverage factor as evaluated in Section 4.2.1. The proposed method uses the ontology-based semantic search engine that results in higher precision, recall and F-measure values compared to other search engines. Also, the provided query results are easily understandable by any end users.

Performance evaluation

The performance metrics are calculated for the developed OWO ontology using the expressions discussed in Section 3.4. A study has been carried out with some existing ontologies O1 (IT Research Sector in Satellite Data Processing n.d.), O2 (Roy 2017), O3 (World Weather Online Developer n.d.) and O4 (Yahoo Weather Developer Network et al. n.d.) as shown in Table 6 on weather domain to demonstrate the proposed method's quality. The performance of any ontology is often evaluated against the other ontology called the GS ontology. The GS ontology of weather domain developed by Automation Systems Group, Technical University of Wein (Kastner 2013) is considered. The elements of ontology like classes, instances, attributes and relations are extracted from the ontologies are identified and presented in Table 7.

Table 7 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 8 Classes defined in ontologies to GS

Defining classes	O1 (WO$_1$)	O2 (WO$_2$)	O2 (WO$_3$)	O4 (WO$_4$)	PO (WO$_5$)
Defined both in GS and WO$_n$	7	9	6	11	12
Defined in GS and derivable from WO$_n$	1	1	0	0	1
Defined in GS but not in WO$_n$	7	6	9	4	2
Defined in WO$_n$ but not in GS	4	8	13	24	24

In some cases, the classes defined in the ontology can also be derivable from other elements defined in that ontology. In that case, those elements have a major contribution in calculating redundancy metrics and semantic coverage. The classes defined in weather ontologies to the golden ontology are analyzed and presented in Table 8.

Evaluation of results

The performance metrics of ontologies expressed in Section 3.4 are evaluated on the sample weather ontologies, and the results are mentioned in Table 9. Using the extracted elements of ontologies, as mentioned in Table 7 and 8, the metrics are estimated through the respective equations. This paper evaluates four major quality factors of ontology, namely completeness (COM), correctness (COR), conciseness (CON), and structural complexity (SC), where the metrics involved are as shown in Tab. 5. The performance metrics are segregated according to the quality factors. Table 10 explains the metrics involved in measuring each quality factor.

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 to the GS ontology than the existing ontologies. It can be noted from Fig. 10 that the proposed ontology results in the highest score of vocabulary coverage with $Cov^\omega(O)$

Table 9 Performance evaluation of various weather ontologies

Factors	Performance metrics	O1	O2	O3	O4	PO
Vocabulary coverage	Cov^ω_C	0.57	0.6	0.4	0.73	0.8
	Cov^ω_I	0.64	0.54	0	0.91	0.91
	Cov^ω_A	0.56	0.61	0.44	0.88	0.94
	Cov^ω_R	0.52	0.48	0.19	0.26	0.66
	Cov^ω	0.54	0.54	0.22	0.61	0.77
Semantic coverage	$SCov^\omega_C$	0.6	0.66	0.4	0.73	0.86
	$SCov^\omega_I$	0.64	0.54	0	0.91	0.91
	$SCov^\omega_A$	0.56	0.61	0.44	0.88	0.94
	$SCov^\omega_R$	0.52	0.48	0.19	0.26	0.66
	$SCov^\omega$	0.54	0.56	0.22	0.61	0.78
Semantic compatibility	RCC^ω	0.38	0.41	0.52	0.6	0.66
	$ARCI^\omega$	0.5	0.55	0.66	0.63	0.75
	$ARCA^\omega$	0.87	0.88	0.83	1	1
	$ARCR^\omega$	0.42	0.46	0.6	0.57	0.62
Redundancy	CR	0	0	0	0	0
	IR	0	0	0	0	0
	AR	0	0	0	0	0
	RR	1	0.97	1	0.93	0.89
Coherence (Relation)	NRN	2	3	1	9	4
	NLN	9	32	20	55	34
	$MaxSPL$	3	2	6	5	3
	NIC	0	0	0	0	0
	$TNRNR$	11	34	35	94	39
	$ANRNR$	5.5	18.02	35	10.44	9.75
Coherence (Acyclic)	$ADLN$	2.33	4.12	4.0	3.27	3.05
	$AWNLN$	2.75	3.26	6.5	1.81	2.09
	$MaxDepth$	3	4	6	5	3
	$MaxWidth$	7	8	7	10	8

Table 10 Quality factors of ontologies

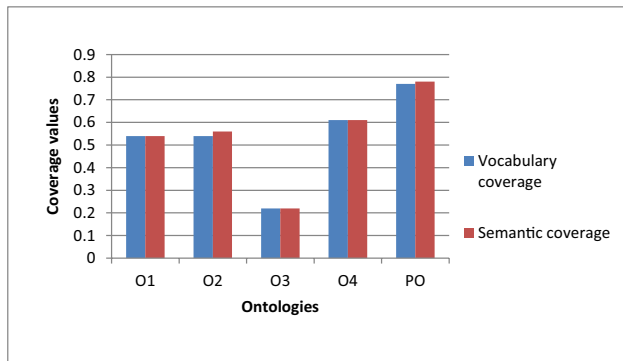| S.no	Quality factors	Performance metrics involved
1	Completeness	$Cov_C^\omega, Cov_I^\omega, Cov_A^\omega, Cov_R^\omega, Cov^\omega,$ $SCov_C^\omega, SCov_I^\omega, SCov_A^\omega, SCov_R^\omega, SCov^\omega$
2	Correctness	$RCC^\omega, ARCI^\omega, ARCA^\omega, ARCR^\omega$
3	Conciseness	CR, IR, AR, RR
4	Structural complexity	$NRN, NLN, MaxSPL, NIC, TNRNR, ANRNR,$ $ADLN, AWNLN, MaxDepth, MaxWidth$

Fig. 10 Semantic coverage of ontologies

as 0.77 and semantic $SCov^\omega(O)$ as 0.79, which is high among the other ontologies. Similarly, ontology of the same domain consists of redundant elements, which illustrate the conciseness of ontology as presented in Fig. 11. Ontology is said to be efficient and concise only if it has minor redundancy elements. Results show that redundant elements are present only in the relations r of ontologies. The value of RR is estimated to be very low in the proposed ontology, 0.89, compared to the ontologies 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 are not domain
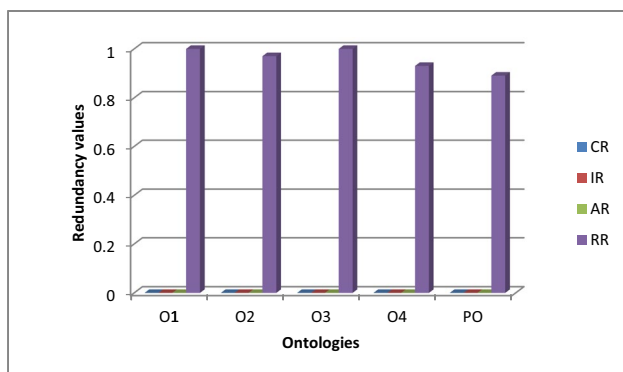
knowledge but rather subject to a particular web service. Hence, all the elements mentioned are not to be equivalent to GS ontology. Figure 12 illustrates that the compatibility of PO is high compared to the other ontologies. Where, $RCC^\omega, ARCI^\omega, ARCA^\omega$ and $ARCR^\omega$ score 0.66, 0.75, 1 and 0.62 respectively. Similarly, cohesion metrics evaluate 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. 13. 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.

The comparison of quality metrics is aggregated into a single value by taking an average and plotted in Fig. 14. The completeness and correctness metrics of the proposed ontology are high at 0.77 and 0.75, respectively; hence, it covers the domain most. Similarly, the redundancy score is 0.89 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 result analyses that the proposed ontology is effective in the ways 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 concluded to be high compared to the existing ontologies.
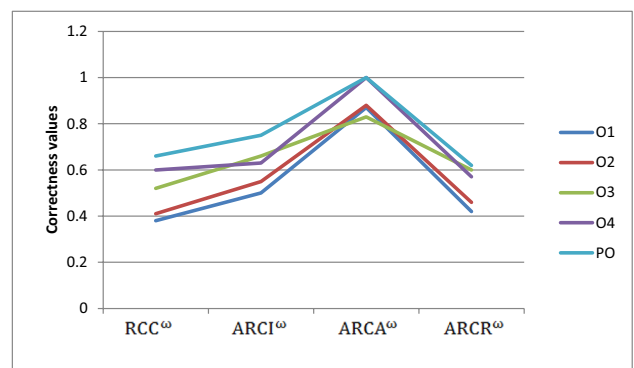

Fig. 11 Redundancy metrics of ontologies

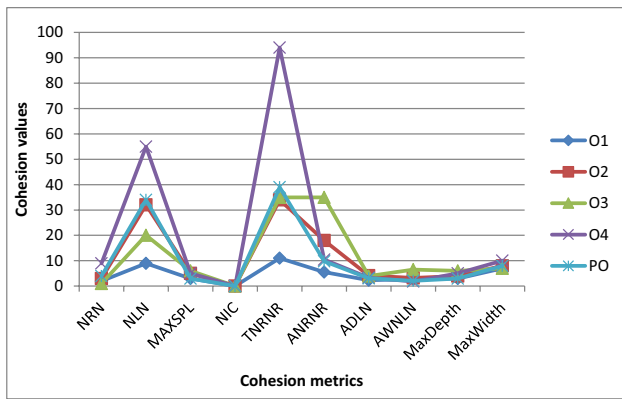**Fig. 12** Compatibility chart of ontologies

Fig. 13 Distribution of coherence metrics of weather ontologies
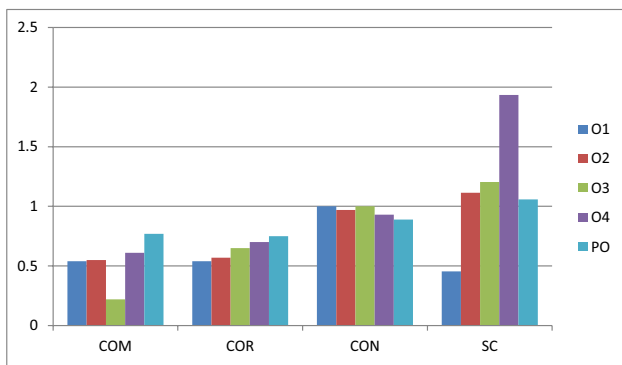

Fig. 14 Comparison of quality factors of ontologies with proposed ontology

Conclusion and future work

This paper presents a weather data model by integrating big data with semantic web technologies and incur with structured, semi-structured and unstructured data. The proposed framework permits the user to aggregate, link, integrate and represent geospatial climatic data from variety of sources using semantic web technologies. The satellite data sources are aggregated semantically and integrated into a machine-understandable format called RDF. Then a knowledge representation of the data is built by using ontology to solve semantic and structural heterogeneity. OWO ontology has been created using protégé 5.1 tool, and the similar attribute values are stored in H2DB written in java. The mapping of ontology with H2 database is carried out using JDBC driver, which helps to query the information via SPARQL. The proposed ontology's performance metrics is analyzed to be 39.28% improved by completeness; 45.29% decreased in

structural complexity, 11%, and 37.7% reduced in conciseness and correctness respectively. This approach sustains various scientific domains, research data sharing, semantic query execution and efficient visualization of results. The future work of this research aims to combine the blockchain technology with semantic web that can solve a wide-range of problems in different domains. Blockchain is a persistent technology implemented in a number of sectors like industry, research and academy. Recently many researchers have shown interest in combining it with semantic web but the implementation has not been done yet. Hence, the future work attempts to combine the semantic web with blockchain technology due to its advantages over the big data by reducing the efforts to user and researches.

Data availability The datasets analyzed during the current study are available in the Meteorological & Oceanographic Satellite Data Archival Centre **(MOSDAC)** repository, [https://www.mosdac.gov.in/open-data].

Code availability Not applicable.

Declarations

Conflicts of Interest/Competing interests The authors declare that they have no conflicts or competing interests.

References

Abburu S, Golla SB (2013) Survey on Ontology Construction Tools. Int J Sci Eng Res 4:1748–1752

Abburu S, Dube N, Miyar RN, Golla SB (2015) An Ontology Based Methodology for Satellite Data Semantic Interoperability. Adv Electr Comput Eng 15:105–110

Ahmed J, Ahmed M (2018) Big data and semantic web, challenges and opportunities a survey. Int J Eng Technol 7:631–633

Ali F, Kwak D, Khan P, EI-Sappagh S, Islam SMR, Park D, Kwak KS (2017) Merged Ontology and SVM-based Information Extraction and Recommendation System for Social Robots. IEEE Access 5:12364–12379

Atemezing G, Corcho O, Garijo D, Mora J, Poveda-Villalón M, Rozas P, Vila-Suero D, Villazón-Terrazas B (2011) Transforming Meteorological Data into Linked Data. Semantic Web Journal 1:1–5

Bai Y, Yan A, Lv B (2012) Modeling Earth Observation Satellite Data Semantically. Proceedings of the 8th IEEE International Conference on eScience (eScience 2012) 1:8–12

Bally J, Boneh T, Nicholson AE, Korb KB (2004) Developing an ontology for the meteorological forecasting process. Decision Support in an Uncertain and Complex World: Proceedings of the 2004 IFIP International Conference on Decision Support Systems (DSS 2004) 70–81.

Bansal SK (2014) Towards a Semantic Extract-Transform-Load (ETL) framework for Big Data Integration. IEEE International Congress on Big Data, 2014

Bansal SK, Kagemann S (2014) Integrating big data: a semantic extract-transform-load framework. Computer 48(3):42–50

Bansal SK, Kagemann S (2015) Integrating Big Data: A Semantic Extract Transform-Load Framework. IEEE Comput Soc 48(3):42–50

Beaufort F Beaufort scale: Royal Meteorological Society. https://en.wikipedia.org/wiki/Beaufort_scale. Accessed 20 June 2018

Belhadi H, Akli-Astouati K, Djenouri Y, Lin JCW (2019) Exploring pattern mining for solving the ontology matching problem. Proc Eur Conf Adv Databases Inf Syst 1:85–93

Belhadi H, Akli-Astouati K, Djenouri Y, Lin JCW (2020) Data mining-based approach for ontology matching problem. Appl Intell 50:1204–1221

Bonstrom V, Hinze A, Schweppe H (2003) 'Storing RDF as a Graph', Proceedings of the first Latin American Web Congress, pp. 27–36

Catherine R, Bernard S, André G, Corcho O, Sousa DG, Boffety D, Chanet JP (2014) Weather station data publication at Irstea: An implementation report. Context-Aware Approach for e-Agriculture 1401:89–104

Chen M, Plale B (2013) From Metadata to Ontology Representation: A Case of Converting Severe Weather Forecast Metadata to an Ontology. Proc Am Soc Inf Sci Technol 49:1–4

Chughtai MW, Elahi F, Akosu N (2017) Semantic web requirement engineering boundaries. J Softw Eng Intell Syst 2:57–61

Cruz IF, Ganesh VR, Mirrezaei SI (2013) Semantic extraction of geographic data from web tables for big data integration. In: Proceedings of the 7th Workshop on Geographic Information Retrieval, Orlando, FL, USA, pp. 19–26. ACM

Devaraju A, Kauppinen T (2012) Sensors Tell More than They Sense: Modeling and Reasoning about Sensor Observations for Understanding Weather Events. Int J Sensors 2:14–26

Engineering Tool Box. Recommended Relative Humidity. https://www.engineeringtoolbox.com/relative-humidity-d_895.html. Accessed 03 Jul 2018

Gillies A (1997) Software quality: Theory and Management. International Thomson Computer press

Gollapudi S (2015) Aggregating financial services data without assumptions: a semantic data reference architecture. In: 2015 IEEE International Conference on Semantic Computing (ICSC), Anaheim, CA, USA, pp. 312–315

Gupta S, Jain I, Johari P, Lal M (2019) Impact of Climate Change on Tropical Cyclones Frequency and Intensity on Indian Coasts. Proceedings of International Conference on Remote Sensing for Disaster Management 359–365.

Haby J (2014) 1001 Haby Hints for Meteorology and Weather Prediction. The Weather Prediction. http://www.theweatherprediction.com/habyhintsbook/. Accessed 18 Jul 2018

Haines S, Subramanian V, Mayorga E, Snowden D, Ragsdale R, Rueda C, Howard M (2012) IOOS Vocabulary and Ontology Strategy for Observed Properties. Oceans 1(9):14–19

Hasany N, Alwatban B (2017) Inference Based Interpretation of Keyword Queries for OWL Ontology. Int J Web Semant Technol 8:1–9

IT Research Sector in Satellite Data Processing. Open Weather Map Ontology. https://openweathermap.org. Accessed 10 Jan 2020

Kaladevi R, Revathi A, Sridhar S (2020) Development of Background Ontology for Weather Systems through Ontology Learning. Int J Recent Technol Eng 8:168–172

Kalibatiene D, Vasilecas O (2011) Survey on Ontology Languages. Proceedings of 10th International Conference on Perspectives in Business Informatics Research 124–141.

Kambli K (2020) Throwback to a Stormy Year: A Look at the 8 North Indian Ocean Cyclones of 2019. The Times of India. https://weather.com/en-IN/india/news/news/2020-01-03-stormy-year-8-north-indian-ocean-cyclones-2019. Accessed 11 Dec 2020

Kang L, Yi L, Dong L (2014) Research on construction methods of big data semantic model. In: Proceedings of the World Congress on Engineering (WCE 2014), vol. 1, London, UK

Kantamneni A, Brown LE (2018) An Ontology for Solar Irradiation Forecast Models. Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management 2:263–270.

Kastner W (2013) Weather ontology developed by Automation Systems Group, The Institute of Computer Aided Automation. https://algo4rail.auto.tuwien.ac.at/~k/thinkhome/ontologies.html. Accessed 05 Apr 2020

Khelkar P (2018) Watch | Huge water tornado at Nazare Dam in Pune stuns locals. India Today. https://www.indiatoday.in/india/story/watch-huge-water-tornado-at-nazare-dam-in-pune-stuns-locals-1256546-2018-06-10. Accessed 08 Jan 2019

Koll RM, Murtugudde R (2019) New IPCC report warns of marine heat waves, extremely severe cyclones. India Science Wire. https://vigyanprasar.gov.in/isw/New-IPCC-report-warns-of-marine-heat-waves-extremely-severe-cyclones.html. Accessed 09 Dec 2020

Kossin JP, Knapp KR, Olander TL, Velden CS (2020) Global increase in major tropical cyclone exceedance probability over the past four decades. PNAS 117:11975–11980

Ma X (2017) Linked Geoscience Data in practice: where W3C standards meet domain knowledge, data visualization and OGC standards. Earth Sci Inform 10:429–441. https://doi.org/10.1007/s12145-017-0304-8

Ma L, Mei J, Pan Y, Kulkarni K, Fokoue A, Ranganathan A (2007) Semantic Web Technologies and Data Management. IBM China Research Laboratory

Measurement of Precipitation (2018) Guide to Instruments and Methods of Observation. World Meteorological Organization 214–245

Mohsen W, Aref M, ElBahnasy K (2020) Cooperative Domain Ontology Reduction Based on Rough Sets. Int J Comput Sci Mob Comput 9:163–174

Morsey M, Lehmann J, Auer S, Stadler C, Hellmann S (2012) DBpedia and the live extraction of structured data from Wikipedia. Program Electron Libr Inf Syst 46:157–181

Naidu TA (2020) Waterspout leaves locals puzzled in Yanam. The Hindu. https://www.thehindu.com/news/national/andhra-pradesh/waterspout-leaves-locals-puzzled-in-yanam/article32118473.ece. Accessed 05 Dec 2020

Neches R, Fikes R, Finin TW, Gruber TR, Patil RS, Senator TE, Swartout WR (1991) Enabling Technology for Knowledge Sharing". AI Mag 12(3):36–56

Nishanbaev I, Champion E, McMeekin DA (2019) A Survey of Geospatial Semantic Web for Cultural Heritage. Heritage 2:1471–1498

Peng P, Zou L, Özsu MT, Lei CL, Zhao D (2016) Processing SPARQL queries over distributed RDF graphs. VLDB J Int J Very Large Data Bases 25:243–268

Pokharel S, Sherif MA, Lehmann J (2014) Ontology Based Data Access and Integration for Improving the Effectiveness of Farming in Nepal. Proceedings of WI-IAT '14: IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) 2:319–326.

Pouchard LC, Branstetter ML, Cook RB et al (2013) A Linked Science investigation: enhancing climate change data discovery with semantic technologies. Earth Sci Inform 6:175–185. https://doi.org/10.1007/s12145-013-0118-2

Poveda-Villaló M, Espinoza-Arias P, Garijo D, Corcho O (2020) Coming to Terms with FAIR Ontologies. Proc Knowl Eng Knowl Manag 12387:255–270. https://doi.org/10.1007/978-3-030-61244-3_18

Raad J, Cruz C (2015) A Survey on Ontology Evaluation Methods. Proceedings of 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) 631:16–23

Rio J, Toma DM, Martinez E, Thomas CR, Delory E, Pearlman JS, Waldmann C, Jirka S (2018) A Sensor Web Architecture for Integrating Smart Oceanographic Sensors into the Semantic Sensor Web. IEEE J Oceanic Eng 43:830–842

Roy RK (2017) Weather Phenomenon Prediction Using Semantic Web. Dissertation, East West University.

Rudnicki R (2019) An Overview of the Common Core Ontologies (Report No. 716–204–5208). CUBRC, Inc

Saber A, Al-Zoghby AM, Elmougy S (2018) Big-Data Aggregating, Linking, Integrating and Representing Using Semantic Web Technologies. In: Hassanien A., Tolba M., Elhoseny M., Mostafa M. (eds) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018). AMLTA 2018. Advances in Intelligent Systems and Computing, vol 723. Springer, Cham. https://doi.org/10.1007/978-3-319-74690-6_33

Sarkar S (2020) Cyclones rise as climate change heats up Indian Ocean. India Climate Dialogue. https://indiaclimatedialogue.net/2020/06/05/cyclones-rise-as-climate-change-heats-up-indian-ocean/. Accessed 20 Dec 2020

Shao C, Hu LM, Li JZ, Wang ZC, Chung T, Xia JB (2019) RiMOM-IM: a novel iterative framework for instance matching. J Comput Sci Technol 31:185–197

Shen YT, Lai JW, Leu LG, Lu YC, Chen JM, Shao HJ, Chen HW, Chang KT, Terng CT, Chang YC, Tseng RS (2019) Applications of ocean currents data from high-frequency radars and current profilers to search and rescue missions around Taiwan. J Oper Oceanogr 12:126–136

Shenoi SSC, Nasnodkar N, Rajesh G, Joseph KJ, Suresh I, Almeida AM (2009) On the diurnal ranges of Sea Surface Temperature (SST) in the north Indian Ocean. J Earth Syst Sci 118:483–496

Staroch P (2013) A Weather Ontology for Predictive Control in Smart Homes [Master's thesis, Vienna University of Technology]. TU wein repositum. https://repositum.tuwien.at/bitstream/20.500.12708/8079/2. Accessed 02 Feb 2019

Stegmaier F, Gröbner U, Döller M, Kosch H, Baese G (2009) Evaluation of current RDF database solutions. Proceedings of the 10th International Workshop on Semantic Multimedia Database Technologies 1–17.

Thenmalar S, Geetha TV (2014) Enhanced ontology-based indexing and searching. Aslib J Inf Manag 66(6):678–696. https://doi.org/10.1108/AJIM-08-2014-0098

Wilson A, Cox M, Elsborg D et al (2015) A semantically enabled metadata repository for scientific data. Earth Sci Inform 8:649–661. https://doi.org/10.1007/s12145-014-0175-1

World Weather Online Developer. World Wide Marine Online. https://www.worldweatheronline.com. Accessed 13 Jan 2020

World Wide Web Consortium (2004) Resource Description Framework (RDF)/ Extensible Markup Language (XML) Syntax Specification. Available from: <www.w3.org/TR/REC-rdf-syntax/>. Accessed 8 Feb 2018

Xiong J, Liu Y, Liu W (2014) Ontology-based integration and sharing of big data educational resources. In: IEEE 11th Web Information System and Application Conference (WISA), Tianjin, China, pp. 245–248

Yahoo Weather Developer Network. Weather Developer ontology. https://developer.yahoo.com/weather. Accessed 21 Jan 2020

Zavitsanos E, Paliouras G, Vouros GA (2011) Gold Standard Evaluation of Ontology Learning Methods through Ontology Transformation and Alignment. IEEE Trans Knowl Data Eng 23:58–71

Zhang F, Li Z, Peng D et al (2021) RDF for temporal data management – a survey. Earth Sci Inform 14:563–599. https://doi.org/10.1007/s12145-021-00574-w

Zhang Y, Chiang YY, Szekely P, Knoblock CA (2013) A semantic approach to retrieving, linking, and integrating heterogeneous geospatial data. In: Joint Proceedings of the Workshop on AI Problems and Approaches for Intelligent Environments and Workshop on Semantic Cities, pp. 31–37. ACM

Zidan RA, Almustaf MM, Tahawi M (2019) Ontology2rdb for Storing Ontology in Relational Database. J Theor Appl Inf Technol 97:2229–2240

Zu H, Liu D, Bayley L, Aldea A, Yang Y, Ying C (2017) Quality Model and Metrics of Ontology for Semantic Descriptions of Web Services. Tsinghua Science and Technology 22:254–272